Mechanisms of Epigenetic Reader Proteins Interacting with Chromatin

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Khan L. Cox, B.S., M.S.
Graduate Program in Physics

The Ohio State University
2024

Dissertation Committee:
Michael G. Poirier, Advisor
Ralf Bundschuh
Ezekiel Johnston-Halperin
Comert Kural

Abstract

DNA is compacted in the eukaryotic cell nucleus into chromatin by wrapping ~147 bp of DNA at a time around protein histone octamer cores composed of 2 each of histones H2A, H2B, H3, and H4. Repeats of this DNA-protein complex, which is referred to as the nucleosome, form higher order chromatin structure, and it plays a large role in gene transcription regulation by controlling access to DNA. Multiple mechanisms exist to overcome the nucleosomal barrier to transcription. Each of the histones in the nucleosome can have post-translational modifications (PTMs) made to specific residues in the histone as a way to directly affect the structure of the nucleosome, or signal for other cellular machinery to open access to the DNA. These modifications can affect each other across the nucleosomes in a cell. Often the deposition of one type of modification will affect the state of another type of histone modification on the surrounding nucleosomes. This network of histone PTMs contributes to the epigenetic landscape of the cell directing cellular machinery in the nucleus to perform its various functions, and has been referred to as the histone code mechanism. In addition, there exist proteins capable of depositing these modifications on histones, binding specifically to the modifications, or removing them from histones, and these proteins are considered the writers, readers, and erasers of the histone code. A certain number of these proteins do not exclusively perform only one of these functions, and will contribute to the network of histone PTMs in a more complex manner by reading and targeting one PTM type while being responsible for depositing or removing another PTM type for instance. The category of histone PTM related proteins is vast, and so this thesis will focus on characterizing histone PTM reader type proteins and their interactions with modified histone chromatin in vitro. Work in this thesis showed that PTM readers are capable of a wide variety of functions and methods of binding to chromatin while increasing or decreasing accessibility of said chromatin 2-fold or more to transcription factors.

Acknowledgements

To my parents, thank you for all the support and encouragement you gave me during this time, and for being there for me during periods of extreme burnout.

To my current and former colleagues in the lab: Kevin Jamison, Nathan Burge, Kristin Chestnut, Blanche Chen, Kalven Bonin, Michael Neuhoff, Ehsan Akbari, Robin Meng, Cassie Robbins, Eun Jun, Carter Mason, Siamak Shokri, Golbarg Mohammadiroozbahani, Ariel Robbins, Ben Donovan, Matt Brehove, Michael Darcy, and Matt Gibson, thank you for being flexible with instrument reservations and for providing assistance or advice at times, as well as all the interesting small talk.

To my advisor, Dr. Michael Poirier, thank you for your guidance and for your efforts in making the lab a productive environment.

To my committee: Dr. Ralf Bundschuh, Dr. Zeke Johnston-Halperin, and Dr. Comert Kural, thank you for your guidance through my graduate career.

To my collaborators: Thank you to Nick Sapp and Dr. Chandravanu for giving me the opportunity to be a part of your HIV-1 study, and thank you to Dr. Kutateladze and the students in her lab for the opportunity to study so many important proteins.

Vita

December, 2014 ..... B.S, University of Illinois at Urbana-Champaign

December, 2017 ..... M.S., The Ohio State University

Aug, 2015 – Dec, 2016 ..... Graduate Teaching Associate, The Ohio State University

Jan, 2017 – present ..... Graduate Research Associate, The Ohio State University

Publications

Gatchalian, J., Wang, X., Ikebe, J., Cox, K. L., Tencer, A. H., Zhang, Y., … Kutateladze, T. G. (2017). Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nature Communications*, 8(1).

Tencer, A. H., Cox, K. L., Di, L., Bridgers, J. B., Lyu, J., Wang, X., … Kutateladze, T. G. (2017). Covalent Modifications of Histone H3K9 Promote Binding of CHD3. *Cell Reports*, 21(2), 455–466. (Co-First Author)

Zhang, Y., Klein, B. J., Cox, K. L., Bertulat, B., Tencer, A. H., Holden, M. R., … Kutateladze, T. G. (2019). Mechanism for autoinhibition and activation of the MORC3 ATPase. *PNAS*, 116(13), 6111–6119.

Tencer, A. H., Cox, K. L., Wright, G. M., Zhang, Y., Petell, C. J., Klein, B. J., … Kutateladze, T. G. (2020). Molecular mechanism of the MORC4 ATPase activation. *Nature Communications*, 11(1), 1–13.

Klein, B. J., Cox, K. L., Jang, S. M., Côté, J., Poirier, M. G., & Kutateladze, T. G. (2020). Molecular Basis for the PZP Domain of BRPF1 Association with Chromatin. *Structure*, 28(1), 105-110.e3. (Co-First Author)

Klein, B. J., Deshpande, A., Cox, K. L., Xuan, F., Zandian, M., Barbosa, K., … Kutateladze, T. G. (2021). The role of the PZP domain of AF10 in acute leukemia driven by AF10 translocations. *Nature Communications*, 12(1).

Klein, B. J., Cox, K. L., Jang, S. M., Singh, R. K., Côté, J., Poirier, M. G., & Kutateladze, T. G. (2021). Structural and biophysical characterization of the nucleosome-binding PZP domain. *STAR Protocols*, 2(2). (Co-First Author)

Sapp, N., Burge, N., Cox, K., Prakash, P., Balasubramaniam, M., Thapa, S., … Dash, C. (2022). HIV-1 Preintegration Complex Preferentially Integrates the Viral DNA into Nucleosomes Containing Trimethylated Histone 3-Lysine 36 Modification and Flanking Linker DNA. *Journal of Virology*, 96(18), 1–30.

Becht, D. C., Klein, B. J., Kanai, A., Jang, S. M., Cox, K. L., Zhou, B. R., … Kutateladze, T. G. (2023). MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nature Communications*, 14(1), 697.

Becht, D. C., Kanai, A., Biswas, S., Halawa, M., Zeng, L., Cox, K. L., … Kutateladze, T. G. (2024). The winged helix domain of MORF binds CpG islands and the TAZ2 domain of p300. *iScience*, 27(4), 109367.

Gaurav, N., Kanai, A., Lachance, C., Cox, K. L., Liu, J., Grzybowski, A. T., … Kutateladze, T. G. (2024). Guiding the HBO1 complex function through the JADE subunit. *Nature Structural and Molecular Biology.*

Fields of Study

Major Field: Physics

Study of histone modification reader proteins: Michael G. Poirier

Table of Contents

List of Figures

Chapter 1. Introduction

Cells have to regulate expression of genes during each step in the cell cycle or under different environmental conditions. They can increase, reduce, activate, or silence expression of certain genes to maintain proper function or adapt to stress. Improper expression of genes results in abnormal or missing/overexpressed proteins in the cell, and is the cause of many human diseases. As such, gene regulation is a highly active field of study for the development of medical drugs and treatments, among other reasons. This thesis will focus on protein-DNA and protein-protein interactions surrounding a fundamental unit of gene regulation, the nucleosome, responsible for compacting DNA in the cell nucleus.

1.1 Chromatin and the nucleosome

Lengthwise, there is about 2 m of DNA contained within the eukaryotic cell nucleus, which only has a diameter of 6 μm. To condense the DNA into the nucleus and keep the genome organized, histone proteins bind the DNA and pack it into a structure termed chromatin (Figure 1.1). Besides reduction of overall volume, organizing the DNA into chromatin has other purposes in the cell. It helps maintain separation of the genome at the largest scale during mitosis, and the different orders of compaction that chromatin introduces adds extra layers of gene expression regulation, DNA repair, and DNA replication [48].

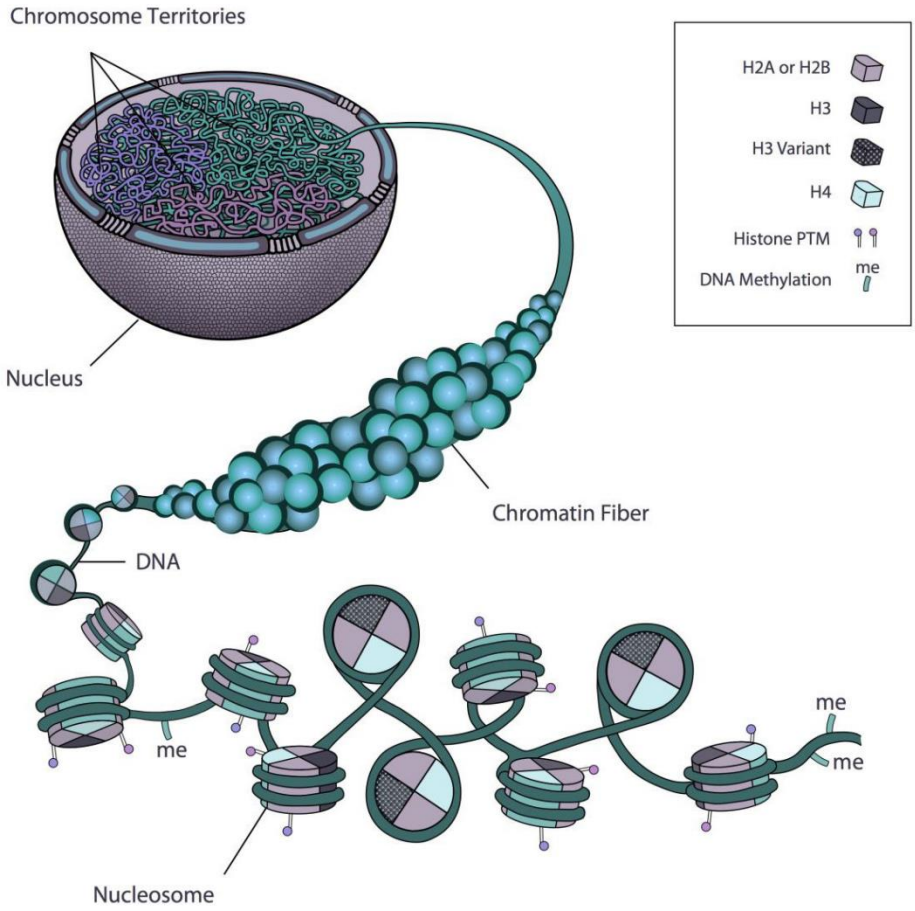


Figure 1.1: Organizational network of chromatin in the cell. Scheme depicting different aspects of chromatin regulation. PTM, post-translational modification. Chromosome territories within the nucleus, shown in different colors, are composed of chromatin fibers, which, in turn, contain packed nucleosomes. Reproduced from [47].

Chromatin is observed in two forms: heterochromatin and euchromatin. Heterochromatin is densely packed and transcriptionally silent, while euchromatin is less compacted, gene rich and made up of genes that are actively transcribed with irregularly spaced nucleosome arrays [47]. Heterochromatin is further divided up into constitutive heterochromatin and facultative heterochromatin. Up to 45% of the human genome can be comprised of constitutive heterochromatin, which contains highly repetitive DNA, making it very gene poor, and is generally found in telomeric and pericentric regions of chromosomes. Facultative heterochromatin does not have repetitive DNA, but contains genes which have been silenced via transcription regulation pathways [49,50].

At the smallest level, the main repeating element of chromatin is the DNA-protein complex termed the nucleosome (Figure 1.2). In each nucleosome there is ~147bp of DNA wrapped ~1.7 times around a histone octamer protein core comprised of two each of histones H2A, H2B, H3, and H4. All of the histones have a conserved histone fold motif of three short alpha helices separated by two unstructured loops. This motif mediates dimerization, H3 with H4 and H2A with H2B. Two H3/H4 dimers form a tetramer via a 4 helix bundle interaction between the two H3 histones. H4 interacts with H2B via a similar 4-helix bundle interaction to form the complete octamer with two H2A/H2B dimers that were formed separately. The negatively charged DNA binds to the histone octamer mainly through electrostatic forces via a high density of positively charged arginines and lysines on the octamer surface [1,2]. Each nucleosome is separated by around 50 bp of linker DNA to form the "beads on a string" structure depicted in Figure 1.1. Nucleosomes and higher order chromatin structures are essential to transcription regulation due to their ability to sterically hinder transcription factors from binding to the DNA and initiating transcription. There are multiple ways in which the nucleosome's barrier to transcription is overcome. Many of these have to do with DNA/histone modifications, and other protein factors that interact with the constituents of the nucleosome. The least well-understood of these aspects of transcriptional regulation, however, would be histone variants. Isoforms of the histones can be deposited in a nucleosome and replace the canonical histones throughout the cell cycle in different gene regions, leading to certain effects.

There exist numerous different histone variants of each of the four histone types excluding H4 [52]. One of the most well-studied variants, H2A.Z, is most commonly found in the first, strongly positioned nucleosome after a promoter sequence [7]. Both H2A.Z and H3.3 have been shown to play a role in nucleosome destabilization, especially when both are found in the same nucleosome [54,55], and are continually expressed throughout the cell cycle while the canonical histones such as H3.1 are expressed only during S-phase [52]. Throughout the cell cycle H3.3 is localized to areas of actively transcribing genes [53]. Deposition of these histone variants is the first level of gene accessibility regulation, and can itself be regulated by modifications made to histones in the receiving nucleosome [56].

In the cell, modifications are made to the chromosomal DNA without altering the DNA sequence. These modifications, in addition to modifications made to the histone proteins, are considered part of the epigenome and play a large part in transcription regulation. The DNA can have methyl groups enzymatically added at the 5'-carbon of the pyrimidine ring in cytosines which results in gene silencing due to the occlusion of DNA binding proteins that act as or recruit transcriptional activators, or due to the recruitment of methyl-binding proteins (MBPs), which recruit transcriptional corepressor complexes [51]. Post translational modifications (PTMs) can also be made to the N-terminal tails of the histones and in the octamer core. These modifications can neutralize charges on the histone residue side chains and lead to destabilization of the octamer-DNA interactions [3]. Chromatin remodeler proteins are capable of binding to certain PTMs and relocating the octamer in order to expose the DNA for subsequent reading by a transcription factor [4, 5].

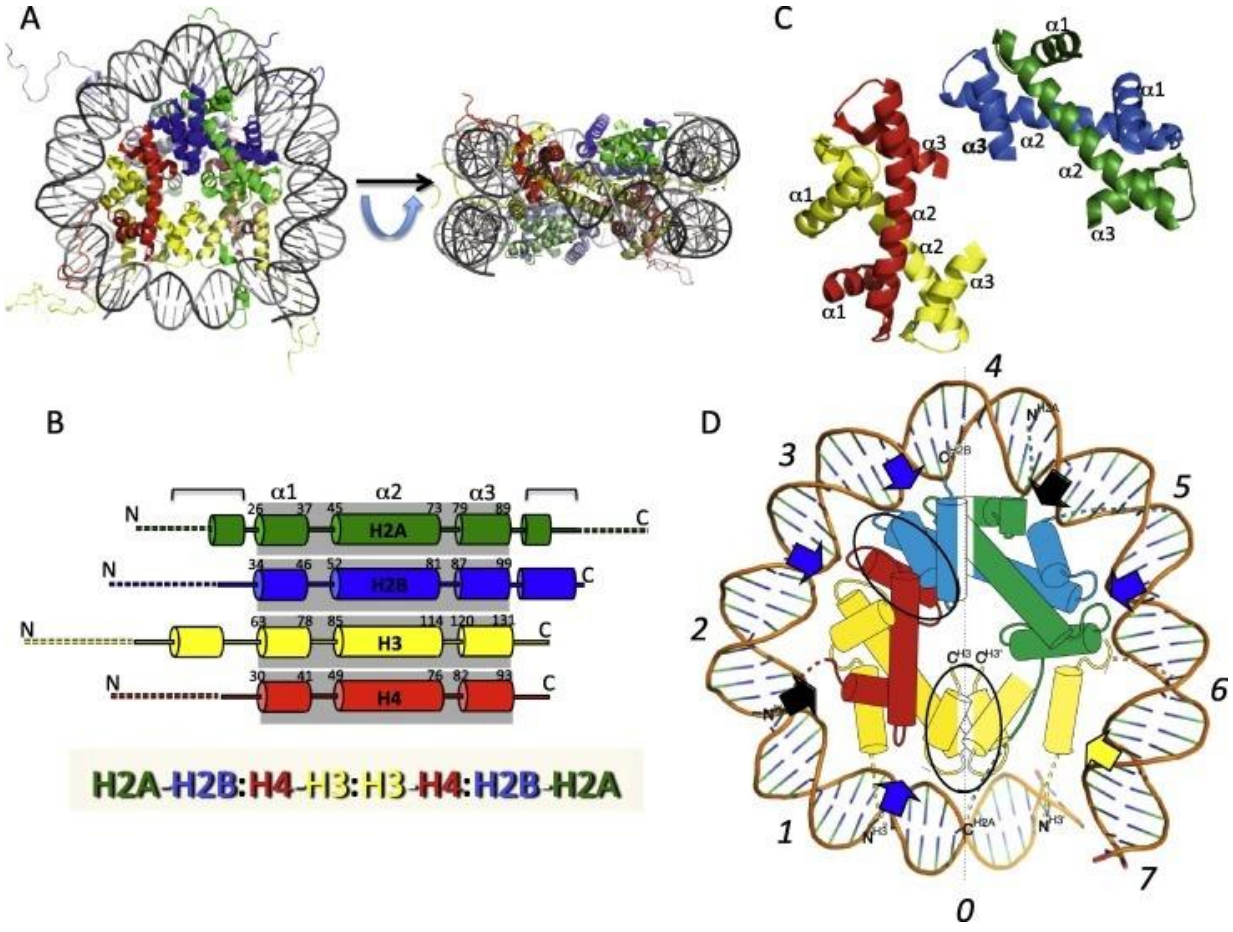


Figure 1.2: Structural details of a nucleosome core, (A) Model of a nucleosome core (PDB: 1KX5). Shown are a view down the superhelical axis, and a view rotated 90° about a horizontal axis, as shown, looking down the dyad axis of the nucleosome. H2A, green, H2B, blue, H3, yellow, H4, red. Proteins in the lower half of the nucleosome are lighter in color. (B) Top: Schematic showing secondary structure of the core histone proteins, with α-helices represented by columns. Dashed lines indicate approximate residues within 'tail' domains; shaded boxes indicate the 3-helix histone fold domains within each protein, with first and last residues within α1, α2 and α3 helices indicated. Additional helices outside the histone fold domain are indicated by brackets, Bottom: Linear representation of primary contacts between the core histone proteins in the nucleosome core. Core histone dimerization partners are separated by dashes; dimer–dimer interactions via 4-helix bundles are indicated by colons. (C) H2A-H2B (green/blue) and H3-H4 (yellow/red) histone fold domain dimers. α1, α2 and α3 helices indicated, corresponding to (B). (D) Schematic showing one-half of nucleosome core, looking down the DNA superhelix axis. Superhelix sites are indicated by italicized numerals; 4-helix bundles between H3:H3 and H4:H2B are highlighted by ovals; blue and black arrows indicate paired loop and paired-end-of-helix DNA interaction sites. Yellow arrow indicates site of interaction centered on the N-helix in H3. Note, a small amount of DNA and H3 from the non-depicted half of the nucleosome core are shown for clarity, lighter in color. Reproduced from [1]

1.2 Histone Modifications and Readers/Writers/Erasers

During the cell cycle there are many different PTMs made to the histone proteins in the nucleosome. These modifications can play many different roles in transcription and gene regulation by recruiting factors to facilitate or suppress transcription [10]. Histone PTMs can also alter the accessibility of the nucleosomes themselves. In the entry/exit region of the nucleosome, PTMs can increase the rate of unwrapping, and PTMs in the dyad region or at specific histone-DNA interaction points in the nucleosome have been shown to destabilize the nucleosome [3, 7]. Currently, the most well-known histone PTMs are lysine acetylation, methylation of lysine and arginine, phosphorylation of serine, threonine, and tyrosine, and ubiquitination (Figure 1.3).

Proteins that target histone PTMs are considered histone readers. In addition, proteins that deposit or remove histone PTMs are considered histone writers and erasers. These proteins utilize functional units in their structure referred to as epigenetic reader domains that contain motifs to recognize and bind specifically to histone methylation, acetylation, and phosphorylation marks. Some highly studied reader domains are the histone methylation reading chromodomain, Tudor, PHD(plant homeodomain) zinc finger, and PWWP(Pro-Trp-Trp-Pro) domain, and of the histone acetylation-binding reader domains, the bromodomain is well-known. Methyl-lysine

readers target methylated sites primarily through an aromatic cage motif formed by two to four aromatic residues which surround the methylated lysine side chain, and the size of the cage determines recognition of mono-, di-, or trimethylation. Histone acetylation and phosphorylation readers bind their mark through similar hydrophobic pockets. Typically a histone tail can contain multiple different PTM marks on the same tail, leading to reader proteins being affected by multiple marks. The interplay between different marks and reader proteins can lead to complicated interactions that are transcriptionally activating or repressive overall [11].

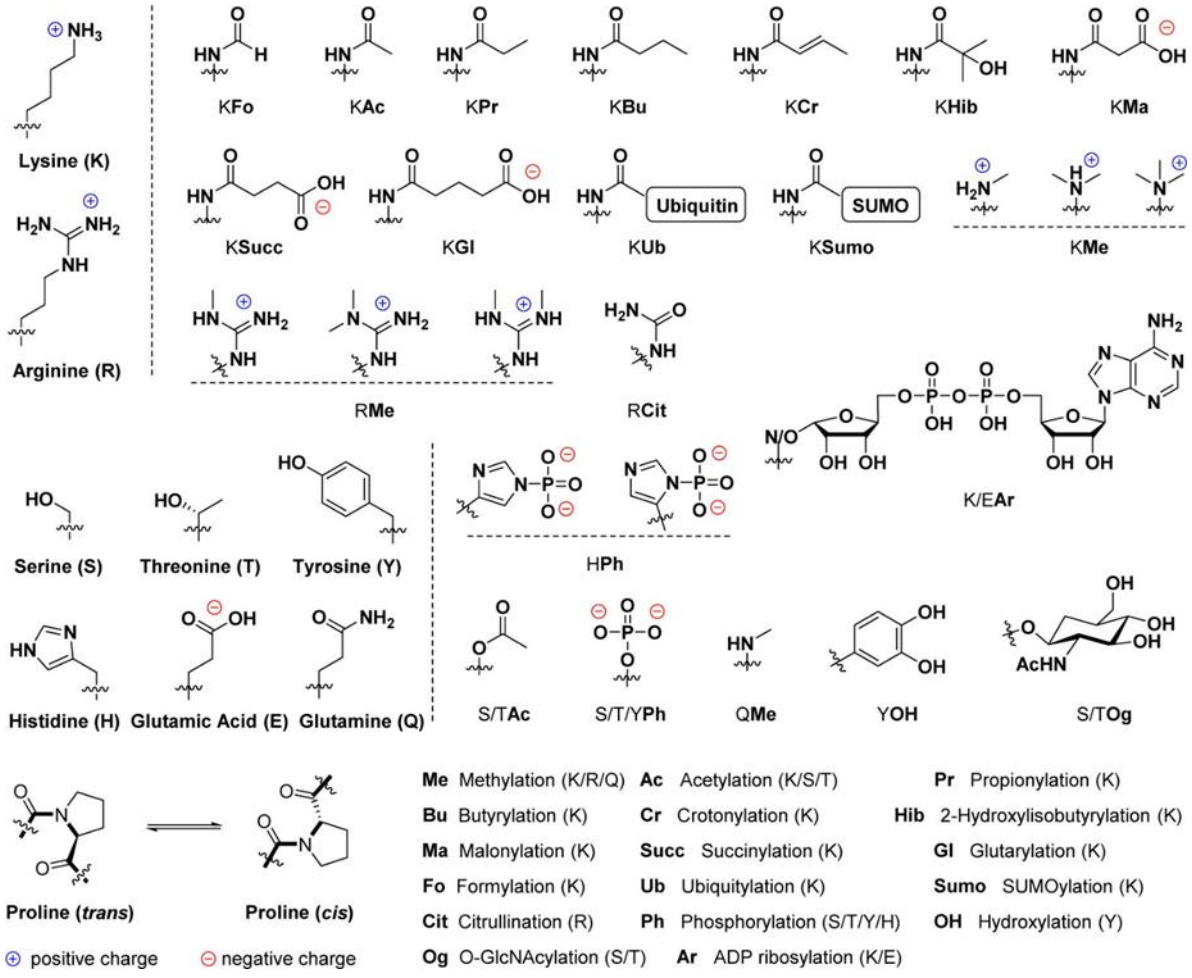


Figure 1.3: Structures of histone post-translational modifications. Reproduced from [6].

This study will indirectly focus on lysine methylation, which is a histone PTM that can add up to three methyl groups to the lysine side chain (e. g., H3K4me/me2/me3), maintaining the positive charge of the side chain, unlike lysine acetylation, but introducing extra bulk and hydrophobicity. Specifically, proteins related to H3K4/9/36me3 will be studied.

Methylation of histones was first discovered in radiolabeling studies using cell extracts [41, 42]. H3K4 methylation in particular was first discovered in trout testes through H3 peptide Edman degradation and carbon-14 radiolabeling [21]. Since then, the H3K4me3 epigenetic mark has been highly associated with actively transcribing gene regions in a variety of eukaryotic species as promoter and enhancer regions of those genes are highly enriched with H3K4me3 [20]. There is a strong positive correlation between H3K4me3, transcription rates, active polymerase II occupancy, and histone acetylation. The mark appears linked to these biological processes by recruiting a wide array of downstream effector proteins that perform their own functions on chromatin [20]. During transcription, promoter-proximal H3K4me3 can interact with the PHD finger domain of transcription initiation factor TAF3 which recruits TFIID and facilitates the assembly of the pre-initiation complex. The ATP-dependent chromatin-remodeling factor CHD1, an important factor in transcription elongation, interacts with H3K4me3 via chromodomains, and the H3K4me3-CHD1 interaction recruits factors required for mRNA maturation [23]. To add to the complexity of the roles H3K4me3 plays in the cell, while considered a mark of active chromatin, the H3K4me3 modified tail serves as a binding site for recruitment of histone deacetylase complexes (HDAC) as well. PHD fingers are a newly studied H3K4me3 binding protein domain that can be found in well-established HDAC complexes found in yeast and humans. Upon DNA damage ING2 recruits the Sin3 HDAC complex to silence transcription of cell proliferation genes, the first instance linking H3K4me3 with active gene repression [20].

At least ten known or predicted H3K4 methyltransferases exist in mammals. The Set1 complex, otherwise known as COMPASS (complex of proteins associated with Set1) is the most well-known methyltransferase, being the first H3K4 methyltransferase to be identified and is the sole enzyme responsible for H3K4 methylation in yeast [20,23]. Set1 is recruited to the 5'-regions of genes by phosphorylated serine 5 in the RNA polymerase II C-terminal domain. It then remains at the promoter regions and produces high levels of H3K4me3 [23].

H3K4me3 is still widely considered to be the 'memory' of recent transcriptional activity of the genes. Many researchers have attempted to show the direct correlation between H3K4me3 and active transcription, but it is still unclear whether the mark triggers active transcription or remains as a mark of transcriptional activity, given that transcription still occurs in the absence of the mark [23].

H3K9me3 is an epigenetic hallmark of transcriptionally inactive chromatin or heterochromatin. In pombe yeast, Clr4 is able to read and write H3K9 methylation, leading to spreading of heterochromatin via a feed-forward mechanism, where binding of Clr4, in addition to HP1 (Heterochromatin protein 1), to existing H3K9 methylation facilitates further propagation of the mark after a critical density of the modification is reached in a genomic region. HP1 dimers can bind to H3K9me3 and bridge nucleosomes while acting as a platform for recruitment of transcriptionally deactivating factors like histone deacetylases, generating the assembly of heterochromatin. Deacetylation of histones maintains high levels of H3K9me3 by preventing destabilizing factors, such as chromatin remodelers, from binding acetylated histones and acting on the chromatin. Absence of histone acetylation also facilitates ubiquitylation of H3K14, another modification that stimulates Clr4 activity [32]. H3K9me3 maintenance has been shown to be a primary factor in epigenetic inheritance of heterochromatin regions in daughter cells, and seems to play a role in stabilizing differentiated cell type identity by acting as a barrier to reprogramming of differentiated cells into stem cells by blocking binding of pluripotency transcription factors like OCT4 and SOX2 to chromatin DNA.

In mammals, Clr4 is replaced with several different histone methyltransferases (HMTs) that perform the function of depositing H3K9 methylation in specific genomic regions to propagate facultative heterochromatin [33]. Mammalian Suv39h was also found to have H3K9me3 read-write activities similar to Clr4 [32]. Typically, proteins that target H3K9me3 contain one or more chromodomains.

H3K36 methylation is another mark commonly found in active chromatin with gradients of H3K36 methylation from mono to trimethylation having been observed starting from the promoter region to the 3' end of the gene in actively transcribing gene regions. H3K36 methylation appears to have a variety of functions, such as playing a role in DNA repair and

recombination, splicing, and transcriptional regulation depending on the genomic region it is located in, when in the cell cycle H3K36 is methylated, and the particular reader proteins interacting with it [40]. To name some examples, upon DNA double-strand breakage from stress agents such as radiation, H3K36me3 is able to recruit LEDGF (Lens-Epithelium Derived Growth Factor) via a PWWP domain found in the protein. LEDGF then anchors CtIP (C-terminal binding protein Interacting Protein), facilitating DNA end resection and repair in S and G2 phase cells. To facilitate alternative splicing, MRG15 (MORF-related gene on chromosome 15) targets H3K36me3 with a chromodomain and recruits splicing regulator PTB1 (Polypyrimidine Tract Binding protein) [43]. H3K36me3's role in transcription regulation seems to be maintaining gene homeostasis and proper function. It can act in a repressive manner by recruiting histone deacetylase complexes like Rpd3S to remove acetylation from histones, leading to recompacted chromatin in the wake of RNA Polymerase II transcription, while also anchoring the Mst2 H3K14 acetyltransferase complex to active genes, preventing mistargeting of its acetyltransferase activity [44].

In cerevisiae yeast, the Set2 (SET domain containing 2) histone methyltransferase has the job of methylating H3K36 with mono-, di-, or trimethylation using its catalytic SET domain. In mammals there are at least eight histone methyltransferases that focus on different methylation states of H3K36. Some act only as monomethylases and some as dimethylases. Some H3K36 methylases have read-write functionality as well, targeting H3K36 methylation while depositing the mark, and a few can do so while also reading other modifications. For example, NSD1 (nuclear receptor SET domain-containing 1) is capable of mono and dimethylating H3K36 while recognizing H3K36me3 with a PWWP domain alongside recognizing H3K4 methylation with its PHD domains. The SETD2 trimethylase in mammals also has a C-terminal domain that interacts with a subunit of RNA polymerase II, linking H3K36me3 to transcription elongation [40].

1.3 Chapter Outlines

There are a large number of proteins capable of reading, writing, and erasing histone PTMs. The rest of this thesis will be primarily concerned only with the reader functionality of certain proteins, studying specifically the binding interactions between them and chromatin, and often probing how this interaction affects the accessibility of DNA.

- In the next chapter I will go over the methods used in this thesis to study binding interactions between mononucleosomes and transcription factors or other chromatin-binding proteins in vitro.
- Chapter 3 describes a study done with LEDGF examining its effects on HIV-1 viral DNA integration into free DNA or nucleosome substrates mediated by either the HIV intasome complex or the full pre integration complex taken from infected SUPT1 cells.
- Chapter 4 describes work on characterizing the PTM reader functionality of the CHD3 PHD domains and their impact on nucleosome accessibility.
- Chapter 5 describes a smaller study focused on the binding interaction of the BRPF1 plant homeodomain-zinc knuckle-plant homeodomain ($BRPF1_{PZP}$) region with mononucleosomes.
- Chapter 6 describes a collaborative effort to learn more about the possible cellular functions of MORC4, a poorly understood member of the MORC family of ATPase-containing chromatin binding proteins.
- Chapter 7 will discuss the results presented in this thesis and potential future directions to take.
- The appendices include protocols for methods I used repeatedly and uniquely in the Poirier Lab, namely anisotropy titrations and creation of H3 trimethylated octamers

Chapter 2. Experimental Methods

2.1 Fluorescence

Once a molecule is excited to a higher energy level by absorbing a photon, it will eventually relax down to the ground state via various radiative and non-radiative processes (Figure 2.1). The radiative process whereby light is emitted is known as fluorescence, and molecules capable of emitting light in this way are called fluorophores. The timeframe between photon absorption and emission is known as the fluorescence lifetime, and is usually on the scale of nanoseconds. Due to fast vibrational relaxation, the emitted photon will have less energy than the absorbed photon, meaning its wavelength will be longer. In addition, after the molecule is excited, it may relax down to its ground state via multiple non-radiative pathways such as thermal dissipation or internal conversion between excited singlet states [12]. As such, the amount of photons emitted is virtually never equal to the amount absorbed. For a fluorophore, the quantum yield (Φ) is defined as the ratio of number of photons emitted over the number absorbed, and is unique to each fluorophore. Typically, we choose fluorophores with a short excitation lifetime and relatively high quantum yield to minimize the effects of photobleaching.

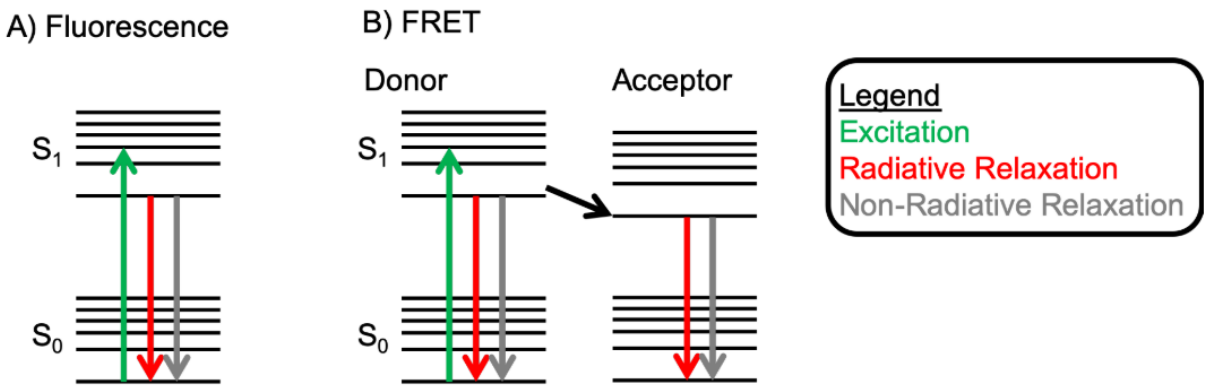


Figure 2.1: Jablonski diagrams. A) diagram of energy states present in a fluorescing molecule. Ground state molecule S_0 is excited to S_1 and relaxes via internal conversion, non-radiative transitions, and fluorescence. B) Additional pathways to relaxation are present when an acceptor fluorophore is introduced. Specifically, energy can be non-radiatively transmitted from the donor excited state to the acceptor excited state. From here, the acceptor molecule may fluoresce. Reproduced from [14]

2.2 Ensemble FRET

In this study we measure transcription factors binding to nucleosomes via FRET or Fluorescence Resonance Energy Transfer by labeling nucleosomes with the Cy3-Cy5 FRET pair. FRET is a non-radiative energy transfer process between two fluorophores that is made possible when the emission spectrum of one of the fluorophores (the donor) overlaps with the excitation spectrum of the other fluorophore (the acceptor), and the two fluorophores are colocalized together typically within a couple nanometers [12]. Then when the donor fluorophore undergoes direct excitation, emission is observed from the acceptor fluorophore instead, with the efficiency of this energy transfer being given as a function of the distance between the two fluorophores by:

$$E = \frac{1}{1 + \left({}^{R}/_{R_0}\right)^6} \tag{2.1}$$

where R_0 is the distance between the two fluorophores at which FRET efficiency is 50%. R_0 is also unique to each donor-acceptor pair, and usually only a few nanometers. The efficiency of FRET is extremely distance dependent due to the dipole-dipole interaction nature of the process, which, combined with the length scale of R_0, makes FRET a very effective method for probing biological systems on the scale of the nucleosome, which has a diameter of ~11 nm. We label the octamer with Cy5 usually at H2AK119C and then label the nucleosomal DNA nearby in the entry region with Cy3 such that the nucleosome will exhibit high FRET (E =70-100%) while the nucleosome is fully wrapped, and low FRET (E=10-40%) while unwrapped (Figure 2.2). Though the nucleosome undergoes spontaneous unwrapping-rewrapping due to thermodynamic fluctuations, such

nucleosome dynamics occur too quickly (~milliseconds) to measure the resulting changes in FRET due to them. We design a transcription factor binding site into the DNA in the entry region so that when the DNA undergoes a transient unwrapping event, the transcription factor can bind its site in the nucleosome and trap it in a low FRET state long enough for us to measure the resulting low FRET of the nucleosome, or in other words, measure binding of a transcription factor to a nucleosome via FRET. This method allows us to probe the accessibility of nucleosomes under many different conditions.

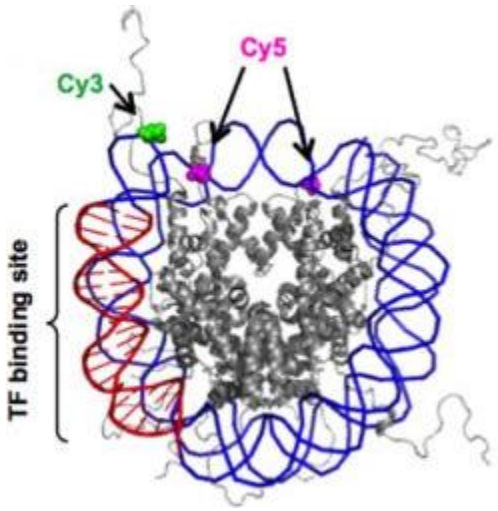


Figure 2.2: Structure of the Cy3-Cy5 labeled nucleosome (PDB: 1KX5), highlighting a TF binding site (red), the Cy3 fluorophore (green) and the Cy5 fluorophore (magenta). Reproduced from [13]

There are a few different ways to measure FRET efficiency. Most of them require a sample with and without the acceptor fluorophore, which doubles the number of samples needed to be made and measured. Instead, we use the $(\text{Ratio})_A$ method [22]:

$$E = \frac{N\varepsilon \frac{F_A{}'}{F_A} - \varepsilon''}{\varepsilon' d^+} \tag{2.2}$$

ε	extinction coefficient of acceptor at acceptor excitation wavelength
ε'	extinction coefficient of donor at donor excitation wavelength
ε''	extinction coefficient of acceptor at donor excitation wavelength
F_A	fluorescence of acceptor when directly exciting acceptor
$F_A{}'$	fluorescence of acceptor when exciting donor (FRET)
d^+	labeling efficiency of donor
N	number of acceptors that can undergo FRET

Table 2.1: factors of the $(Ratio)_A$ method, the different extinction coefficient factors account for the difference in energy absorption of the fluorophores. $F_A{}'$ = fluorescence of the acceptor when exciting the donor after subtracting any contribution from the donor to the acceptor fluorescence intensity peak. N is typically 2 for a nucleosome labeled with a Cy5 acceptor on both H2As.

To perform a $(Ratio)_A$ measurement of sample FRET, the donor is excited in the presence of the acceptor, leading to observed emission from both the donor and acceptor in the fluorescence spectrum. Then the emission spectrum obtained from exciting the donor without the presence of the acceptor is subtracted from the first spectrum in order to remove donor emission bleed-through to the acceptor signal (this spectrum only needs to be measured once and then can be used to correct each FRET spectrum measurement as long as it's between the same donor-acceptor pair). After that, the spectrum of acceptor emission via acceptor excitation is obtained and compared to acceptor emission via donor excitation to acquire a value for the sample's FRET with $(Ratio)_A$.

2.3 Single Molecule TIRF

2.3.1 Background/Motivation

In the previous section we discussed a way to measure the binding of transcription factors to nucleosomes utilizing FRET via the $(Ratio)_A$ method, which partially integrates the emission spectrum of a sample population of typically Cy3-Cy5 labeled nucleosomes plus some type of transcription factor. This measurement of an ensemble of biomolecules only describes the equilibrium state of the system. It does not provide a detailed view of the kinetics of the system, the different states that the system can enter, and the transition rates between them. To do this, we use prism-based TIRF (Total Internal Reflection Fluorescence) microscopy (Figure 2.3). TIRF on a slide surface creates an evanescent wave that penetrates only roughly 100 nm past the surface, exciting individual fluorophore labeled biomolecules on the surface without causing a significant amount of background fluorescence [25]. This allows us to monitor the FRET state of single molecules over time to detect state transitions and measure transition rates, discover hidden states, show order of events in a binding reaction, etc.

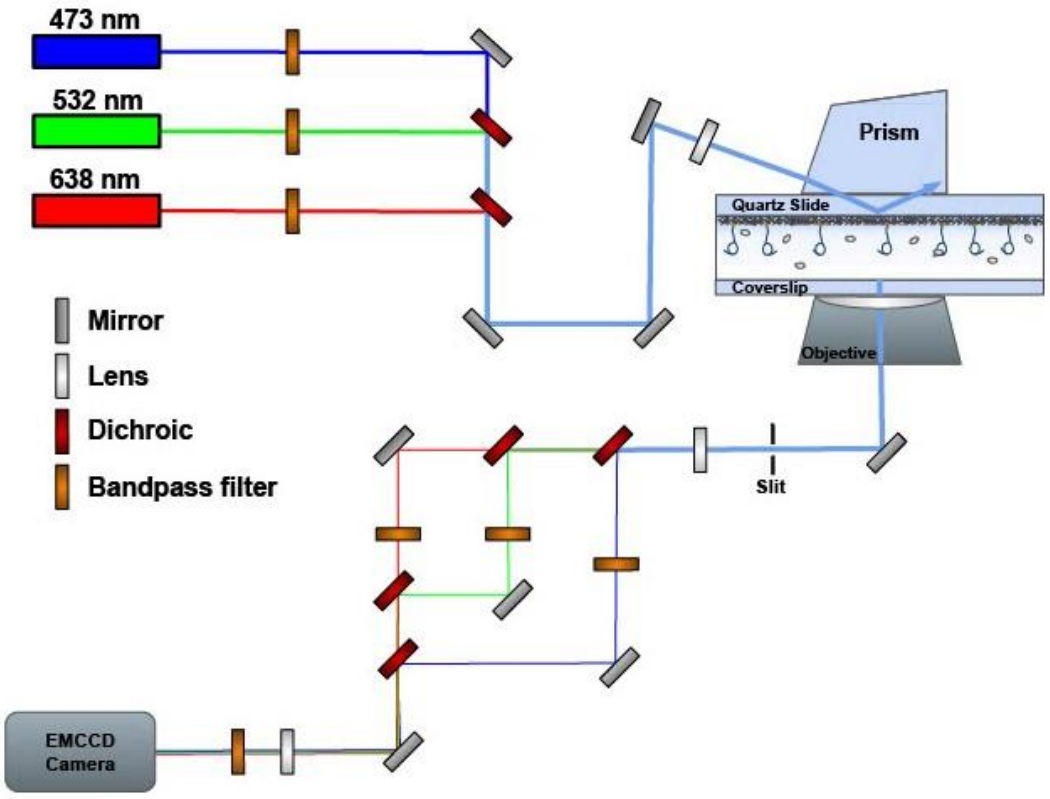


Figure 2.3: Diagram of microscope setup for three-color single-molecule measurements.

2.3.2 Preparing Functionalized Microscope Slide Surfaces

Pictured in Figure 2.3, to perform single-molecule experiments, we passivate and functionalize quartz slide surfaces by coating them with PEG (Polyethylene glycol) to prevent proteins and DNA from sticking to charged groups on the surface [45]. A tiny fraction of the PEG molecules also have a biotin conjugated to their structure in order for us to take advantage of the strong biotin-neutravidin interaction to tether nucleosomes to the passivated slide surface. The microscope slides are created by sandwiching the quartz slides and PEG functionalized glass coverslips together with parafilm in between (coverslip PEG surface not pictured). Sample chambers are cut out of the parafilm beforehand to create flow cells for sample insertion. Before nucleosomes are added to the slide flow cell, 1 mg/ml BSA is incubated in the flow cell for five minutes to further cover charged groups, followed by 20 μg/mL neutravidin for 5 minutes. After each incubation,

excess BSA/neutravidin is fully removed from the cell with wash buffer. Once neutravidin has been put down, nucleosomes at a very low concentration of picomolar are added to the flow cell and allowed to incubate for five minutes before being washed out as well. Finally, imaging buffer is added just before the single-molecule experiment. If the experiment is meant to measure a transcription factor's binding kinetics to a nucleosome, the transcription factor is then included in the imaging buffer.

2.3.3 Imaging Buffer

During a single-molecule experiment where fluorophores are in use, more needs to be taken into consideration than just the pH and concentration of ions. Emission of light from a fluorophore occurs from electron singlet-singlet state transitions, but due to interactions with molecules in the buffer, conversion to a relatively long-lasting triplet state can occur which results in fluorophore blinking. To reduce photobleaching and blinking of the fluorophore emission, we use a mix of reagents known to quench the triplet state in single-molecule FRET studies: Trolox, Cyclooctatetraene, and 3-nitrobenzyl alcohol [26]. Oxygen in the buffer can also cause problems. When excited into a singlet state, it can react with and damage biomolecules in the buffer, including the fluorophores, which causes photobleaching. To prevent this, we utilize an oxygen scavenging system involving glucose oxidase which converts glucose and oxygen into D-gluconolactone and hydrogen peroxide. Added catalase then converts the hydrogen peroxide into water [27].

2.3.4 Microscope Setup

We use three lasers, 473 nm, 532 nm, and 638 nm, that are first directed through bandpass filters (473/10, 531/22, 640/22) and lined up together along the same path through the use of dichroic mirrors to contact the prism at a single point. A plano-convex lens is used to focus the light before

it contacts the prism, which is a Pellin-Broca prism with the same index of refraction as the quartz slide (n=1.55). Emission from samples on the slide surface excited by the resulting evanescent field is then collected through a 60x silicone objective lens and directed through a custom built emission path meant to separate the emission from Cy3 and Cy5 signals. The light is first focused through a slit meant to set the width of the image downstream. The light is collimated then separated into different wavelengths again by dichroic mirrors corresponding to 473 nm, 532 nm, or 638 nm emission, before hitting an EMCCD camera chip. The beams are aligned so that the different emission paths hit the chip side by side instead of at the same point on the chip, so that emission from Cy3 and Cy5, for instance, can be viewed simultaneously.

In prism based TIRF, the evanescent field is created by directing the excitation path through a prism above the quartz microscope slide. Emission is then collected through the objective below the slide. TIR can be explained by using Snell's law for light entering or exiting different mediums:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_t}{n_i} \tag{2.3}$$

θ_i is the angle of incidence with respect to the perpendicular of the surface upon which the light is entering, θ_t is the transmitted angle with respect to the perpendicular, and n_t / n_i are the indexes of refraction for the medium the light is entering and exiting respectively at the interface. To achieve TIR, from Snell's law we can see that light transitioning to a lower index of refraction medium bends the light away from the surface perpendicular, and so if we increase the angle of incidence at the interface leading into the flow cell, at some point the incident light will be at or greater than the critical angle at which the beam is reflected propagating parallel with the slide surface (Figure 2.4). In our case, TIR occurs at the interface of quartz (n_i = 1.55) and essentially water (n_t = 1.3). From Snell's law, taking $\theta_t = \pi/2$, the critical angle required to produce TIR then is given by $\theta_c = \sin^{-1}(n_i / n_t) \approx 60^\circ$.

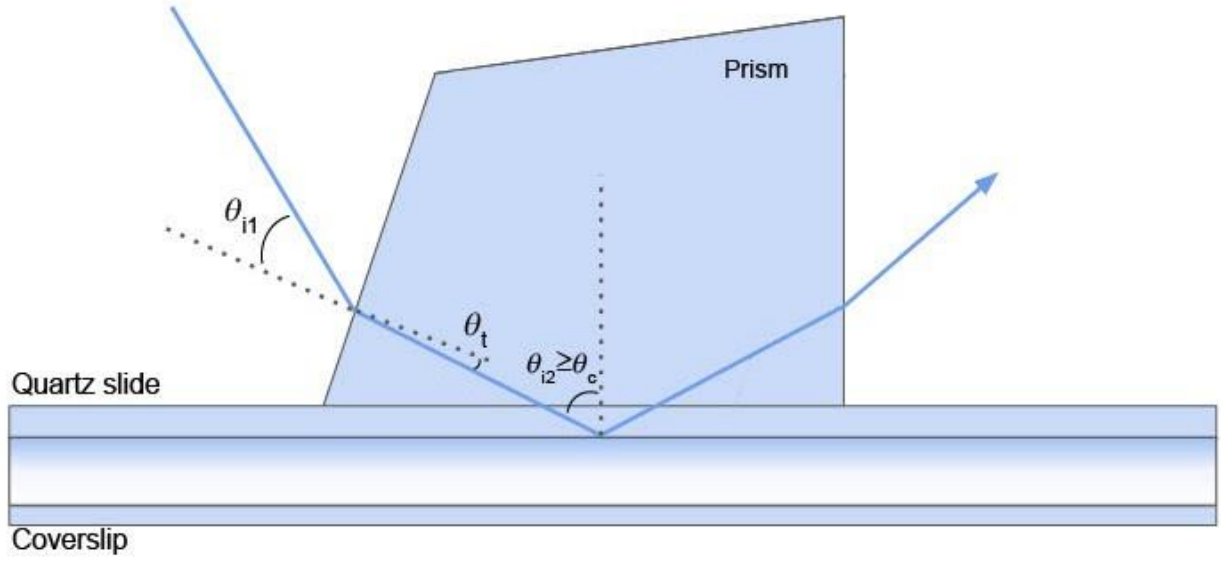


Figure 2.4: Diagram of prism orientation and incident angles leading to total internal reflection at the quartz slide-flow cell interface.

2.3.5 Data Acquisition/Analysis

The samples in this work use only Cy3 and Cy5 for imaging single molecules on the quartz slide surface. Once the slide is in place above the objective and TIRF evanescent field excitation has been set up, the 638 nm laser is engaged at a low power for ~10 frames in order to excite Cy5 and locate nucleosomes on the slide surface. Cy5 excitation is used instead of Cy3 since typically we label H2AK119C with Cy5, which is located on the dimer of the nucleosome. Since loss of the dimers is the first step in nucleosome disassembly, the presence of Cy5 suggests a fully formed nucleosome. Cy3 excitation also tends to excite much more background fluorescence on the slide surface compared to Cy5. Then after the nucleosomes are located, the 532 nm laser is engaged to begin Cy3 excitation. Movies of the slide surface are taken as stacked TIF images at 0.5-20 frames per second for five minutes up to an hour, whatever is most appropriate for the timescale of the binding being observed.

ImageJ in combination with Matlab is then used to process the movies. First the movies are binned by a factor of 2 in ImageJ to increase signal-to-noise ratio in the movies, then

background subtraction using a rolling ball radius of 50 is performed on the movies. The first ~10 frames of a movie showing the locations of the nucleosomes are combined and averaged to form a single TIF image which we ultimately use to pinpoint the nucleosome positions on the slide surface by looking for the intensity maxima in ImageJ. We then use an in-house coded Matlab program to average a 3x3 pixel square around those maxima at each frame throughout the movie to generate Cy3-Cy5 emission time traces of individual nucleosomes associated with those maxima on the slide surface.

At this point we sort through the raw time traces looking for nucleosomes exhibiting consistent anti-correlated Cy3-Cy5 FRET fluctuations that could be attributed to binding of a transcription factor to the nucleosome, which traps it in a low FRET state, observed as a drop in Cy5 emission and a rise in Cy3 emission for a period of time in the raw time traces. We do not worry about the changes in FRET associated with the nucleosome unwrapping/rewrapping due to thermodynamic fluctuations since these occur on timescales faster than our microscope can measure. First, we determine how many of the time traces come from a nucleosome undergoing FRET instead of an artifact on the slide surface, the clearest way to do this is to only include traces that show a clear FRET photobleaching event where the Cy5 signal suddenly drops followed by an immediate rise in the Cy3 signal during the time trace. Of the traces that show nucleosomes undergoing FRET, the traces are discarded according to a couple more criteria: there are no fluctuations, the fluctuations are correlated or uncorrelated, there is blinking occurring during the trace, or photobleaching occurred too quickly to measure enough binding events.

Once we have finished sorting through the traces for ones viable for analysis, we truncate the traces. Cutting off the sections showing photobleaching and the initial Cy5 excitation before utilizing the vbFRET Matlab program developed by the Ruben Gonzalez lab to gather information on kinetics [46]. The raw time trace is fit to a Hidden Markov model of 2 states, a high FRET unbound state and low FRET bound state, to generate an idealized FRET time trace. Then from these idealized time traces we can determine binding and dissociation rates by building histograms that fit to an exponential distribution.

For our single molecule data we have two states, high FRET and low FRET or unbound [U] and bound [B] states. Molecules transition between the two states with rate k_{UB} to the bound state and rate k_{BU} back to the unbound state. This system follows the set of differential equations:

$$\frac{d[B]}{dt} = k_{UB}[U] - k_{BU}[B] \tag{2.4}$$

$$\frac{d[U]}{dt} = -k_{UB}[U] + k_{BU}[B] \tag{2.5}$$

We can choose to focus on either the bound or unbound state in solving the differential equations. Choosing the bound state and solving under the condition that there are only transitions out of the bound state, we get that the probability distribution of a molecule to be in the bound state for a period of time is a decaying exponential proportional to the transition rate out of the state $P_{bound} = e^{-k_{BU}t}$. We set all the bound times we measure to start at the same point in time to generate histograms that we then fit to a decaying exponential. The bound and unbound times are then determined from the characteristic rate of decay, followed by the rates of binding and dissociation. Movies are taken typically for four different concentrations of the transcription factor in part to show that the binding rate is concentration dependent while the dissociation rate remains constant, which shows that the measured rates follow first order binding kinetics.

2.4 Anisotropy

When a fluorophore is excited by polarized light, the resulting fluorescence emission will be more or less polarized depending on the relation between the excitation lifetime of the fluorophore and its rotational mobility (speed and rotational freedom). In other words, if the excitation lifetime of a fluorophore is much faster than the tumbling speed, then it is most probable that by the time the fluorophore re-emits light after excitation, the polarized emitted light will be oriented in a

different direction, leading to the fluorescence of the fluorophore becoming isotropic due to the effects of rotational diffusion. However if the opposite is true, the tumbling time becomes much slower than the excitation lifetime. Then the emitted fluorescence will remain anisotropic. This fluorescence technique has been used to probe interactions between biomolecules by having a ligand molecule bind to a (usually of the same size or smaller) fluorophore-tagged substrate molecule in order to observe a resulting rise in the anisotropy of the emitted fluorescence caused by the shift to the resulting complex's lower rotational diffusion [16-18]. As such, fluorescein is usually a good choice as a biomolecule labeling fluorophore since its long excitation lifetime helps ensure the baseline emitted fluorescence is isotropic. The nucleosome presents difficulties with this technique due to its relatively large size of 200 kDa. Most protein-nucleosome interactions considered in this study look at proteins much smaller in size than the nucleosome, such that normally when one of these proteins would bind the fluorescein labeled nucleosome, the resulting shift in anisotropy would be small since the tumbling time of the nucleosome itself is not significantly affected. However, anisotropy techniques have been performed successfully with nucleosomes by labeling the nucleosomal DNA with fluorescein in the DNA entry region which initially allows the fluorophore a high freedom of movement. Then once a chromatin-binding protein capable of clamping down on the DNA restricts the movement of the attached fluorophore, a shift in anisotropy is observed, allowing one to detect binding of a protein to the nucleosome. As expected, this strategy does not seem to be effective for chromatin binding proteins that do not interact with the nucleosomal DNA, but only bind the histone tails or interact with acidic patches on the octamer.

To take an anisotropy measurement, two polarizers are required. One to polarize the excitation light and another to measure the amount of polarization of the emitted fluorescence via different axis [19]. The anisotropy of a sample is defined as:

$$\langle r \rangle = \frac{I_{VV} - G I_{VH}}{I_{VV} + 2G I_{VH}} \tag{2.6}$$

$$G = \frac{I_{HV}}{I_{HH}} \tag{2.7}$$

Where the H/V subscripts indicate whether one of the two polarizers are in a horizontal or vertical position with regards to a common axis. G is called the 'G factor' and corrects for potential differences in the detection efficiency of the different light polarizers. The polarization of a sample in terms of the sample's anisotropy is defined as:

$$P = \frac{3\langle r \rangle}{2 + \langle r \rangle} \tag{2.8}$$

Polarization or anisotropy can be used. Typically polarization is used and reported as mP (polarization x 1000), but using anisotropy instead leads to simpler relations.

2.5 Binding of a Protein Ligand to the Nucleosome

2.5.1 The Hill Equation

In this thesis, the dynamics of mononucleosomes are studied in the context of protein transcription factors accessing the nucleosomal DNA. The initial roughly 20bp of DNA in the entry-exit region of the nucleosome are not tightly bound to the octamer and can become transiently unwrapped due to thermodynamic fluctuations, granting access to the nucleosomal DNA and whatever transcription factor binding sites may lie there [34]. The dynamics of nucleosome unwrapping/rewrapping trend very much towards the wrapped state, with the rate of unwrapping being around $4s^{-1}$ and the rate of rewrapping being around $100s^{-1}$ [35]. These rates are currently too fast for us to measure, and so we rely on transcription factor binding to the transiently unwrapped DNA, which traps the nucleosome in a partially unwrapped state via steric blocking. Typically the nucleosome is then occupying a partially unwrapped state long enough for us to detect, allowing us to ultimately probe nucleosome dynamics in the presence of other factors.

Before we study the binding of transcription factors and other proteins to nucleosomes, we need a model to functionally describe the system. For a two state system in dynamic equilibrium describing a single ligand binding to a substrate with one binding site we have:

$$[\boldsymbol{P}] + [\boldsymbol{S}] \leftrightharpoons [\boldsymbol{PS}] \tag{2.9}$$

[P] and *[S]* are the concentrations of the binding protein and substrate respectively, and *[PS]* is the concentration of the bound complex. The binding affinity of the protein to the substrate can be quantified in the form of the dissociation constant K_D defined as:

$$\boldsymbol{K_D} = \frac{[\boldsymbol{P}][\boldsymbol{S}]}{[\boldsymbol{PS}]} \tag{2.10}$$

Experimentally we determine the K_D from the fraction of bound complexes compared to the total amount of free protein and substrate:

$$\emptyset = \frac{[\boldsymbol{PS}]}{[\boldsymbol{P}] + [\boldsymbol{S}]} \tag{2.11}$$

From (3.2) we can rewrite (3.3) as:

$$\emptyset = \frac{\mathbf{1}}{\mathbf{1} + (\boldsymbol{K_D}/[\boldsymbol{P}])} \tag{2.12}$$

For n multiples of the same binding protein (if the substrate has multiple binding sites or non-specific binding occurs), (3.4) becomes:

$$\emptyset = \frac{\mathbf{1}}{\mathbf{1} + (\boldsymbol{K_D}/[\boldsymbol{P}])^{\boldsymbol{n}}} \tag{2.13}$$

Binding curves developed from concentration titrations of the protein in the presence of a constant concentration of the substrate can be fit to (3.5) to determine the K_D of the reaction. The factor n is called the Hill coefficient and is a gauge of the cooperativity of the protein binding to its substrate. For $n > 1$, binding of the first protein (and successive proteins) assists binding of further proteins to the same substrate, naturally $n < 1$ indicates the opposite effect. Binding of each proteins impedes binding of further proteins. At $n = 1$ the binding reaction is uncooperative [36].

To stay consistent in characterizing the binding affinity of proteins to their substrates throughout this thesis. The $S_{1/2}$ is found for each binding curve as a way to quantify the binding affinity of the protein to the substrate, which is the concentration of binding protein at which half of the substrates are bound. For (3.5), this is when $[P] = K_D$, which can only be said when it is reasonable to make the approximation that the total amount of protein is roughly equal to the

amount of free unbound protein. An approximation only possible if the concentration of substrate in the reaction is much less than the reaction's K_D. To illustrate, if $[S]_{total} \ll K_D$ then since $[S]_{total} = [S]+[PS]$, $[S]$ or $[PS]$ is much less than the K_D. Meaning that if $K_D = [P][S]/[PS]$ and $[S] \ll K_D$ then $[PS] \ll [P]$ and $[P]_{total} = [P] + [PS] \approx [P]$.

For the ensemble FRET and anisotropy titrations performed later in this thesis, the fraction of nucleosome substrates bound with protein can be written as:

$$\emptyset = \frac{E - E_1}{E_2 - E_1} \tag{2.14}$$

E is the current FRET efficiency or fluorescence anisotropy of the nucleosome substrates for each binding protein concentration in a titration. E_1 is the FRET efficiency/anisotropy with no protein added, and E_2 is the FRET efficiency/anisotropy at protein binding saturation when all of the nucleosomes are bound. Setting (3.6) equal to (3.5) and solving for E gives us:

$$E = E_1 + \frac{E_2 - E_1}{1 + \left(S_{1/2}/[P]\right)^n} \tag{2.15}$$

Which is an equation that can be fit to ensemble FRET and anisotropy binding curves seen later in this thesis in order to experimentally determine the $S_{1/2}$ of a protein binding to nucleosome substrates using E_1, E_2, $S_{1/2}$, and n as the fitting parameters.

2.5.2 Nucleosome Dynamics

To describe the binding of a transcription factor protein to a site in the entry-exit region of the nucleosome, which undergoes transient unwrapping, exposing the nucleosomal DNA. We turn to a three-state model, the "site exposure model," developed previously in nucleosome restriction enzyme digestion and transcription factor binding studies [37,38]. Wherein binding of the transcription factor becomes less probable the further into the nucleosome the factor's binding site is located.

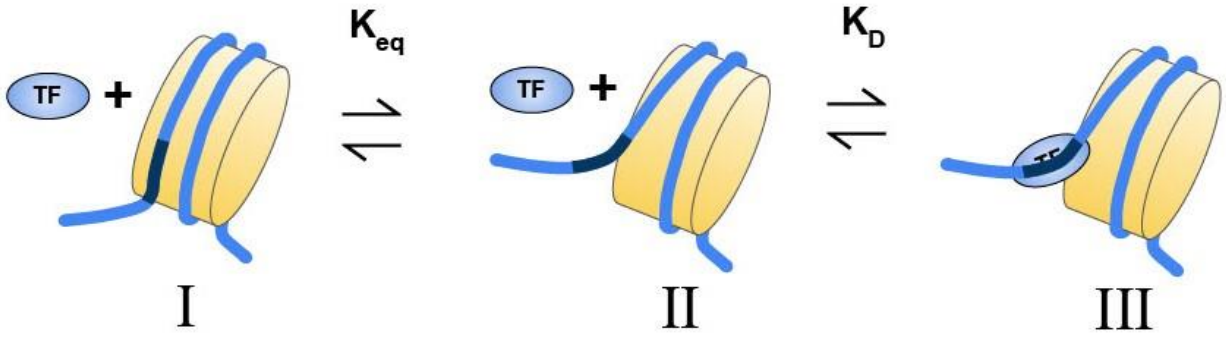


Figure 2.5: Three state site exposure model of transcription factors binding to a site in the nucleosomal DNA. State I represents the fully wrapped, closed nucleosome. The nucleosome becomes partially unwrapped in state II, leading to the transcription factor binding to its site in state III

In this model, the nucleosome transitions back and forth between the unwrapped/wrapped states with equilibrium constant $K_{eq} = [II]/[I]$, and once the nucleosome becomes partially unwrapped, the transcription factor protein binds and disassociates from the nucleosome with dissociation constant $K_D = [P][II]/[III]$.

To describe the impact of site exposure on transcription factor binding. We can start by writing the fraction of nucleosomes in the protein bound state in terms of the concentrations of each state:

$$\emptyset = \frac{[\boldsymbol{III}]}{[\boldsymbol{I}] + [\boldsymbol{II}] + [\boldsymbol{III}]}$$

$$\emptyset = \frac{K_{eq}[I][P]/K_D}{[I] + K_{eq}[I] + K_{eq}[I][P]/K_D}$$

$$\emptyset = \frac{K_{eq}[P]}{K_D + K_D K_{eq} + K_{eq}[P]} \tag{2.16}$$

If we then want to find the concentration of protein at which half of the nucleosomes are bound, we set $[P] = S_{1/2}$ and $\emptyset = ½$ and solve for $S_{1/2}$:

$$\boldsymbol{K_D + K_D K_{eq} + K_{eq} S_{1/2}} = \boldsymbol{2K_{eq} S_{1/2}}$$

$$S_{1/2} = \frac{K_D}{K_{eq}}\left(1 + K_{eq}\right)$$

$$S_{1/2} \approx \frac{K_D}{K_{eq}} \text{ for } K_{eq} \ll 1 \quad (2.17)$$

Transcription factor binding affinity to its site in the nucleosomal DNA is reduced by a factor of K_{eq}, which is the probability of site exposure. Previous in vitro studies have shown this amounts to a reduction in binding affinity of around 2-3 orders of magnitude to the nucleosomal entry-exit region DNA compared to naked DNA [39].

Chapter 3. HIV-1 Preintegration Complex Preferentially Integrates the Viral DNA into Nucleosomes Containing Trimethylated Histone 3-Lysine 36 Modification and Flanking Linker DNA

Reproduced (with additions) from: Sapp, N., Burge, N., Cox, K., Prakash, P., Balasubramaniam, M., Thapa, S., … Dash, C. (2022). HIV-1 Preintegration Complex Preferentially Integrates the Viral DNA into Nucleosomes Containing Trimethylated Histone 3-Lysine 36 Modification and Flanking Linker DNA. *Journal of Virology*, *96*(18), 1–30.

This chapter covers a study that examined LEDGF's importance to HIV-1 viral DNA integration into chromatin in the context of the pre-integration complex collected from cell extracts, and in the context of the HIV-1 intasome complex of HIV integrase bound to viral DNA. I was responsible for the Gal4+LEDGF FRET titrations and LEDGF fluorescence polarization data. In section 3.9 I discuss unpublished smFRET data taken with LEDGF.

3.1 Abstract

HIV-1 DNA is preferentially integrated into chromosomal hot spots by the preintegration complex (PIC). To understand the mechanism, we measured the DNA integration activity of PICs—extracted from infected cells—and intasomes, biochemically assembled PIC substructures using a number of relevant target substrates. We observed that PIC-mediated integration into human chromatin is preferred compared to genomic DNA. Surprisingly, nucleosomes lacking histone modifications were not preferred integration compared to the analogous naked DNA.

Nucleosomes containing the trimethylated histone 3 lysine 36 (H3K36me3), an epigenetic mark linked to active transcription, significantly stimulated integration, but the levels remained lower than the naked DNA. Notably, H3K36me3-modified nucleosomes with linker DNA optimally supported integration mediated by the PIC but not by the intasome. Interestingly, optimal intasome-mediated integration required the cellular cofactor LEDGF. Unexpectedly, LEDGF minimally affected PIC-mediated integration into naked DNA but blocked integration into nucleosomes. The block for the PIC-mediated integration was significantly relieved by H3K36me3 modification. Mapping the integration sites in the preferred substrates revealed that specific features of the nucleosome-bound DNA are preferred for integration, whereas integration into naked DNA was random. Finally, biochemical and genetic studies demonstrate that DNA condensation by the H1 protein dramatically reduces integration, providing further evidence that features inherent to the open chromatin are preferred for HIV-1 integration. Collectively, these results identify the optimal target substrate for HIV-1 integration, report a mechanistic link between H3K36me3 and integration preference, and importantly, reveal distinct mechanisms utilized by the PIC for integration compared to the intasomes.

3.2 Importance

HIV-1 infection is dependent on integration of the viral DNA into the host chromosomes. The preintegration complex (PIC) containing the viral DNA, the virally encoded integrase (IN) enzyme, and other viral/host factors carries out HIV-1 integration. HIV-1 integration is not dependent on the target DNA sequence, and yet the viral DNA is selectively inserted into specific "hot spots" of human chromosomes. A growing body of literature indicates that structural features of the human chromatin are important for integration targeting. However, the mechanisms that guide the PIC and enable insertion of the PIC-associated viral DNA into specific hot spots of the human chromosomes are not fully understood. In this study, we describe a biochemical mechanism for the preference of the HIV-1 DNA integration into open chromatin. Furthermore,

our study defines a direct role for the histone epigenetic mark H3K36me3 in HIV-1 integration preference and identify an optimal substrate for HIV-1 PIC-mediated viral DNA integration.

3.3 Introduction

The human immunodeficiency virus 1 (HIV-1) has infected approximately 80 million people worldwide, with ~38 million people currently living with the virus (UNAIDS, 2021). Highly potent antiretroviral therapy (ART) has rendered HIV-1 infection a chronic condition and has significantly reduced the burden of AIDS (UNAIDS, 2021). However, ART is not curative, faces drug resistance, and lifelong ART can cause severe toxicity and comorbid conditions [57]. Therefore, a clear understanding of the mechanism of HIV-1 replication is critical for the continued development of novel and improved therapeutic strategies.

HIV-1 replicates in immune cells expressing the CD4 receptor and the CCR5 or CXCR4 chemokine coreceptors [58]. Binding of the HIV-1 glycoprotein(s) to these receptors and coreceptors initiates the fusion of the viral membrane to the cellular plasma membrane [59]. Thereafter, the viral capsid containing two copies of the linear single-stranded viral RNA genome and a number of viral/cellular factors is released into the cytoplasm of the host cell [60–65]. The cytoplasmic release of the capsid allows the reverse transcription complex (RTC) to convert the viral RNA genome into a double-stranded viral DNA copy [69]. Subsequently, by a poorly understood mechanism, the RTC transitions into a preintegration complex (PIC). The PIC containing the viral DNA, viral integrase (IN) enzyme, and associated cellular/viral proteins [70] carries out integration of the viral DNA into the host chromatin to establish the proviral genome [71–73]. The proviral genome is required for the production of progeny virions and the establishment of viral reservoirs [74–86]. Because integration is critical for HIV- 1 infection [67, 77, 78], this step has been the target of highly potent antiviral drugs [79,80].

HIV-1 DNA integration is dependent on the activity of the PIC-associated IN enzyme [81]. First, IN carries out the 3'-processing of the viral DNA ends and then, via a strand-transfer step, the viral DNA is inserted into the host genomic DNA [82–85]. Even though HIV-1 integration is

not dependent on the target DNA sequence [86], the viral DNA is selectively inserted into specific "hot spots" of human chromosomes [87]. Notably, the protein-coding genes account for <2% of the entire human genome. However, this tiny fraction of the human genome contains over half of the reported HIV-1 integration sites [87–93]. Despite HIV-1 IN being the primary viral factor required for inserting the viral DNA into chromosomal hot spots, a growing body of literature indicates that structural features of the human chromatin are important for integration targeting [89, 93–98]. For instance, HIV-1 integration site analysis reveals selective integration patterns into the intragenic regions of the open chromatin [91, 92, 96, 99]. These preferred regions for integration are often characterized by active transcription, gene density, and epigenetic factors characteristic of open chromatin [89, 93, 96, 98–102]. In particular, deep-sequencing analysis coupled with the advances in human genome annotation have provided evidence that HIV-1 integration is favored near chromatin features associated with active transcription [89, 103–49]. Consistent with the model that open chromatin is favored for HIV-1 integration, heterochromatic regions at human centromeres and telomeres are disfavored for integration [87, 96, 98, 100]. However, the mechanisms that guide the PIC and enable insertion of the PIC-associated viral DNA into specific hot spots of the human chromosomes are not fully understood.

To insert the viral DNA into the genomic hot spots, the PIC must engage/overcome the structural barriers of the chromosomal landscape. For instance, a single copy of the human genome can extend over 2 m, yet it is packaged into a nucleus with an average diameter of 10 μm [106]. This complex and poorly understood genome compaction process is achieved by organizing the DNA into a nucleoprotein polymer called chromatin. The basic repeating unit of chromatin is a nucleosome, which contains a nucleosome core formed by an octamer of histone proteins containing two copies of H2A, H2B, H3, and H4 that wraps a 147 bp of the genomic DNA [106, 107]. The nucleosome core is connected to the adjacent nucleosome core by a segment of linker DNA. The linker DNA is often associated with the histone protein (H1) that further compacts the genome [108–110]. Finally, an array of epigenetic modifications on the histone proteins and the DNA adds more complexity to the structural and functional landscape of the human chromatin [111–113].

A direct role of chromatin structure in PIC-mediated viral DNA integration targeting has remained elusive. However, studies of purified HIV-1 IN suggest a potential functional link. For example, the DNA wrapped within nucleosomes is preferred for retroviral integration [114–118]. Particularly, HIV-1 and other retroviral INs preferentially target the outward-facing major groove of the nucleosomal DNA [116–119]. Evidently, the distortion of the nucleosomal DNA as it bends around the histone octamer is thought to facilitate integration into the nucleosomes. However, retroviral INs differ in how they integrate viral DNA within chromatinized DNA templates [94, 97, 110–114]. Although HIV and murine leukemia virus (MLV) target nucleosome cores for integration, nucleosome arrays are less preferred for avian sarcoma virus (ASV) integration compared to analogous naked DNA [95, 97, 115]. In addition, ASV IN-mediated integration is preferred into a compacted chromatin than with naked DNA or an extended nucleosome array [94]. In contrast, HIV-1 IN-mediated integration was reduced into compacted chromatin. In addition, specific epigenetic modifications that are known to alter chromatin structure have also been linked to integration targeting [89, 93]. For example, HIV-1 integration is positively associated with a group of histone posttranslational modifications that are linked to active transcription [89, 93, 125]. Conversely, HIV-1 integration is negatively associated with modifications that inhibit transcription, such as H3 K27 trimethylation and DNA CpG methylation. The effects of epigenetic modifications on HIV-1 integration contrast with other retroviruses such as MLV, which prefers to integrate into CpG islands, and ASV, which integrates within runs of alternative CpG islands [136, 137]. Similarly, cellular proteins such as LEDGF/p75, BET, and others that are recruited by viral proteins also play key roles in viral DNA integration, especially into chromatin substrates [94, 95, 97, 120, 121, 128–130]. In particular, the interaction between HIV-1 IN and LEDGF/p75 is linked to integration site selection [131–142]. Still, the molecular and biochemical details underlying the preference of retroviral DNA integration into specific regions of chromatin are not fully understood. There is currently a lack of studies of PIC-mediated viral DNA integration preference into physiologically relevant targets such as isolated chromatin and dechromatized genomic DNA (gDNA).

In this study, we describe a biochemical mechanism for the preference of the HIV-1 DNA integration into open chromatin. We used HIV-1 PICs extracted from acutely infected T-cell lines

and measured the ability of these viral replication complexes to integrate the viral DNA into isolated chromatin, genomic DNA substrates, biochemically assembled nucleosomes, and analogous naked DNA. To study whether viral DNA integration by PICs is distinct, we carried out comparative analysis with intasome (INS)-mediated integration. Our results demonstrated that PIC-mediated integration into biochemically assembled nucleosomes without histone tail modifications was lower compared to the naked DNA substrates. Notably, the addition of a trimethylated histone tail modification H3K36me3 significantly enhanced PIC-mediated integration into nucleosomes. The addition of linker DNA to the modified nucleosomes optimally supported PIC-mediated integration but not INS-mediated integration. Surprisingly, the cellular cofactor LEDGF/p75 had distinct effects on PIC compared to the INS. Furthermore, chromatin compaction by the linker histone H1° negatively regulated HIV-1 integration. Finally, using sequencing analysis, we identified integration preferences within specific regions of the nucleosomal DNA. Overall, our study provides critical biochemical evidence for HIV-1 PIC-mediated integration preference into open chromatin.

3.4 Results

3.4.1 Chromatin is the preferred substrate for HIV-1 PIC-mediated viral DNA integration

HIV-1 infection is dependent on the integration of the viral DNA into host chromosomes by the PIC. There is strong evidence that HIV-1 DNA is preferentially integrated into actively transcribing genes [87, 86, 92, 101]. However, the biochemical determinants of integration site preference within the chromatin are not fully understood. To better understand the mechanism of HIV-1 integration into chromatin hot spots, we measured viral DNA integration into target substrates such as isolated chromatin, genomic DNA, biochemically reconstituted nucleosome core particles with or without linker DNA, and the analogous naked DNA sequences. To study viral DNA integration in a physiologically relevant system, we extracted HIV-1 PICs from SupT1 cells acutely infected with wild-type enveloped HIV-1 particles. The extracted PICs retain robust

DNA integration activity in vitro (see Fig. S3.1A and B in the supplemental material) [143–147], and the assay is specific since the HIV-1 integrase inhibitor raltegravir (RAL) significantly inhibited viral DNA integration (see Fig. S3.1B).

Next, we extracted chromatin from HEK293T cells by adopting a protocol that yields high-quality chromatin through isolation and purification of clean nuclei [148]. The chromatin preparation contained the canonical histone proteins and lacked detectable levels of cytoplasmic protein markers (Fig. 3.1A and B; see also Figure S3.1C). Importantly, a partial micrococcal nuclease digestion of the isolated chromatin resulted in a "ladder-like" pattern with the smallest DNA band at 150 bp (Fig. 3.1C; see also Fig. S3.13C). The ~150-bp length is equivalent to the length of DNA (147 bp) wrapped around a nucleosome [106, 107, 149]. Collectively, these results demonstrate the integrity of the intact nucleosomes in the chromatin preparation. In parallel, the analogous dechromatinized genomic DNA (gDNA) substrate was prepared by deproteination of the isolated chromatin (Fig. 3.1C). To probe the integration into these substrates, PICs were incubated with either the chromatin or gDNA preparations containing equivalent amounts of DNA (300 ng). PIC-mediated DNA integration activity was measured by using Alu-based nested qPCR [143, 145–147]. The results from these measurements revealed that PIC-mediated viral DNA integration levels were significantly higher with the chromatin substrate compared to the gDNA (Fig. 3.1D and E). Even though substrate preference studies of HIV-1 PICs are limited, these results are consistent with prior studies using recombinant integrase showing that chromatinized substrates can support higher quantities of integration relative to naked DNA [114, 115, 117, 118, 120, 130]. Notably, the chromatin substrate used in our assay contains the host factor LEDGF/p75 and a specific histone modification that are positively implicated in HIV-1 integration targeting (see Fig. S3.1C). Therefore, the preference of PIC-mediated integration into the chromatin is most likely driven by both the host factors and the structural elements within the nucleosomes.

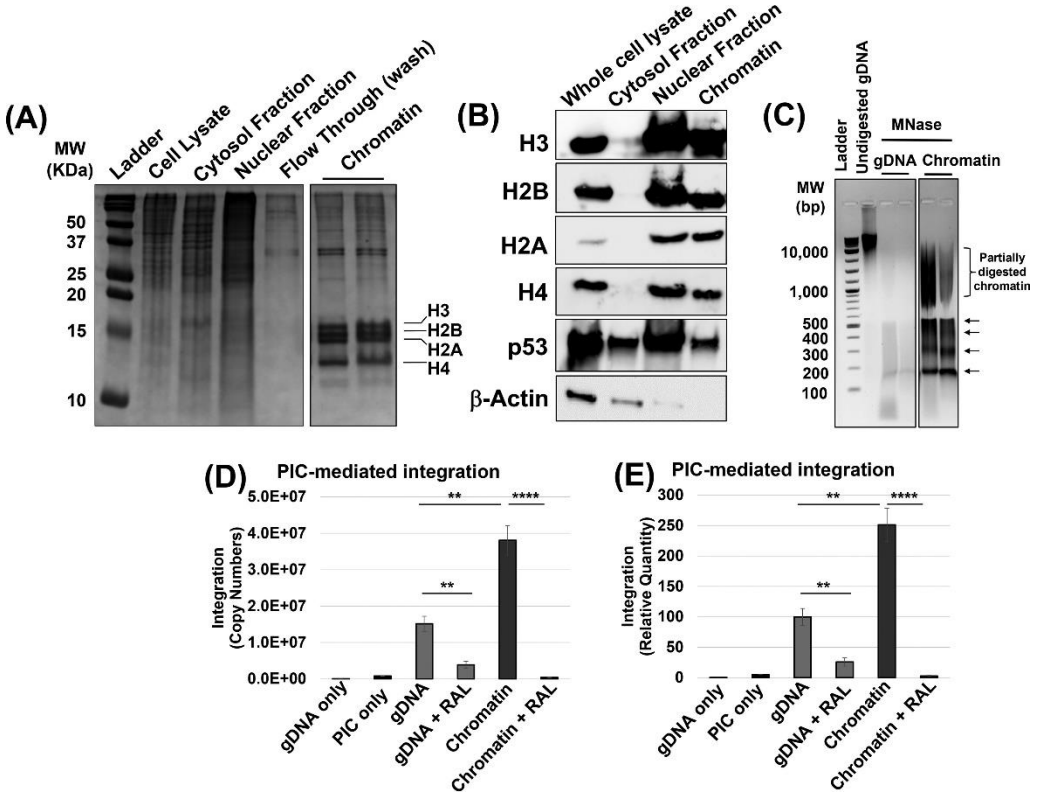


Figure 3.1: Chromatin is the preferred substrate for HIV-1 PIC-mediated integration. Chromatin was isolated from HEK293T cells and assessed for the histone proteins and DNA. (A) Fractions from various steps of the chromatin preparation were analyzed by using a 15% acrylamide gel and visualized by Coomassie staining. The chromatin fraction containing the canonical histone proteins H2A, H2B, H3, and H4 is shown. (B) The histone proteins were detected by Western blotting in the whole-cell lysate, the cytosol fraction, the nuclear fraction, and chromatin (left panel). In addition, cytoplasmic and nuclear proteins were probed in each fraction (right panel). (C) Protection of the chromatin DNA within nucleosomes was assessed by partial micrococcal nuclease (MNase) digestion. The arrows to the right of the gel indicate discrete DNA bands of nucleosome-mediated protection. (D) PIC-mediated integration was measured by nested-PCR and represented as the copy numbers of integrated viral DNA, using 300 ng of chromatin and deproteinated genomic DNA (gDNA) as targets. As a negative control, 1 μM RAL (the integrase strand transfer inhibitor) was used. (E) The copies of HIV-1 DNA integration were plotted relative to the integration into gDNA. The results in panel E are shown as means of the viral DNA copy numbers of at least three replicates, with the error bars indicating the standard errors of the mean (SEM). *, $P < 0.05$; **, $P = 0.01$ to 0.05; ***, $P = 0.01$ to 0.001; ****, $P = 0.001$ to 0.0001; *****, $P < 0.0001$.

3.4.2 Nucleosomes without histone tail modifications are not preferred for PIC-mediated viral DNA integration.

The basic repeating unit of chromatin is a nucleosome, consisting of a core histone octamer that wraps 147 bp of DNA [50, 51]. A number of published studies suggest that nucleosomes are the preferred substrates of retroviral DNA integration [115–118, 120, 150]. In particular, there is evidence that the outward-facing DNA within the nucleosome is targeted for retroviral DNA integration [115–118]. To probe whether the nucleosome structure contributed to the enhanced integration into chromatin (Fig. 3.1), we measured PIC-mediated integration using biochemically assembled nucleosomes. To generate nucleosome substrates, we used the Widom 601 nucleosome positioning sequence DNA [151] and purified human recombinant histones H2A, H2B, H3C110A, and H4 (Fig. 3.2A). First, histone octamers were generated from purified histone proteins by a well-established stepwise biochemical assembly approach [152]. Then, the 147-bp Widom 601 DNA (Fig. 3.2B) was added to the assembled histone octamers (Fig. 3.2C). Formation of the nucleosome core particle was confirmed by an electromobility shift assay (EMSA), and our purified nucleosome preparation was devoid of any free DNA (Fig. 3.2D).

Then, equivalent amount of the nucleosomes or the analogous naked Widom DNA (300 ng) was used as the substrate to measure PIC-mediated viral DNA integration. We used specific primers that amplify the junctions of integrated viral DNA into these particular target substrates by nested qPCR. Surprisingly, PIC-mediated integration into the naked DNA was significantly higher (~7-fold) compared to the nucleosomes (Fig. 3.2E). The specificity of the assay was demonstrated by the inhibition of PIC-mediated integration activity by 1 μM RAL with both the nucleosome and naked DNA (Fig. 3.2E). These results deviate from prior studies showing that nucleosomes are preferred substrates for integration compared to naked DNA. A key difference is that our study uses PICs extracted from infected cells compared to the in vitro studies of purified IN enzymes [94, 95, 115–118, 120, 121, 130]. Therefore, it is likely that the PIC-associated factors play key roles in integration preference beyond what might be driven by the IN alone.

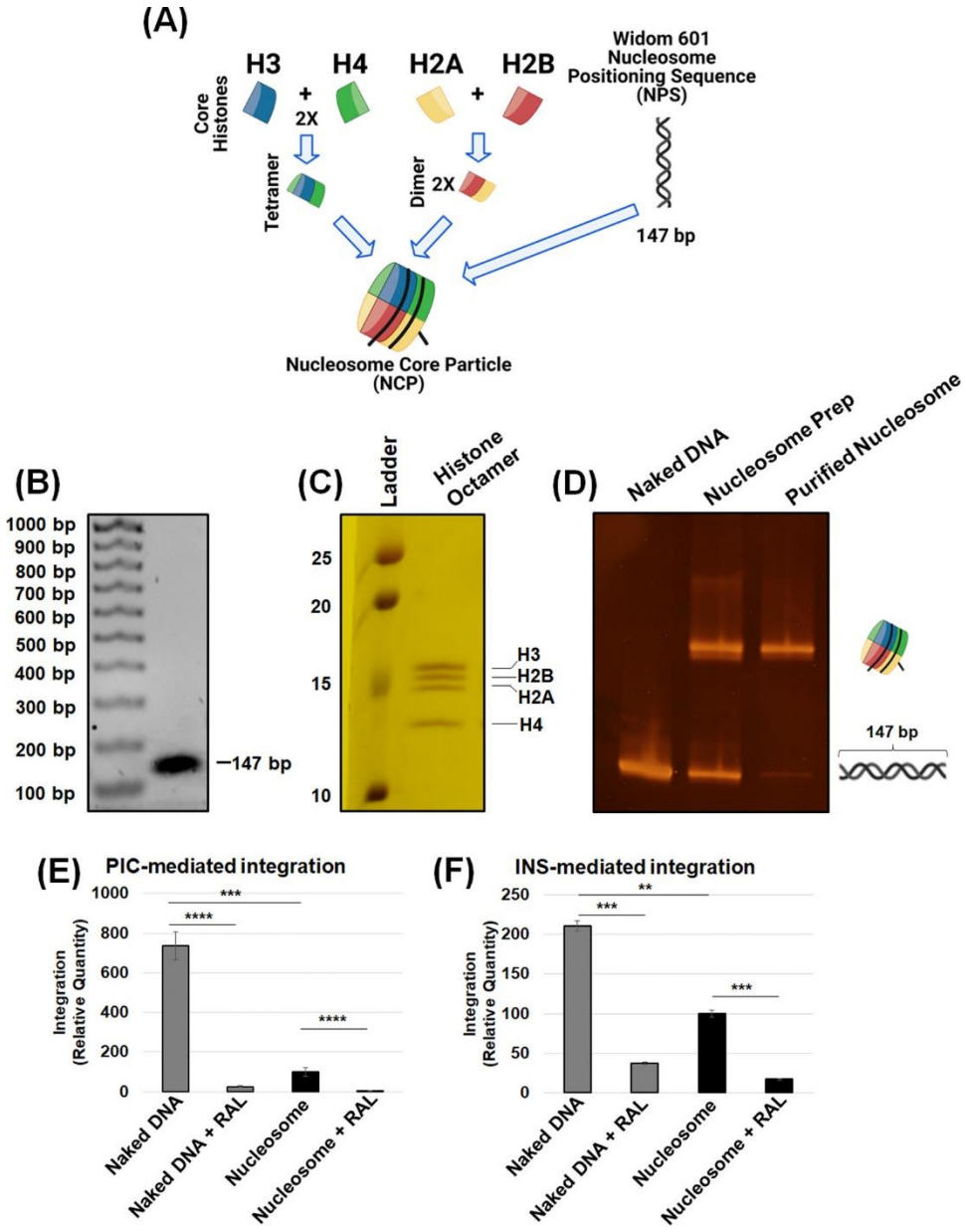


Figure 3.2: The nucleosome is a barrier to HIV-1 integration. (A) Schematic of the biochemical assembly of nucleosomes. The nucleosomes are assembled with the Widom 601 nucleosome positioning sequence and a recombinant human histone octamer. The histone octamer was assembled first by an equimolar addition of histone protein dimers H2A and H2B with the H3/H4 tetramer. (B) The 147-bp NPS (naked DNA) DNA used for nucleosome assembly was analyzed on a 1.5% agarose gel and visualized by ethidium bromide staining. (C) SDS-PAGE and Coomassie staining visualized the presence of individual histone proteins in the octameric histone assembly.

(D) The biochemical assembly of the nucleosome was analyzed by an EMSA. (E) PIC-mediated integration was assessed using 300 ng of the 147-bp naked DNA and the analogous nucleosome as targets, with RAL serving as a specific inhibitor for HIV-1 integration. (F) Intasome (INS)-mediated integration was measured using qPCR, with the naked DNA and the nucleosome serving as targets. The data are represented as the relative quantity of viral DNA integration in reference to the naked DNA, and error bars were generated from the SEM of at least three independent experiments. *, $P < 0.05$; **, $P = 0.01$ to 0.05; ***, $P = 0.01$ to 0.001; ****, $P = 0.001$ to 0.0001

To better understand the PIC-mediated DNA integration into nucleosomes, we probed whether the reduced levels of integration is PIC specific. To test this, we used HIV-1 intasomes (INS), which are formed by IN and viral DNA sequences and constitute the minimal substructures of PICs capable of carrying out viral DNA integration [153, 154]. We assembled HIV-1 INS through an established biochemical approach that uses purified IN and viral DNA sequences from the long terminal repeat (LTR) region [153, 154]. We used the purified recombinant HIV-1 IN containing a Ssod7 domain for solubility [98] and viral DNA mimics (25 and 27 bp) (see Fig. S3.2B). We evaluated INS-mediated viral DNA integration by using a qPCR-based assay to amplify the integration junctions into these specific target substrates (see Fig. S3.2B). Notably, this qPCR-based assay is designed to detect both half-site and full-site strand transfer. Our results show selective amplification of the viral and target DNA junction only when both the INS and the target substrate were added to the reaction mixture (see Fig. S3.2C). Since RAL inhibited INS-mediated integration activity, the assay is specific (see Fig. S3.2C). Next, INS activity was measured with the naked DNA and nucleosome substrates (Fig. 3.2F). We observed that INS-mediated integration was significantly higher with the naked DNA compared to the nucleosome (Fig. 3.2F), an observation similar to the PIC activity (Fig. 3.2E). Interestingly, the level of reduction in INS activity into nucleosomes was ~2-fold compared to a reduction of ~7-fold in PIC activity. These observations suggest that nucleosomes confer a stronger barrier for the PICs to insert the viral DNA. Even though previous studies suggest that nucleosomes are preferred for HIV-1 integration, the source of nucleosomes in our study differs from those of previously published studies. For instance, published studies used nucleosomes and chromatinized substrates, assembled using histones derived from cellular sources [94, 95, 97, 106, 120, 121, 155]. In contrast, our study used biochemically assembled recombinant nucleosomes assembled with purified histone proteins and

the synthetic Widom 601 sequence that forms highly stable nucleosomes compared to native DNA sequences [150, 151, 156]. Thus, the nucleosomes in our study lack any histone tail modifications, whereas the cellular histones used in the previously published studies most likely contain physiologically relevant and possibly integration-stimulating histone modifications. Taken together, these observations suggest that highly stable nucleosomes without any histone tail modifications are not preferred for HIV-1 DNA integration over the analogous naked DNA by both the PIC and the INS.

3.4.3 Nucleosomes containing a trimethylated histone 3 at lysine 36 enhanced PIC-mediated viral DNA integration.

In chromatin, nucleosomes are decorated with chemical modifications primarily in the extended tails of the histone proteins [112, 157, 158]. In the case of HIV-1, histone modifications associated with active transcription are positively correlated with integration preference, whereas histone marks indicative of heterochromatin and transcriptionally silent genes are negatively associated with DNA integration [89–91, 93, 96, 101]. Given our unexpected results that nucleosomes without histone modifications are not preferred for HIV-1 DNA integration in vitro (Fig. 3.2), we assessed whether nucleosomes containing specific histone modifications are required for optimal integration. One of the histone modifications most commonly associated with HIV-1 integration preference in sequencing studies is the trimethylated histone 3 at lysine 36 (H3K36me3) [89, 93, 96, 98, 101, 102, 122, 159, 160]. As a proof of concept, we tested whether nucleosomes containing the H3K36me3 mark affects integration activity of both PICs and INS. We assembled nucleosomes with the biochemical mimetic (H3K36Cme3) of the H3K36me3 modification by an alkylation reaction (Fig. 3.3A). The H3K36Cme3 mark itself is not known to enhance nucleosomal DNA unwrapping [161, 162] and is biochemically indistinguishable from the posttranslational modification found in cells [163]. Our Western blot analysis confirmed the presence of H3K36me3 in isolated chromatin and H3K36Cme3-modified nucleosomes (Fig. 3.3B). When we assessed PIC-mediated integration activity using the H3K36Cme3-modified nucleosomes as the substrate, we observed a significant increase in integration compared to the analogous

unmodified nucleosome substrate (Fig. 3.3C). However, PIC activity with the H3K36Cme3-modified nucleosome substrate remained significantly lower compared to the analogous naked DNA substrate (Fig. 3.3C; see also Fig. S3.3B and C). To further study this phenotype, we measured INS-mediated integration activity using the naked DNA, nucleosomes, and H3K36Cme3-modified nucleosomes as target substrates. In stark contrast to the PIC activity (Fig. 3.3C), INS activity with the H3K36Cme3-modified nucleosome substrate showed no measurable

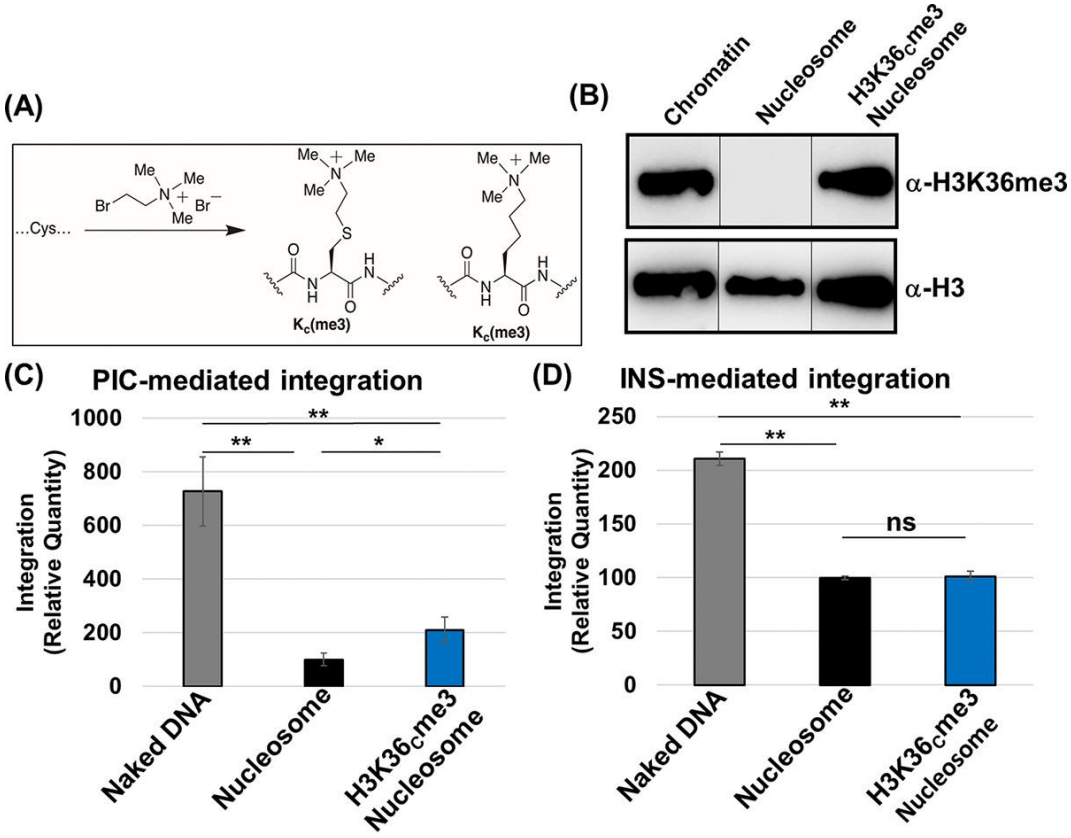


Figure 3.3: Nucleosomes containing trimethylated histone 3 at lysine 36 enhanced PIC-mediated integration. (A) Schematic illustrating the insertion of the H3K36me3 mimetic by site-directed mutagenesis of the H3 K36 to cysteine (K36C). The K36C H3 undergoes alkylation to functionalize the K36C to create a biochemical mimic of the histone 3 trimethylation (H3KC36me3). (B) Western blot analysis of the H3K36me3 and histone H3 in the H3K36Cme3 nucleosome compared to the unmodified nucleosome and the isolated chromatin. (C) PIC-mediated integration with the naked DNA, nucleosome, and H3K36Cme3 nucleosome. The data were analyzed with reference to the naked DNA. (D) The INS-mediated integration was then assessed by comparing the naked DNA, nucleosome, and H3K36Cme3 nucleosome. The error bars were determined by the SEM of at least three independent replicates. *, $P < 0.05$; **, $P = 0.01$ to 0.05.

difference relative to the unmodified nucleosomes (Fig. 3.3D; see also Fig. S3.3D). Interestingly, the INS activity remained higher with the naked DNA substrate compared to either the modified or the unmodified nucleosome substrates (Fig. 3.3D; see also Fig. S3.3D). These observations strongly suggest that the H3K36me3 modification specifically promotes viral DNA integration, by the PIC but not by the INS, into nucleosomes.

3.4.4 Nucleosomes harboring H3K36me3 and linker DNA optimally promoted PIC-mediated viral DNA integration.

Our results show that unmodified nucleosomes without flanking linker DNA are not preferred for integration compared to the analogous naked DNA (Fig. 3.3C and D). In the cellular context, nucleosomes are connected to adjacent nucleosomes by a linker DNA [149, 164]. Therefore, we used nucleosomes mimicking the chromatin by assembling nucleosomes containing an additional 50 bp of linker DNA flanking both sides of the 147-bp Widom 601 DNA (Fig. 3.4A and B; see also Fig. S3.4A). Imperfect placement of DNA longer than 147 bp in the nucleosomes [165] most likely explains the presence of more than one population of nucleosomes (Fig. 3.4B). First, we measured INS-mediated integration into the 247-bp naked DNA and compared it to the nucleosomes assembled with the same DNA (Fig. 3.4C). We observed a reduction in the integration with the nucleosomes compared to the naked DNA, similar to our results with 147-bp naked DNA and nucleosome substrates (Fig. 3.2E and F). Moreover, results with the linker nucleosome without or with H3K36Cme3 modification indicated that histone trimethylation has a negligible effect on INS activity (Fig. 3.4D). Comparison of all three linker substrates surprisingly showed the preference of INS-mediated integration into the naked DNA (Fig. 3.4D). Next, we tested PIC-mediated integration with all the linker substrates. The naked 247-bp DNA was preferred for PIC activity over the linker nucleosomes (Fig. 3.4E), similar to the data of INS activity (Fig. 3.4C and D) and PIC activity with nonlinker substrates (Fig. 3.3C and D). Interestingly, the results of PIC activity with the 247-bp DNA and linker nucleosomes with H3K36Cme3 revealed a significant increase in integration with the H3K36Cme3-modified nucleosome substrate (Fig. 3.4F; see also

Fig. S3.4D and E). Comparative analysis of PIC activity with all of the linker substrates showed that the linker H3K36Cme3 nucleosomes supported significantly higher integration than the linker nucleosomes and naked DNA (Fig. 3.4F). This is in stark contrast to the results with INS (Fig. 3.4D), strongly suggesting that the combination of the H3K36me3 and the linker DNA provides a biochemically favorable environment for the HIV-1 PIC-mediated viral DNA integration. Furthermore, these data also indicate that only the PIC, but not the PIC-substructure INS, contains the necessary factors/mechanisms to mediate HIV-1 DNA integration into H3K36me3-modified nucleosomes.

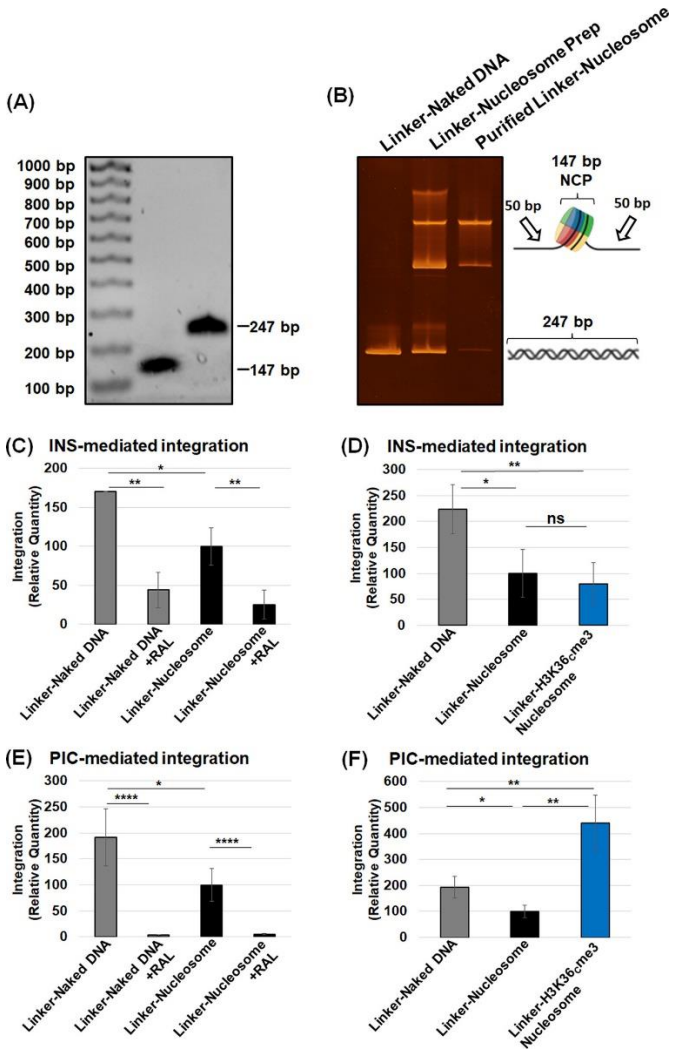


Figure 3.4: The nucleosome containing linker DNA and the H3K36Cme3 is an optimal substrate for PIC-mediated integration. The nucleosome containing 50 bp of linker DNA flanking the Widom 601 NPS (linker-naked DNA) was assembled, resulting in a nucleosome with a linker substrate that is 247 bp. (A) The 147- and 247-bp nucleosome positioning sequence DNA was analyzed on a 1.5% agarose gel and visualized by ethidium bromide staining. (B) The biochemical

assembly of nucleosomes with the 247-bp DNA was analyzed by EMSA. (C) The relative quantity of INS-mediated integration was measured with the linker-naked DNA, compared to the linker-nucleosomal DNA. The integrase inhibitor, RAL, was included as a control for integration. (D) The linker-naked DNA, linker-nucleosome, and linker-H3K36Cme3 were compared for INS-mediated integration relative to the linker-naked DNA. (E) PIC-mediated integration was assessed with the linker-naked DNA and linker-nucleosome. (F) PIC-mediated integration was measured by comparing the linker-naked DNA, linker-nucleosomes, and linker-H3K36Cme3. All the results are shown as the relative integration quantity and represent the means of at least three independent experiments. Error bars represent the SEM (*, $P < 0.05$; **, $P = 0.01$ to 0.05; ****, $P = 0.001$ to 0.0001).

3.4.5 LEDGF/p75 addition stimulated INS-mediated viral DNA integration but reduced PIC-mediated viral DNA integration.

The PIC utilizes several host proteins to efficiently integrate the viral DNA into the host chromosomes, most notably CPSF6 and LEDGF/p75 [129, 166–169]. LEDGF/p75 is known to bind mitotic chromatin [135, 170]; however, the consequence of this binding is just recently beginning to emerge, particularly in the context of transcription [171]. Early studies identified LEDGF/p75 as the key host protein to support HIV-1 integration [131–143, 139–141, 155]. In particular, LEDGF/p75 binds to an integrase dimer at the catalytic core domain, plays a role in targeting integration toward actively transcribing genes, and stimulates the strand transfer activity of lentiviral integrases, including HIV, with both naked and chromatinized substrates [172–175]. Therefore, to better understand our contrasting results in Fig. 3.4, we tested the effects of LEDGF/p75 on INS- and PIC-mediated viral DNA integration using nucleosome and naked DNA substrates. First, increasing amounts of purified LEDGF/p75 were supplemented to the INS reaction mixture containing the nucleosomes or naked DNA substrates. We observed a dose-dependent enhancement of INS-mediated integration into the naked DNA with increasing concentrations of LEDGF/p75 (Fig. 3.5A). LEDGF/p75 addition also stimulated INS activity with the nucleosomes in a concentration-dependent manner (Fig. 3.5B). Surprisingly, increasing amounts of LEDGF/p75 did not enhance PIC activity with the naked DNA albeit a slight but nonsignificant decrease at the highest concentrations (Fig. 3.5C). Furthermore, LEDGF/p75 addition significantly reduced PIC activity with the unmodified nucleosome substrate (Fig. 3.5D), in contrast to the enhanced effect observed with the INS (Fig. 3.5A and B). Interestingly, Western

blot analysis showed that endogenous LEDGF/p75 is present in our extracted PIC preparation (Fig. 3.5E). Therefore, it is plausible that the PIC-associated LEDGF/p75 levels are sufficient for viral DNA integration and that additional amounts of the protein confer an inhibitory effect.

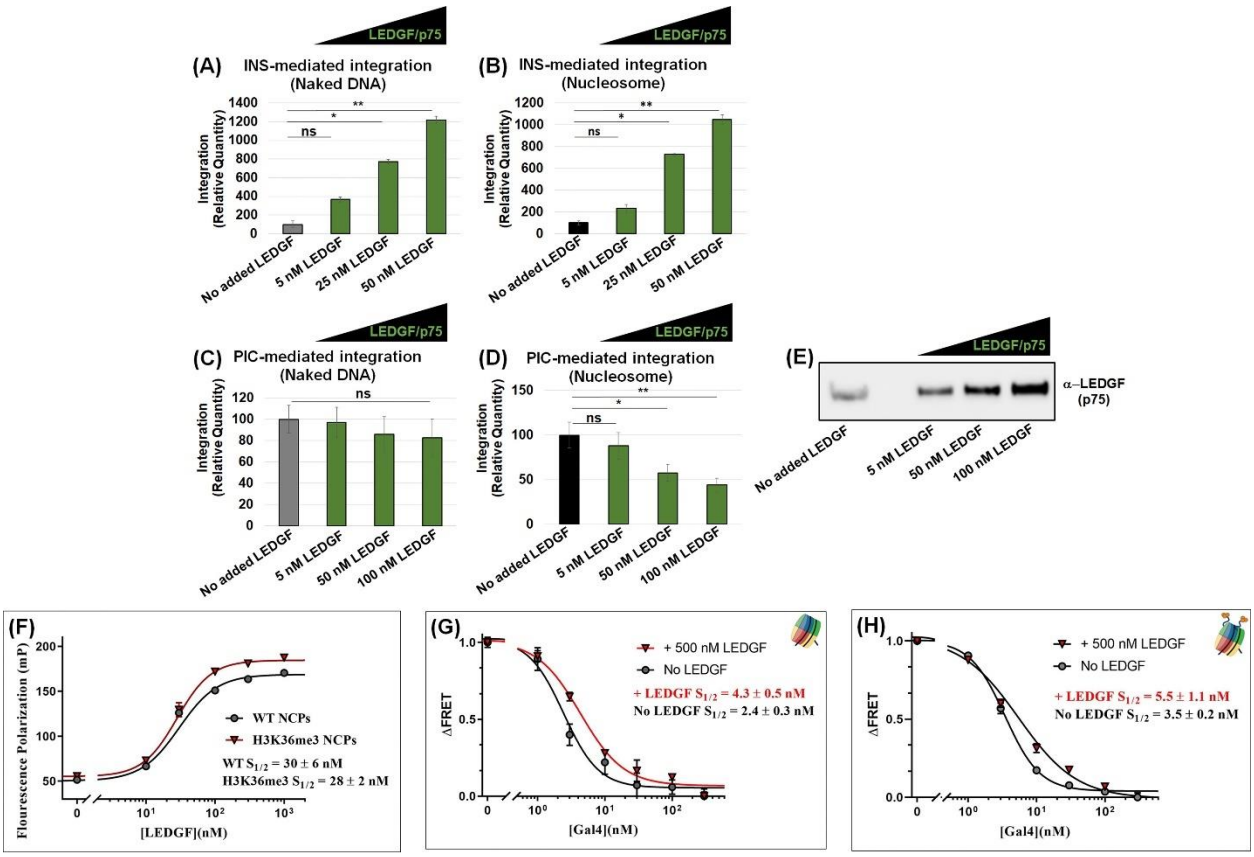


Figure 3.5: LEDGF/p75 addition stimulated INS-mediated integration but reduced PIC-mediated integration with nucleosomes. (A and B) INS-mediated integration (25 nM) was measured in the presence of the integration cofactor LEDGF/p75. LEDGF/p75 was added to the indicated reaction mixtures at 5, 10, and 25 nM to both the naked DNA and the nucleosome substrate. (C and D) PIC-mediated integration was measured in the presence of LEDGF/p75 addition with both the naked DNA and the nucleosome substrate. (E) Western blot analysis for LEDGF/p75 in the isolated PIC preparation and in PICs supplemented with the recombinant LEDGF/p75 protein. (F) A fluorescence polarization assay was performed with fluorescein-labeled nucleosome core particle (NCP) to determine the LEDGF/p75 binding kinetics (S1/2) to the nucleosomes. (G and H) The ensemble FRET measurements were detected by Cy3-Cy5 NCP (Cy3 at the end of the NPS and Cy5 at the H2A K119C) and with titrations of GAL4 in the presence of LEDGF/p75 at saturating amounts of the unmodified NCP and H3K36Cme3-NCP. The error bars represent the SEM of at least three independent experiments. *, $P < 0.05$; **, $P = 0.01$ to 0.05.

To understand the mechanism by which LEDGF/p75 reduced PIC-mediated integration into the nucleosome, we employed a fluorescence resonance energy transfer (FRET)-based binding assay. This FRET assay utilizes a 5′-Cy3-labeled nucleosome positioning sequence reconstituted into a nucleosome containing a Cy5 label on the H2A K119C within the histone octamer [176, 177]. The positioning of these two fluorophores allowed for the measurement of nucleosome accessibility near the DNA entry/exit site. Since LEDGF/p75 is known to bind to the nucleosome at the entry/exit site [178], this assay is uniquely suited for our study. In this assay, a GAL4 binding sequence was cloned into the Widom 601 DNA (at nucleotide positions 6 to 25) to determine GAL4 accessibility to the nucleosome in direct competition with LEDGF/p75 (see Fig. S3.5C) [161, 165, 176]. Initially, we demonstrated that LEDGF/p75 binds to the nucleosome by EMSA (see Fig. S3.5A), consistent with earlier studies [178, 179]. This LEDGF/p75 binding to the nucleosome requires the DNA-binding domains, PWWP and AT-hooks, since a LEDGF/p75 mutant containing only the integrase-binding domain (IBD) failed to bind the nucleosome (see Fig. S3.5B). We observed that in the absence of GAL4, the addition of increasing amounts of LEDGF/p75 did not alter the FRET efficiency (see Fig. S3.5D). These results established that upon binding to the nucleosome, LEDGF/p75 addition minimally affects nucleosomal DNA unwrapping. Using fluorescence polarization analysis, we observed that LEDGF/p75 binds to both the unmodified nucleosome and H3K36Cme3 nucleosome substrates with comparable affinities of 30 and 28 nM, respectively (Fig. 3.5F). We then determined the nucleosome accessibility to GAL4 in the presence of LEDGF/p75 (Fig. 3.5G and H). LEDGF/p75 addition reduced the accessibility of GAL4 to the nucleosome by ~2-fold. Likewise, GAL4 accessibility to the DNA in the nucleosome containing H3K36me3 was also reduced by LEDGF/p75. Collectively, these results show that LEDGF/p75 can physically or sterically block trans-factors from gaining access to the DNA within the nucleosome. Thus, an excess of LEDGF/p75 likely reduces PIC-mediated integration by blocking access of the PIC-associated viral DNA to the nucleosomal DNA.

3.4.6 The H3K36me3 modification relieved the LEDGF/p75-mediated reduction of PIC-mediated viral DNA integration.

Our results indicated that the nucleosome features (histone tail modification and DNA length) and the chromatin-binding protein LEDGF/p75 directly influence viral DNA integration. The combination of the H3K36Cme3 and linker DNA within the nucleosome has an additive effect on PIC-mediated viral DNA integration (Fig. 3.4F). In contrast, PIC activity with the nonlinker nucleosome showed a LEDGF/p75-mediated reduction (Fig. 3.5D). To reconcile these disparate

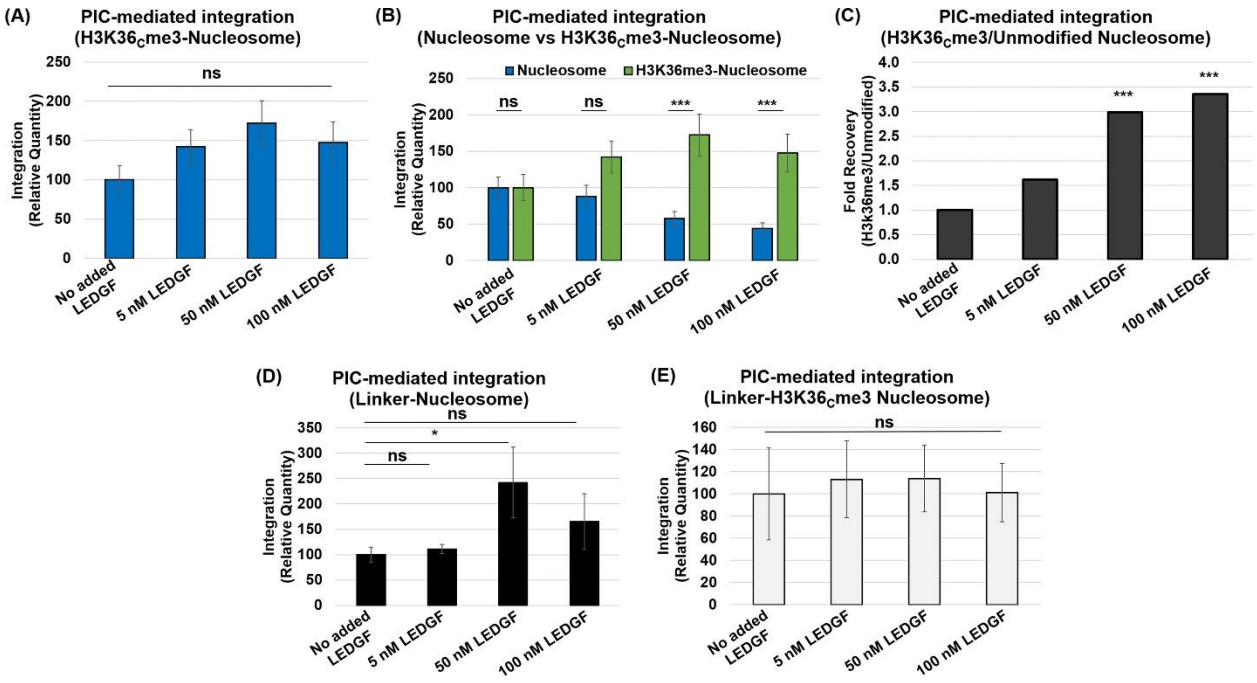


Figure 3.6: The H3K36Cme3 nucleosome and linker DNA supported PIC-mediated integration in the presence of LEDGF/p75. (A) PIC-mediated integration was measured using H3K36me3-nucleosomes as targets in the presence of LEDGF/p75 (5, 50, and 100 nM). Integration data are presented relative to the PIC reactions without added LEDGF/p75. (B) Comparative analysis of the relative PIC-mediated integration with the unmodified nucleosome to the H3K36Cme3-nucleosome in the presence of LEDGF/p75. (C) The fold change was calculated between the unmodified nucleosome and the H3K36me3-modified nucleosome effects on PIC-mediated integration in the presence of LEDGF/p75. (D) PIC-mediated integration was measured with LEDGF/p75 supplementation to the unmodified nucleosomes containing linker DNA. (E) PIC-mediated integration with the H3K36Cme3 nucleosome containing linker DNA in the presence of LEDGF/p75 is plotted relative to the assay without LEDGF/p75 supplementation. The error bars represent the SEM of at least three independent experiments. *, $P < 0.05$; ***, $P = 0.01$ to 0.001.

results of LEDGF/p75 with distinct substrates, we measured PIC activity using the linker substrates in the presence of LEDGF/p75. First, we observed that with the H3K36Cme3-modified nucleosome without the linker DNA, LEDGF/p75 addition did not inhibit PIC activity (Fig. 3.6A). Interestingly, when we juxtaposed the LEDGF/p75 addition data of the unmodified nucleosomes to the H3K36Cme3-modified nucleosome, the data indicated that the H3K36me3 mark relieves the LEDGF/p75-mediated reduction of PIC activity (Fig. 3.6B). Further analysis revealed a significant ~3-fold recovery of PIC activity with the H3K36Cme3-modified substrate over the unmodified substrate in the presence of 50 to 100 nM LEDGF/p75 (Fig. 3.6C). Intriguingly, the presence of linker DNA also relieved the LEDGF/p75-mediated reduction of PIC activity (Fig. 3.6D), an observation in clear contrast to the dose-dependent reduction of PIC activity with the unmodified nucleosome (Fig. 3.5D). Finally, PIC activity with the linker H3K36Cme3-modified nucleosome showed no significant effect of LEDGF/p75 addition on integration (Fig. 3.6E). Notably, the raw integration copy numbers indicated that the integration quantity is relatively high (see Fig. S3.6C), suggesting that in vitro integration occurs at an optimal level with the nucleosome substrate containing both the H3K36Cme3 and linker DNA, with the endogenous LEDGF/p75. These observations establish that the presence of linker DNA and the H3K36me3 mark within a nucleosome are biochemically favored for HIV-1 integration by the PIC.

3.4.7 The nucleosome core is preferentially targeted for integration when modified with H3K36me3.

Our results identified that nucleosomes with linker DNAs and histone tail modification akin to open chromatin enhanced PIC-mediated viral DNA integration (Fig. 3.4F; see also Fig. S3.4D and E). In the human genome, the H3K36me3 mark is primarily located within the gene body of actively transcribing genes [180–182]. The H3K36me3 mark is associated with promoting polymerase II elongation through the gene body, gene splicing, suppression of cryptic transcription, and DNA damage repair [183–186]. While sequencing studies have correlated the proximity of H3K36me3 to HIV-1 integration targeting [159, 187, 188], a direct role of this

epigenetic mark in the local viral DNA targeting within a nucleosome remains unknown. Therefore, to understand the consequence of the H3K36Cme3 mark in viral DNA integration, we carried out deep sequencing of the integration reactions with the linker substrates [155]. Duplicate integration reactions of the naked DNA with linker, unmodified nucleosome with linker, and H3KC36me3 nucleosome with linker substrates were PCR amplified, concentrated, and spectrophotometrically analyzed. Amplicons were deep sequenced, and we quantified reads that contained sequences of the HIV-1 5′ LTR adjoined to the DNA sequence for the naked DNA, nucleosome, and H3K36Cme3 nucleosome targets with linkers. Analysis of the sequencing data identified integration junctions and revealed distinct integration patterns in each of the substrates (see Fig.

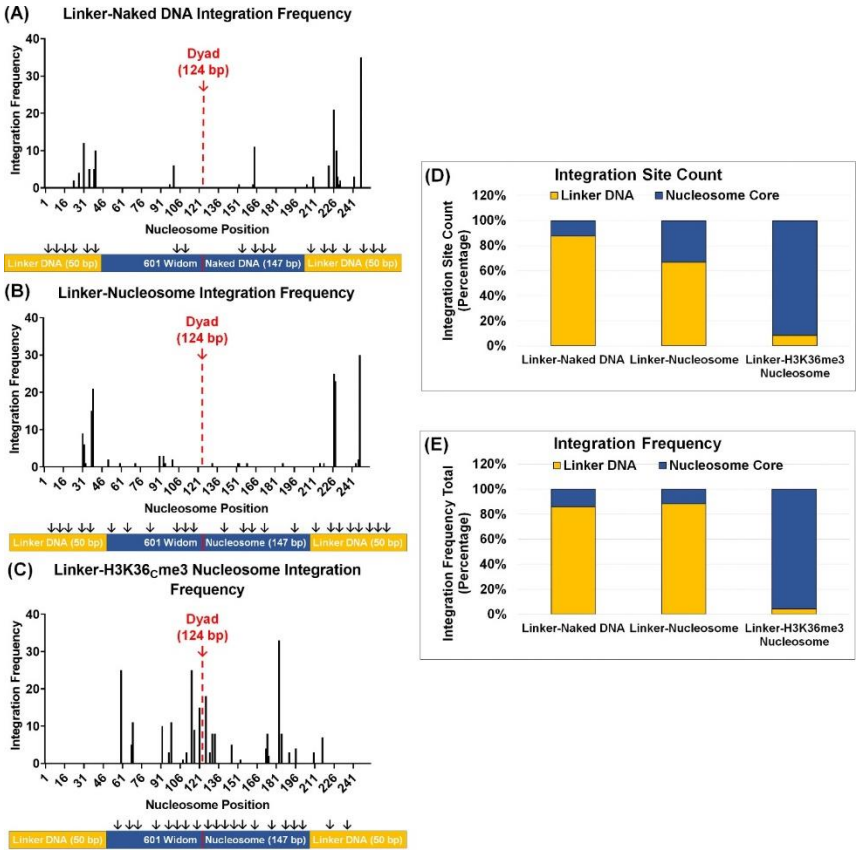


Figure 3.7: HIV-1 DNA integration is preferentially directed into the core of the H3K36Cme3 containing nucleosomes. To study HIV-1 integration preference, the DNA from the integration reactions was PCR amplified and subjected to next-generation sequencing. The integration frequency within the linker substrates was determined by quantifying the integration junctions. The integration frequency is plotted as a histogram for the naked DNA with a linker (A), the nucleosome with the linker (B), and H3K36Cme3 with linker substrates (C). (D) After the integration sites within the linker containing DNA substrates were quantified, the percentages of

sites within the linker sequences and nucleosome core sequence were plotted for comparative analysis. (E) The frequency of integration junctions at a particular site within the sequence was then determined. The integration frequency in the linker sequences and the nucleosome core sequence was quantified as a percentage of the total integration sites.

S3.8 and 3.10). With the linker containing naked DNA, integration junctions were identified throughout the length of the sequence with an overrepresentation of viral DNA integration near the target DNA ends (Fig. 3.7A; see also Fig. S3.8). Similarly, the integration junctions were mostly found in the linker DNA regions of the nucleosome (Fig. 3.7B; see also Fig. S3.9). In contrast, the integration junctions within the linker containing H3K36Cme3 nucleosome were identified predominantly in the nucleosome core sequence (Fig. 3.7C; see also Fig. S3.7 and Fig. S3.10). The location of the integration sites within the linker H3K36Cme3 nucleosome was mapped near the entry/exit sites of the nucleosome. Importantly, this is the location of the H3-tail protrusion between the nucleosomal DNA [111, 176, 178], indicating that the HIV-1 PIC is drawn toward the H3K36me3 mark when engaged with a nucleosome during an integration event. To further understand these sequencing data, we quantified the integration site counts as a percentage of the total integration events. These analyses revealed that the linker naked DNA and linker nucleosome comprised ~87% and ~67% of the junctions within the linker DNA, respectively (Fig. 3.7D). Distinctly, the linker H3K36me3 nucleosome contained over ~90% of the integration junctions adjoined in the nucleosome core sequence. Integration frequencies varied considerably between unmodified and H3K36Cme3-modified nucleosomes, wherein the nucleosome core of the linker H3K36Cme3-modified nucleosome contained ~95% of the integration junctions (Fig. 3.7E; see also Fig. S3.7B). However, the frequency of integration within the nucleosome core was only ~15% with the linker DNA and linker nucleosome lacking any histone modifications (Fig. 3.7E; see also Fig. S3.7B). Collectively, these results indicate that H3K36me3 modification promotes HIV-1 integration into specific regions of the nucleosome.

3.4.8 Chromatin compaction by histone H1 inhibited HIV-1 DNA integration.

To further understand the role of chromatin structure on HIV-1 integration, we tested the effects of H1 protein on INS and PIC activity. Linker histone H1 reduces chromatin accessibility by facilitating chromatin compaction [89, 189] through the direct interaction with the linker DNA near the entry/exit location of the nucleosome and at the nucleosome dyad [108, 190]. A previous study has reported that H1-mediated compaction of nucleosome arrays can reduce HIV-1 integrase activity relative to open (noncondensed) chromatinized substrates [94]. Therefore, we measured HIV-1 DNA integration by both PIC and INS in the presence of increasing concentrations of purified recombinant histone H1° protein. Our results revealed that INS-mediated integration was significantly reduced in the presence of nanomolar concentrations of H1° protein (Fig. 3.8A to C). The inhibitory effect of H1 was universally observed with the linker naked DNA, as well as, with nucleosomes without or with linker DNA. Surprisingly, PIC-mediated integration was not inhibited in the presence of nanomolar concentrations of H1° protein (Fig. 3.8D to F). Rather, with the linker substrates, H1° either modestly enhanced PIC activity with the linker naked DNA or had a minimal effect with both the unmodified and the H3K36me3-modified linker nucleosome substrates (Fig. 3.8D to F). Then, we tested the effects of H1° at saturating concentrations of 287 and 1,474 μM, which correspond to 1:1 and 1:5 molar ratios to the substrate concentrations, respectively. At these concentrations, the PIC activity was significantly inhibited with the naked DNA and unmodified and H3K36Cme3-modified nucleosome substrates containing the linker DNA (Fig. 3.8G to I). Finally, we tested the effects of H1° protein addition on PIC activity using chromatin and dechromatinized gDNA substrates. After preincubating the substrates with saturating concentrations of H1°, we observed significantly reduced PIC-mediated integration with both substrates (Fig. 3.8J and K; see also Fig. S3.13E and F). Notably, nanomolar amounts of H1° addition did not reduce PIC activity with the genomic DNA and chromatin substrates (see Fig. S3.12A to D).

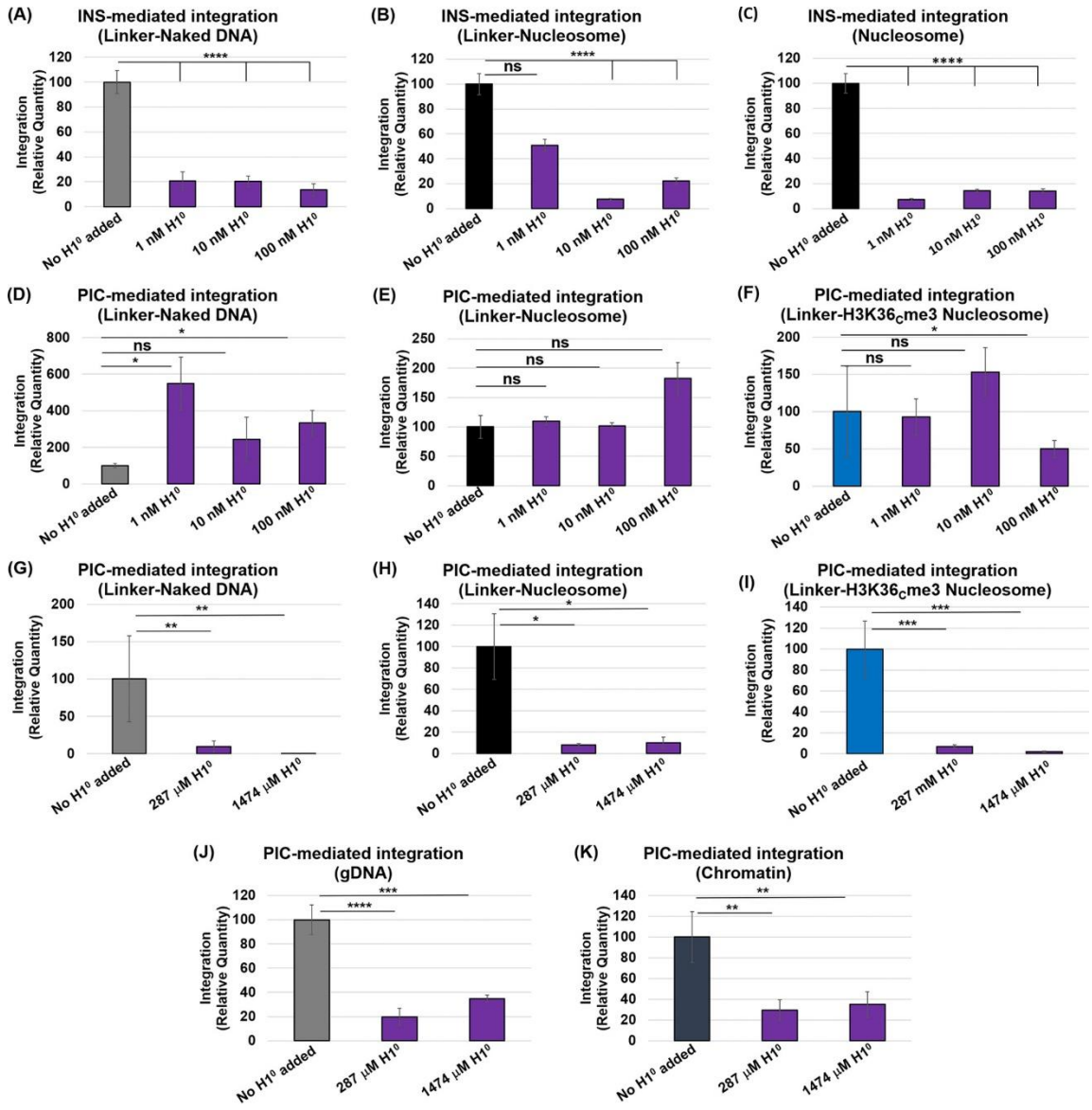


Figure 3.8: H1° reduces HIV-1 DNA integration. To probe the effects of H1 on HIV-1 integration, INS-mediated integration [25 nM] was first tested with the linker-naked DNA (A), the linker-nucleosome (B), or the non-linker-nucleosome (147 bp) (C) preincubated with recombinant H1° (1, 10, or 100 nM). The results are the average relative quantities with reference to the assay lacking H1° addition. The PIC-mediated integration was then measured with linker-naked DNA (D), linker-nucleosome (E), and linker-H3K36Cme3 nucleosome (F) that were preincubated with 1, 10, or 100 nM H1°. (G to I) Next, either 287 or 1,474 mM H1° was added to the PIC-mediated integration measurements with the linker-naked DNA, the linker-nucleosome, and the linker-H3K36Cme3 nucleosome substrates. The mM concentrations of H1° reflect amounts that show, respectively, 1:1 and 1:5 (wt/wt) stoichiometry of the substrate concentrations. The data shown are the relative quantities of the PIC integration relative to the assay lacking H1° addition. PIC-mediated integration was measured with gDNA (J) and chromatin (K) that were incubated on ice with 287 and 1,474 mM H1°. All data represent the means of at least three independent experiments, with the error bars representing the SEM (*, $P < 0.05$; **, $P = 0.01$ to 0.05; ***, $P = 0.01$ to 0.001; ****, $P = 0.001$ to 0.0001).

Finally, to determine whether these in vitro biochemical effects translate into a cellular context, we probed whether H1 expression regulates HIV-1 integration into human chromosomes in infected cells. To assess this, we overexpressed H1° in HEK293T cells before inoculation with pseudotyped HIV-1 virions. H1° overexpression in these cells was confirmed by Western blot analysis (Fig. 3.9A). Importantly, there is evidence that H1° overexpression does not alter cell viability [191]. The genomic DNA was extracted from the WT and H1°-overexpressing cells, and the proviral integration (PVI) was measured using the nested Alu-Gag PCR method [87, 90, 136]. Our results revealed that proviral DNA integration was reduced by ~50% in the H1°-overexpressing cells relative to the control infected cells (Fig. 3.9B), implying a negative correlation between H1 expression and HIV-1 integration. Collectively, these results provide evidence that chromatin compaction by H1 reduces HIV-1 integration and, by extension, supports the model that open chromatin structure is preferred for viral DNA integration.

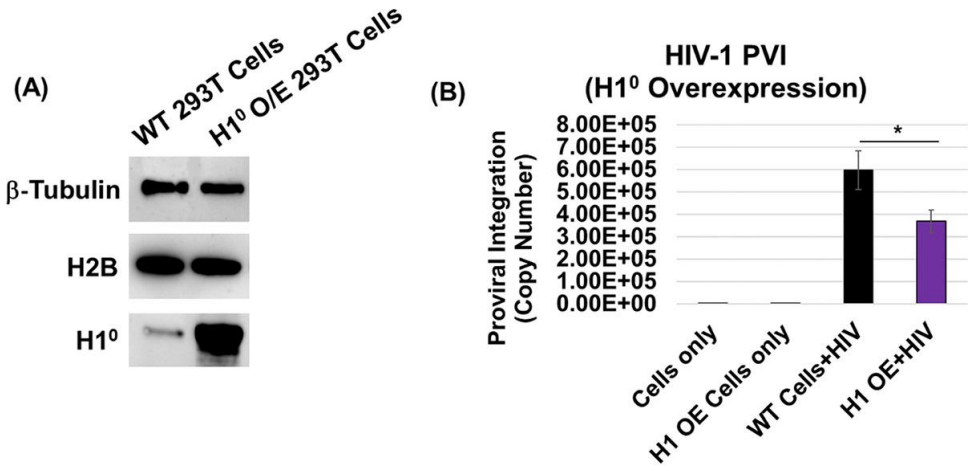


Figure 3.9: Histone H1 expression negatively regulates HIV-1 integration. (A) Linker histone H1° was overexpressed in HEK293T cells, H1° protein was probed by Western blot analysis, and the same blot was reprobed for β-tubulin and histone protein H2B as loading controls. (B) Proviral integration in the HEK293T cells overexpressing H1°. Cells were transfected with the pEV833 (GFP) expression construct for H1° and then inoculated with VSVg-dEnv-HIV-1 (GFP) particles. At 24 h postinoculation, HIV-1 proviral DNA integration (PVI) was quantified by nested Alu-PCR. All data represent at least three independent experiments, and the error bars represent the SEM (*, $P < 0.05$)

3.5 Discussion

HIV-1 infection is dependent on the integration of the viral DNA into genic hot spots of host chromosomes [67]. These hot spots contain actively transcribing genes that are epigenetically marked with histone tail modifications representative of open chromatin structure [89, 93, 96, 98, 101, 102]. However, the exact role of viral and host factors that program the PIC to integrate the viral DNA into these hot spots remains largely unclear. While specific chromatin features are associated with HIV-1 integration targeting, establishing a functional link between the markers of open chromatin and HIV-1 integration preference has been challenging [122, 124, 130, 193]. To address this long-standing challenge, we employed a biochemical approach involving the extraction of PICs from acutely infected cells and quantified viral DNA integration activity in vitro (see Fig. S3.1A and B) [143, 145, 146]. We used isolated chromatin, genomic DNA, in vitro assembled nucleosomes, and the analogous 147-bp naked DNA (Widom 601) to define the target substrate preference for HIV-1 DNA integration. Using this approach, we identified that the H3K36me3 histone modification, an epigenetic histone mark of active transcription, plays a direct role in HIV-1 DNA integration preference. In particular, our results show that nucleosomes assembled with the histone H3K36me3 are a preferred target and HIV-1 DNA integration predominantly occurred within the DNA wrapped around the nucleosome core. The H3K36me3 mark also relieved an unexpected LEDGF/p75-mediated block on HIV-1 DNA integration into the nucleosome. Finally, the histone H1 protein that induces compaction of chromatin significantly reduced HIV-1 DNA integration into nucleosomes, naked DNA, and chromatin substrates. Collectively, these studies establish a biochemical link between open chromatin structure and HIV-1 integration preference.

Retroviruses must complete an obligate integration step in order to propagate. Different genera of retroviruses employ distinct mechanisms to direct viral DNA integration into specific genomic regions of the host [88, 93, 121]. However, the molecular and biochemical mechanisms that drive HIV-1 DNA integration into specific genomic regions in the human chromatin are not fully understood. This is in part due to the complexity of the structural landscape of the human chromatin coupled with the lack of complete knowledge of the PIC function and composition.

Nucleosomes, the organizing unit of chromatin, regulate access to the genomic DNA through numerous intrinsic biochemical and structural mechanisms [111–113]. These structures also allow access to the nucleosomal DNA for genomic functions such as DNA replication, RNA transcription, and DNA damage repair. The near-meta-stable nature of the nucleosomes, driven by structural and conformation mechanisms involves (i) the intrinsic binding affinity of the histone octamers, (ii) post-translational modifications of the histone proteins and the DNA, (iii) the competitive or cooperative binding of other chromatin binding factors, and (iv) active translocation by ATP-dependent remodeling complexes [194–196]. With the exception of prototype foamy virus IN [150, 197–202], the mechanisms by which most retroviruses counter the intrinsic structural and conformational barriers to access the nucleosomal DNA to insert their viral DNA are not fully understood. Nonetheless, biochemical studies of MLV PICs, together with purified MLV and HIV IN, show a preference for the nucleosomal DNA relative to naked DNA as targets of integration [115–118]. The curvature of nucleosomal DNA bending around the histone octamers or even the histone proteins themselves seem to impact viral DNA integration preference [118, 119, 124, 193, 203]. Conversely, the integration activity of recombinant HIV-1 IN can also be reduced with biochemically assembled target substrates such as nucleosomes and chromatin arrays [95, 97, 121]. Our results showing enhanced viral DNA integration into chromatin-bound DNA relative to the dechromatinized genomic DNA (Fig. 3.1D and E) provide biochemical evidence for the substrate preference of the HIV-1 PIC. Therefore, we sought to tease out the contribution of the nucleosome structure and host factors to integration preference.

To understand the mechanism of chromatin preference for HIV-1 DNA integration, we assembled nucleosomes with the 147-bp Widom 601 nucleosome positioning DNA sequence (Fig. 3.2A to D) [151, 152]. This approach generates compositionally uniform and precisely positioned nucleosomes as opposed to the cellular sources for histones/chromatin that contain highly diverse nucleosomes marked with an array of histone and DNA modifications [112, 204, 205]. Most importantly, the biochemically assembled nucleosome is uniquely suited to test whether the nucleosomes without any histone tail modification and the analogous naked DNA can serve as the substrates for HIV-1 DNA integration. Given the abundance of studies reporting enhanced HIV-1 integration with nucleosomes [115–117, 119, 120, 130], we anticipated that HIV-1 PIC activity

would be stimulated with nucleosomes. To our surprise, we observed higher viral DNA integration with the naked DNA rather than the nucleosomes by both HIV-1 PICs and INS, the preassembled and purified PIC substructures (Fig. 3.2E and F). Since the nucleosomes used in our assay are devoid of any histone tail modifications, our results suggest that the DNA wrapped around naked nucleosomes lacks the biochemical and/or structural determinants to support efficient HIV-1 DNA integration. Moreover, there are studies that report nucleosome dense chromatin arrays are less preferred for HIV-1 integrase activity relative to the analogous naked DNA [95, 97, 121]. Furthermore, the Widom 601 DNA used in these experiments is optimized for binding to the nucleosome with high affinity [150, 151, 156]. This could impose a strong steric hindrance on the access and insertion of the viral DNA into nucleosome-bound DNA, whereas the naked DNA substrate possesses no such biochemical or structural limitation. Even though HIV-1 DNA integration is not specific for target DNA sequence [86], future studies using nucleosomes assembled with human DNA sequences like the D02 [150] would be valuable to determine the impact of DNA-histone octamer binding on PIC- and INS-mediated integration. In addition, it is important to note that the previous studies showing enhanced integration into nucleosome-bound DNA were derived from disparate chromatinized substrates, such as chicken erythrocytes, yeast, and SV40 minichromosomes, a mouse mammary tumor virus sequence mononucleosome, and a heterogeneously positioned chimeric dinucleosome [59–62]. Moreover, the histones used in subsequent nucleosome reconstitution studies were from a cellular source (HeLa cells or chicken erythrocytes) that most likely retained the native posttranslational modifications that could influence viral DNA integration [94, 95, 120, 121, 130, 155]. Therefore, our results suggest that the DNA wrapped within the nucleosomes lacking any histone tail modifications or linker DNA sequences is not a preferred substrate for HIV-1 DNA integration.

Chromatin structure is broadly categorized into open or euchromatin and closed or heterochromatin state [149, 206]. Heterochromatin represents a highly condensed, gene-poor, and transcriptionally silent state, whereas euchromatin is less condensed, gene-rich, and more easily transcribed [207]. Multiple features of chromatin, including histone modifications, DNA methylation, and small RNAs, are involved in higher-order chromatin structure and thus facilitate or prevent access to the nucleosomal DNA [208, 209]. Notably, histone tail modifications

characteristic of open chromatin are positively correlated with HIV-1 DNA integration preference [89, 92, 93, 96, 98, 100, 101, 160, 167]. Conversely, histone modifications that define DNA compaction into a heterochromatin state are negatively associated with HIV-1 integration. In particular, the H3K36me3 modification is positively correlated with HIV-1 DNA integration across various published data sets [89, 93, 96, 98, 101, 102, 160, 210]. Although the underlying mechanism by which H3K36me3 affects HIV-1 integration targeting is unclear, H3K36me3 is an abundant and highly conserved chromatin modification within the body of transcriptionally active genes [180–182]. H3K36me3 also plays critical roles in the regulation of transcription elongation, DNA repair, the prevention of cryptic start sites, pre-mRNA splicing, and processing [183–196]. Notably, H3K36me3 may play a role in tethering the PIC to the actively transcribing genes to facilitate HIV-1 DNA integration through the host factor LEDGF/p75 [98, 101, 122, 178, 186, 187]. Genetic domain swamping and pharmacological inhibition of LEDGF/p75 binding to the HIV-1 IN results in a shift in integration sites away from the H3K36me3 mark [187, 188, 211, 212]. Accordingly, we observed higher PIC-mediated DNA integration into the nucleosomes containing H3K36Cme3 compared to the unmodified nucleosomes (Fig. 3.3C and 4F; see also Fig. S3.3C and S4D). The level of viral DNA integration into the H3K36Cme3 nucleosome (147 bp) was lower compared to the naked DNA (Fig. 3.3C; see also Fig. S3.3B). Even though our results cannot fully explain the exact mechanism, the nucleosomes used in our study contain a single histone modification in contrast to the nucleosomes within native chromosomes (or natively sourced histones) that are decorated with epigenetic histone and DNA modifications [94, 95, 97, 115–118, 120, 121, 130, 155]. Nevertheless, the presence of a linker DNA optimally stimulated PIC-mediated DNA integration into the H3K36Cme3-modified nucleosomes (Fig. 3.4F; see also Fig. S3.4E). The preference for the H3K36Cme3-modified nucleosomes with a linker DNA by the PICs was not observed with the INS (Fig. 3.3D and Fig. 3.4D). These data are consistent with other biochemical studies that reported a negligible effect by histone methylation marks associated with active transcription on HIV-1 IN or preassembled INS activity [122, 124, 130]. Notably, these results strongly suggest that the recognition of the H3K36me3 mark is specific to the PIC and is probably mediated by a viral and/or host factor(s) other than the IN enzyme alone. The HIV-1 PIC preference for integration into the linker H3K36me3-modified nucleosomes also

aligns with studies that demonstrated HIV-1 IN preferentially targets regions within chromatin arrays containing less nucleosome density and consequently longer linker DNA [95, 121]. Structural studies of LEDGF/p75 with H3K36Cme3 nucleosomes revealed that LEDGF/p75 bound nucleosomes containing both linker DNA and the H3K36me3 mark more efficiently than nucleosomes lacking linkers [178]. Consequently, the H3K36me3-modified nucleosome with linker DNA may serve as a better binding partner for LEDGF/p75, which, in turn, supports greater quantities of PIC-mediated integration (see Fig. S3.4D and E). Taken together, our results suggest that the H3K36me3 mark in the nucleosomes with a linker DNA contains the biochemical and molecular determinants for efficient HIV-1 DNA integration.

During HIV-1 infection, a number of host factors are utilized to complete the early and late stages of the viral replication cycle [157]. HIV-1 DNA integration into the host chromosomes by the PIC is dependent on host factors such as LEDGF/p75, INI1, CPSF6, and others [167–169, 214]. The most studied PIC-associated host factor is LEDGF/p75, which engages with HIV-1 IN via the IN binding domain (IBD) [135, 172, 174]. Concomitantly, LEDGF/p75 interacts with the H3K36me2/3 marked chromatin, and other methylated histones, through the PWWP domain and AT-hook motifs [178, 179, 215]. It has been reported that LEDGF/p75 facilitates HIV-1 DNA integration by tethering the PIC to actively transcribing genes [120, 133, 135, 139, 174, 216]. Therefore, we hypothesized that the combination of LEDGF/p75 and the H3K36me3 would yield an additive effect on HIV-1 DNA integration. As expected, we observed a significant increase in HIV-1 DNA integration into both the nucleosome and the analogous naked DNA with the INS in the presence of LEDGF/p75 (Fig. 3.5A and B). Surprisingly, LEDGF/p75 significantly reduced PIC-mediated integration into the nucleosomes and had a minimal effect on integration into the naked DNA (Fig. 3.5C and D). Intriguingly, Lapaillerie et al. recently reported that LEDGF/p75 preincubation with a nucleosome target DNA resulted in reduced viral DNA integration with preassembled HIV-1 INS [122]. A lack of such block to the INS-associated viral DNA points to the involvement of additional factors and/or mechanisms specific to the PIC. In addition, IN activity was disrupted when the LEDGF/p75 amounts exceeded a >2-fold amount of IN [175]. Since, LEDGF/p75 is present in the extracted PICs (Fig. 3.5E) and the addition of exogenous LEDGF/p75 blocked access to the nucleosome-bound DNA (Fig. 3.5G and H), we predict that

excessive LEDGF/p75 can sterically hinder access of the PIC-associated viral DNA to the nucleosome-bound target DNA. We acknowledge that the stoichiometry of PIC-associated IN and LEDGE/p75 in our PIC extracts is unknown. Nevertheless, excessive LEDGF/p75, overexpression of the LEDGF/p75-IBD domain, and certain LEDGF-derived peptides are reported to inhibit HIV-1 integration [134, 135, 175, 217–219]. Notably, the LEDGF/p75 inhibition of PIC-mediated integration was relieved with the H3K36me3-modified nucleosomes and allowed recovery of the PIC-mediated integration (Fig. 3.6A to C). The presence of linker DNA also relieved the LEDGF/p75 mediated reduction of PIC-mediated integration (Fig. 3.6D). Collectively, these data implicate that structural and/or conformational features of nucleosomes rendered by linker DNA and the H3K36me3 modification are essential for LEDGF/p75 to promote of PIC-mediated viral DNA integration.

Our biochemical data show that the H3K36me3-modified nucleosome containing linker DNA is an ideal substrate for PIC-mediated integration. However, our in vitro integration assay is not designed to identify the targets of integration into specific regions of the nucleosome. The nucleosome core particle-bound DNA is highly contorted and bent into ~1.75 left-handed superhelical turns of ~80 bp/turn [106, 107]. This bending is a consequence of the minor and major groove structures of the DNA where the minor grooves are preferentially oriented toward the histone octamer surface [106, 151, 220, 221]. There is evidence that the outward-facing DNA integration occurs at the sites of the major grooves of the target DNA within the nucleosome. Even though, retroviral INs lack target DNA sequence specificity for integration [222, 223], studies of mononucleosome particles [117] or minichromosomes [115, 116] have shown that the DNA wrapped around the nucleosome is preferentially targeted compared to nucleosome-free DNA. Therefore, to better understand the consequence of the H3K36me3-modified nucleosomes on viral DNA integration targeting, we carried out deep sequencing of the viral DNA integrants from PIC assays with the linker naked DNA, linker nucleosome, and the linker H3K36me3-modified nucleosome. The PIC-mediated viral DNA integration that was observed within the linker-DNA preferably occurred throughout the length of the DNA sequence, and primarily within the linker DNA sequence-positions from bp 1 to 50 and from bp 197 to 247 (Fig. 3.7A; see also Fig. S3.7 and

S3.8 in the supplemental material). Similarly, integration within the linker unmodified nucleosomes occurred most frequently in the linker DNA and a slight uptick in the integration sites throughout the nucleosome core positioning sequence, bp 50 to 197 (Fig. 3.7B; see also Fig. S3.7 and S3.9). Remarkably, the integration into the nucleosome core DNA increased within the linker H3K36me3-modified nucleosomes. Both the integration count and the frequency occurred overwhelmingly in the nucleosome core sequence of the H3K36me3-modified nucleosomes with linker DNA (Fig. 3.7C; see also Fig. S3.7 and S3.10). It should be noted that these integration site analyses were from sequencing reads of <500 due to the technical challenges of shorter DNA substrates. Therefore, we are cautious about the statistical and physiological significance of these observations. Nevertheless, qualitative analysis of the integration sites within the linker H3K36me3 nucleosomes indicates that the integrants abutted within the superhelical locations (SHL) −1, −7, +7, and +1, locations that correspond to the location where the H3-tail protrudes through the nucleosomal DNA [111, 161, 178]. This striking observation suggests that the H3K36me3 mark can perform a functional role in directing the HIV-1 PIC-associated DNA into specific regions of the nucleosomal DNA.

Our results provided biochemical evidence that the nucleosome structure characteristic of open chromatin enhances HIV-1 DNA integration. Open chromatin represents a minority of the human genomic DNA, whereas closed chromatin or heterochromatin represents up to 80% of the genomic DNA [206, 224]. However, it is unclear why HIV-1 integration is disfavored in these vast DNA regions of the human chromosomes [89, 96, 98, 100, 159]. To interrogate the role of closed chromatin on HIV-1 integration, we probed the effects of the linker histone H1° due to its relatively high abundance among H1 variants [225] and well-characterized ability to condense chromatin fibers and form rigid, stable chromatosome structures [108, 190]. As expected, H1° addition protected isolated chromatin from the partial micrococcal nuclease digestion (see Fig. S3.13A to D) and, at nonsaturating amounts, inhibited HIV-1 INS-mediated integration (Fig. 3.8A to C). This is consistent with the published studies with HIV-1 IN and H1-mediated condensation of recombinant nucleosome arrays [94, 189, 226]. Surprisingly, nonsaturating amounts of H1° failed to inhibit PIC-mediated integration into the linker substrates (Fig. 3.8D to F; see also Fig. S3.11A to C). However, at stoichiometric saturating levels relative to the DNA

substrate, PIC-mediated integration was significantly reduced with all the linker substrates (Fig. 3.8G to I; see also Fig. S3.11D to F). Interestingly, H1 similarly reduced PIC-mediated integration into both genomic DNA and chromatin (Fig. 3.8J to K; see also Fig. S3.13E and F). These observations suggest that mechanisms other than nucleosome structure could be involved in PIC-mediated integration inhibition. While H1 specifically binds nucleosomes at the dyad while interacting with linker DNAs, H1 has also the ability to bind to naked DNA substrates [227–229]. In addition, our data indicated that preincubation with the target DNA is requisite for the H1-mediated inhibition of PIC-mediated integration (see Fig. S3.13G and H). Finally, in line with our in vitro data, proviral DNA integration was reduced by almost 50% in H1° O/E cells (Fig. 3.9B), suggesting that the histone H1 can negatively regulate HIV-1 DNA integration through chromatin compaction. Studies interrogating the consequence of excess H1 and salt-mediated compaction on HIV-1 integration, as well as the subsequent integration site selection, will further enhance our knowledge of the mechanisms of reduced retroviral DNA integration into compacted chromatin.

Collectively, our study provides biochemical evidence and mechanistic insights into HIV-1 integration preference into open chromatin. Particularly, these results reveal a direct role of the H3K36me3 epigenetic mark in HIV-1 integration preference and identify an optimal substrate for HIV-1 PIC-mediated viral DNA integration. Finally, our study highlights the utility of PIC studies, since viral DNA integration by the PIC utilizes distinct mechanism compared to the biochemically assembled INS.

3.5 Materials and Methods

Cell culture.

HEK293T and SupT1 cell lines were obtained from the American Type Culture Collection (Manassas, VA). The TZM-bl reporter cell line (catalog no. ARP-8129) was obtained through the NIH HIV Reagent Program, Division of AIDS, NIAID, NIH. TZM-bl cells were contributed by John C. Kappes, Xiaoyun Wu, and Tranzyme, Inc. [230]. HEK293T and TZM-bl cells were

cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 2 mM glutamine, 1,000 U/mL penicillin, and 100 mg/mL streptomycin. SupT1 cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 2 mM glutamine, 1,000 U/mL penicillin, and 100 mg/mL streptomycin. All cells were cultured at 37°C with 5% CO2.

Virus stocks and other reagents.

High titer virus stocks were generated by transient transfection of HEK293T cells with the HIV-1 plasmid construct pNL4-3 (NIH AIDS Reagents) with polyethyleneimine (PEI) as per our published protocol [143]. Briefly, 3×10^6 cells were seeded per 10-cm culture dish and cultured overnight. The following day, cells in each dish were transfected using PEI and 10 to 15 μg of pNL4-3 DNA. At 16 h posttransfection, the cells were washed with phosphate-buffered saline (1× phosphate-buffered saline [PBS]) and replenished with 6 mL of DMEM. After 36 h, the virus-containing culture supernatant of transfected cells was harvested, cleared of debris by low-speed centrifugation, filtered through 0.45-μm filters, and treated with DNase I (Calbiochem; 20 μg/mL of supernatant) in the presence of 10 mM magnesium chloride (MgCl2) for 1 h at 37°C. The virus infectivity was determined by using TZM-bl indicator cells, as previously described [231]. Raltegravir (RAL) (Isentress, MK-0518), ARP-11680, was obtained from the NIH HIV reagent program, Division of AIDS, NIAID, NIH, and was contributed by Merck & Company, Inc.

Chromatin isolation.

The chromatin isolation protocol is adapted from a published protocol [148]. HEK293T cells were cultured to near confluence, harvested, and then washed once with ice cold 1× PBS. The cells were manually lysed on ice with a Dounce homogenizer (Wheaton) in native lysis buffer (NLB; 5 mM MgCl2, 10 mM KCl, 20 mM HEPES [pH 7.5], 250 mM sucrose, 0.5 mM dithiothreitol [DTT], 0.5 mM phenylmethanesulfonyl fluoride [PMSF]) for 20 to 30 strokes. The lysate was washed and homogenized with NLB twice by centrifugation at 3,000 × g for 15 min at 4°C. After manual lysis

with NLB, the pellet was washed with modified buffer B (MBB; 5 mM MgCl2, 10 mM KCl, 20 mM HEPES [pH 7.5], 0.2 M EGTA, 0.5 mM DTT, 0.5 mM PMSF, protease inhibitor cocktail [Promega, Madison, WI]). The resulting pellet was resuspended in 2 volumes of MBB, in reference to the pellet volume, and then an equivalent volume of MBB–0.6 M KCl–10% glycerol was added in a dropwise manner. The lysate was incubated at 4°C for 10 min with consistent rocking. After incubation, the nuclear fraction was pelleted at a precooled (4°C) centrifuge at maximum speed. The pelleted nuclear fraction was resuspended in 20 times volume of medium salt buffer (MSB; 20 mM HEPES [pH 7.5], 0.4 M NaCl, 1 mM EDTA, 5% glycerol [vol/vol], 0.5 mM DTT, 0.5 mM PMSF) and manually homogenized. The nuclear pellet was centrifuged for 10 min at 11,000 × g at 4°C and then resuspended in 4 pellet volumes of high salt buffer (HSB; 20 mM HEPES [pH 7.5], 0.65 M NaCl, 1 mM EDTA, 0.34 M sucrose, 0.5 mM DTT, 0.5 mM PMSF). Oligonucleosomes were released from the nuclear debris by manual homogenization on ice and fully separated from the debris by maximum-speed centrifugation at 4°C to collect the chromatin-containing supernatant. The chromatin was dialyzed for 16 h against low-salt buffer (LSB; 20 mM HEPES [pH 7.5], 0.1 M NaCl, 1 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF). The dialyzed chromatin was tested for nucleosome protection by using partial micrococcal nuclease digestion, and, subsequently the presence of histone proteins was tested via Coomassie staining and Western blotting.

Immunoblotting.

To detect proteins by immunoblot, samples were prepared in radioimmunoprecipitation assay (RIPA) buffer (Santa Cruz Biotechnologies, Inc., Dallas, TX) or NLB buffer supplemented with protease inhibitor cocktail and 10 μg/μL PMSF (Sigma-Aldrich, Saint Louis, MO) according to the standardized protocol. Protein concentrations were determined using BCA (bicinchoninic acid) protein assay reagent (Thermo-Fisher Scientific, Waltham, MA) according to the manufacturer's specifications. Equivalent amounts of protein from cellular lysates or fractions were electrophoresed on precast gels (Bio-Rad Laboratories, Hercules, CA) or in-house prepared 15%

polyacrylamide gels and electrophoretically transferred to nitrocellulose membranes using Trans-Blot SD semidry transfer cells (Bio-Rad Laboratories). The membranes were incubated in blocking buffer (5% [wt/vol] nonfat milk in Tris-buffered saline containing 0.05% Tween 20 [TBST]; pH 8.0). After blocking, the membranes were then probed with the following indicated antibodies diluted in blocking buffer: H2A (ab18255), H2B (ab1790), H3 (ab1791), H4 (ab7311), and H1° (ab218417) (all from Abcam, Boston, MA); H3K36me3 (Active Motif, Carlsbad, CA; ab_2615073); LEDGF/P75 (Bethyl Laboratories, Inc., Montgomery, TX; A300-848A); p53 (Cell Signaling Technology, 2527S); and GAPDH (60004-1; Proteintech, Rosemont, IL) and subsequently with secondary antibody conjugated to horseradish peroxidase (anti-rabbit 1:10,000 [vol/vol]; anti-mouse 1:10,000 [vol/vol]). The membranes were washed with 1× TBST buffer at least three times for 15 min for each wash, and immunocomplexes were detected by a clarity enhanced chemiluminescence method (Bio-Rad Laboratories).

Isolation of PICs.

HIV-1 PICs were isolated from HIV-1-infected T cells, as described in our published methods [143–147]. Briefly, 8 × 107 of SupT1 cells were spinoculated (480 × g) with DNase I-treated wild-type virions for 2 h at 25°C and then cultured for 5 h at 37°C. The infected cells were then harvested by centrifugation (300 × g) for 10 min at room temperature. The cell pellet was washed twice with 2 mL of buffer K−/− (20 mM HEPES [pH 7.5], 150 mM potassium chloride [KCl], 5 mM MgCl2). Subsequently, the cell pellet was gently resuspended in 2 mL of ice-cold buffer K+/+ (20 mM HEPES [pH 7.5], 150 mM KCl, 5 mM MgCl2, 1 mM DTT, 20 mg/mL aprotinin, 0.025% [wt/vol] digitonin). The cell suspension was transferred to a chilled 2-mL microcentrifuge tube and incubated on a rocking platform (60 to 80 rocking motions/min) for 10 min at room temperature. The cell lysate was centrifuged (1,500 × g) for 4 min at 4°C to separate the cytoplasmic and nuclear fractions. The supernatant (cytoplasmic fraction) was transferred to a fresh 2-mL microcentrifuge tube and centrifuged again (16,000 × g) for 1 min at 4°C to clear residual nuclear debris. The resulting cytoplasmic fraction was treated with RNase A (Invitrogen,

Waltham, MA) at a final concentration of 20 mg/mL and then incubated for 30 min at room temperature. Finally, 60% (wt/vol) sucrose was added to a final concentration of 7%, and the contents were thoroughly mixed. Aliquots of the cytoplasmic fraction were flash-frozen in liquid nitrogen and stored at −80°C for long-term storage. The cytoplasmic fraction aliquots were used as the source of HIV-1 PICs in assays.

Assay for measuring PIC-mediated viral DNA integration.

In vitro integration assays were performed using a modified version of a protocol from the Chandravanu laboratory [143, 145–147]. The in vitro integration reaction was carried out by mixing 50 µL of PIC with 300 ng of the indicated target DNA and then allowing the mixture to incubate at 37°C for 45 min. The integration reaction was stopped by adding SDS, EDTA, and proteinase K to final concentrations of 0.5%, 8 mM, and 0.5 mg/mL, respectively, followed by overnight incubation at 55°C. The deproteinized reaction was brought to 200 µL, mixed with an equal volume of phenol (pH 8.0), thoroughly mixed by vortexing, and then centrifuged (17,000 × g) for 2 min at room temperature. The aqueous phase is extracted once with an equal volume of phenol-chloroform (1:1) mixture, followed by an equal volume of chloroform. The DNA was precipitated by adding 2.5 volumes of 100% ice-cold ethanol in the presence of sodium acetate (0.3 M, final concentration) and the coprecipitate glycogen (25 to 100 µg, final concentration), followed by a minimum incubation for 2 h at −80°C. The sample was centrifuged at maximum speed for 30 min at 4°C, and the resultant DNA pellet was washed once with 70% ethanol using maximum-speed centrifugation for 15 min at 4°C. The precipitated DNA was air-dried at room temperature, resuspended in 50 µL of nuclease-free water, and used as the template DNA for the nested qPCR to measure PIC-mediated viral DNA integration.

The nested PCR consists of two rounds of PCR to amplify the junction between the viral DNA and target DNA, followed by quantitative PCR (qPCR) specific to the viral DNA. A first-round standard PCR, designed to amplify only the integrated virus-target DNA junctions, was performed in a final volume of 50 µL containing 5 µL of purified DNA product from the

integration reaction, 500 nM concentrations of each primer against the indicated target DNA, and the viral LTR (5′-GTGCGCGCTTCAGCAAG-3′), 1× Bestaq PCR buffer (Applied Biological Materials, Inc., Richmond, BC, Canada), a dNTP nucleotide mix containing 200 mM concentrations of each nucleotide (Promega, Madison, WI), and 1.25 U of Bestaq DNA polymerase (Applied Biological Materials, Inc.) with the following thermocycling conditions: 95°C for 5 min, followed by 25 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s (for the Widom 601 DNA) or 2 min (for chromatin or genomic DNA) and a final extension at 72°C for 10 min. The second-round qPCR designed to amplify the viral LTR-specific region contained one-tenth (5 μL) the volume from the first-round PCR product as the template DNA, 1× iTaq Universal Probe Supermix (Bio-Rad Laboratories), 300 nM concentrations each of the viral LTR-specific primers that target the R region (5′-TCTGGCTAACTAGGGAACCCA-3′) and the U region (5′-CTGACTAAAAGGGTCTGAGG-3′), and a 100 nM concentration of TaqMan probe (5′-6-carboxyfluorescein [FAM]-TCAGCATTATCAGAAGGAGCCACC-6-carboxytetramethylrhodamine [TAMRA]-3′). The qPCR run included an initial incubation at 95°C for 3 min, followed by 39 cycles of amplification and acquisition at 94°C for 15 s and 58°C for 30 s and a final incubation at 72°C for 30 s. During qPCR analyses of the samples, a standard curve was generated under the same conditions in parallel using 10-fold serial dilutions of known copy numbers (100 to 108) of the HIV-1 molecular clone plasmid. Data were analyzed using CFX Manager software (Bio-Rad Laboratories), and integrated viral DNA copy numbers were determined by plotting the qPCR data against the standard curve.

Nucleosome positioning sequence DNA preparation.

The DNA molecules used for nucleosome assembly were generated by PCR from a pUC19 clone containing the Widom 601 nucleosome positioning sequence. Labeled DNA molecules were produced using PCR with fluorescein- or Cy3-labeled oligonucleotides (primers) amplified from a plasmid containing the Widom 601 sequence or a plasmid containing the Widom 601 sequence with a GAL4 binding site at bases 6 to 25 as previously reported [161, 162, 165]. DNA products from a 48-well PCR plate (50 to 100 μL reaction per well) were pooled and injected into an anion

exchange column (MonoQ-GL; GE Healthcare Life Sciences, Marlborough, MA) for purification. After at least 5 column volumes were washed with 0.5× Tris-EDTA (TE) buffer, the DNA was eluted with a 0 to 1.0 M NaCl linear gradient in 0.5× TE buffer (pH 7.5). The fractions containing the PCR product were pooled and concentrated by a minimum cutoff 10 MWCO spin column (Amico Ultra; Millipore Sigma, Burlington, MA) in 0.5× TE.

Preparation of histone octamers.

Human recombinant histones were expressed and purified as previously described [152]. Mutation H3 (C110A) was introduced by site-directed mutagenesis (Agilent). Human H2A, H2A K119C, and H2B were expressed in Rosetta BL21(DE3)/pLysS cells, while human H3 C110A and H4 were expressed in BL21(DE3)/pLysS cells. Histones were purified separately under denaturing conditions using size exclusion and cation exchange chromatography. The histone octamer was also prepared as previously described [152]. Lyophilized histones were resuspended in unfolding buffer (20 mM Tris-HCl [pH 7.5], 7 M guanidinium, 10 mM DTT) at 5 mg/mL and mixed together at a ratio of 1.2:1 [(H2A and H2B)/(H3 and H4)]. The histone octamer was formed by performing double dialysis with the mixture into refolding buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 2 M NaCl, 5 mM BME). The octamer was removed from dialysis, and H2A K119C containing histone octamer was labeled with Cy5-maleimide (GE Healthcare) as previously described [176]. The histone octamer was refolded by adding each of the histones together at an equal molar ratio and purified with size exclusion chromatography directly after refolding dialysis.

Deposition of methyl-lysine analog on histone H3.

Human H3.2 (C110A, K36C) was created by site-directed mutagenesis using a Stratagene QuikChange XL kit. The mutant histone was then expressed in E. coli BL21(DE3)/pLysS cells and purified as previously described [152]. The methyl-lysine analog was then deposited onto H3.2(C110A, K36C) based on a previously reported protocol [162, 163]. Initially, 5 mg of the mutant histone was suspended in 980 μL of alkylation buffer (1 M HEPES [pH 7.8], 4 M

guanidine-HCl, 10 mM d/l-methionine) and allowed to unfold for 1 h. The histones were then reduced with 6.66 mM DTT for 1 h at 37°C. After reduction, 100 mg of (2-bromoethyl)trimethylammonium bromide was added to the histones. The reaction was then covered and stirred for 5 h at 50°C. The reaction was quenched with 50 μL of 14.3 M 2-mercaptoethanol. Histones were then desalted using a PD 10 desalting column and dried via vacuum concentration. Labeling efficiency was confirmed by matrix-assisted laser desorption ionization–time of flight mass spectrometry.

Nucleosome reconstitution.

Nucleosomes were prepared as previously described (152). DNA was mixed with histone octamer at a molar ratio of 1.25:1 in 0.5× TE (pH 8), 2 M NaCl, and 1 mM benzamidine hydrochloride and dialyzed via double dialysis into 0.5× TE (pH 8) and 1 mM benzamidine hydrochloride. Reconstituted nucleosomes were separated from free DNA by ultracentrifugation with a 5 to 30% (wt/vol) sucrose gradient spun at 41,000 rpm for 22 h. Gradients were fractioned and ran on a 0.3× TBE–5% acrylamide gel at 300 V for 1 h to determine nucleosome-containing fractions. Selected fractions were concentrated, the buffer was exchanged into 0.5× TE (pH 8.0) using centrifugal filters, and final concentrated nucleosomes were run on acrylamide gels as before to verify the purity. Of note, multiple bands can be seen on the gels of linker DNA containing nucleosomes due to the imperfect nucleosome positioning of the Widom 601 DNA longer than 147 bp [165].

HIV-1 intasome preparation and assembly.

The HIV-1 intasomes used in our study were prepared by using published methods [153, 154]. Briefly, a double-stranded preprocessed viral DNA substrate was used as the donor DNA (Integrated DNA Technologies, Coralville, IA). A 25-base oligonucleotide (5′-AGCGTGGGCGGGAAAATCTCTAGCA-3′) was synthesized with a complementary 27-base oligonucleotide (5′-ACTGCTAGAGATTTTCCCGCCCACGCT-3′) to provide the 3′-OH recessed

end for the preprocessed intasomes. The recombinant Sso7d-integrase (Sso7d-IN) was expressed and purified in E. coli BL21(DE3), and the cells were lysed in a buffer containing 20 mM HEPES (pH 7.5), 10% glycerol, 2 mM β-mercaptoethanol (β-ME), 20 mM imidazole, and 1 M NaCl. The protein was purified by nickel affinity chromatography and subsequently by gel filtration on a Hi-Load 26/60 Superdex-200 column (GE Healthcare, Milwaukee, WI) equilibrated with 20 mM HEPES (pH 7.5), 10% glycerol, 2 mM DTT, and 500 mM NaCl. The protein was concentrated using an Amicon centrifugal concentrator (EMD Millipore, Burlington, MA) and then used for intasome assembly or flash-frozen in liquid nitrogen and stored at −80°C. Intasomes were assembled by incubating 3 mM HIV-1 integrase and 1 mM 3′-processed viral DNA substrate in 20 mM HEPES (pH 7.5), 20% glycerol, 5 mM 2-mercaptoethanol, 5 mM CaCl2, 2 mM ZnCl2, 100 mM NaCl, and 50 mM 3-(benzyldimethylammonio)propane sulfonate (NDSB-256) at 30°C for 1 h. Assembled intasomes were purified as described previously [153, 154], aliquoted, flash frozen in liquid nitrogen, and stored at −80°C.

Assay to measure HIV-1 INS-mediated viral DNA integration.

Integration assays to measure INS-associated viral DNA integration were assembled on ice with 25 nM INS in 20 mM HEPES (pH 7.5), 20% glycerol, 10 mM DTT, 5 mM MgCl2, 4 mM ZnCl2, and 100 mM NaCl in a 20-μL reaction volume. In vitro integration reactions with INS were carried out with 300 ng of target DNA, followed by incubation at 37°C for 1 h. The reaction was terminated with 0.5% SDS and 8 mM EDTA, together with 0.5 mg/mL protease K, and deproteinated for 1 h at 55°C. The DNA was recovered by ethanol precipitation, as described above, and resuspended in the original reaction volume of nuclease-free dH2O. The recovered DNA was used as the template DNA for an SYBR green-based quantitative PCR to amplify the junctions between the viral DNA substrate and the target DNA for the indicated assay. The qPCR was carried out in a 20-μL volume containing 1 to 2 μL (5 to 10 ng) of purified DNA product from the INS-mediated integration assay, a 300 nM concentration of each primer targeting the viral

DNA substrate (5′-AGCGTGGGCGGGAAAATCTC-3′) and the indicated target DNA, and 1× iTaq Universal SYBR green Supermix (Bio-Rad Laboratories). The qPCR cycling conditions for quantifying the INS activity included an initial incubation at 95°C for 3 min, followed by 39 cycles of amplification and acquisition at 94°C for 15 s, 55°C for 30 s, and 72°C for 30 s. The thermal profile for melting-curve analysis was obtained by holding the reaction at 65°C for 30 s, followed by a linear ramp in temperature from 65 to 95°C at a ramp rate of 0.5°C/s and acquisition at 0.5°C intervals. The Maestro CFX program was used for data analysis.

Fluorescence polarization.

Fluorescence polarization measurements were acquired with a Tecan infinite M1000Pro plate reader by exciting samples at 470 nm and then measuring polarized emission at 519 nm with 5 nm excitation and emission bandwidths. The fluorescence polarization (P) of each sample was calculated as follows:

$$P = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}}$$

where $I_{||}$ and $I_{\perp}$ are the emission polarized parallel and perpendicular to the polarized excitation light, respectively. The data were then fit to a binding isotherm using GraphPad 9.1.2 to determine the $S_{1/2}$ values.

Titrations used a 30-μL reaction volume with 5 nM nucleosomes, assembled with fluorescein-labeled Widom 601 DNA and mixed into increasing concentrations of LEDGF. The conditions for each reaction were 75 mM NaCl, 15 mM Tris-HCl (pH 7.5), 0.00625% Tween 20, and 3 mM DTT. The samples were loaded into a Corning round-bottom polystyrene plate and allowed to incubate at 4°C for 30 min before measurements were taken. Titrations were repeated in triplicate to acquire standard deviation errors.

FRET titrations.

FRET efficiency measurements were carried out using a Horiba Scientific Fluoromax 4 spectrofluorometer. Samples were excited at 510 and 610 nm, and the photoluminescence spectra were measured from 530 to 750 nm and from 630 to 750 nm for donor and acceptor excitations, respectively. FRET efficiencies from measured spectra were computed through the (ratio) A method [176] using an in-house Matlab program. FRET efficiencies were then normalized as changes in FRET and fit to a binding isotherm using GraphPad 9.1.2 to determine $S_{1/2}$ values, the concentrations at which 50% of the nucleosome is bound by LEDGF/p75. Titrations used a 60-µL reaction volume with 1 nM nucleosomes mixed into increasing concentrations of GAL4. Each titration was repeated in the presence or absence of binding saturation concentrations of LEDGF (500 nM). The conditions in each reaction were 75 mM NaCl, 10 mM Tris-HCl (pH 8), 0.25% Tween 20, and 10% glycerol. Titrations were repeated in triplicate to acquire standard deviation errors.

Gal4 expression and purification.

The Gal4 DNA binding domain (amino acids 1 to 147) was expressed in Rosetta BL21(DE3)/pLysS cells as previously reported [161, 162, 165]. Cells were grown in 2×YT and induced at optical density at 600 nm of 0.5 with 1 mM IPTG and 100 mM zinc acetate. After 3 h of expression, cells were spun at 4,000 × g for 15 min, the supernatant was removed, and cell pellets were frozen using liquid nitrogen. Cells were thawed and resuspended in 30 mL of buffer A (50 Mm Tris-HCl [pH 8], 200 mM NaCl, 10 mM imidazole, 10 mM BME, 20 mM zinc acetate, 1 mM DTT, 1 mM PMSF) with leupeptin and pepstatin added at 20 µg/mL final concentration. Resuspended cells were sonicated and spun at 23,000 × g for 15 min, and the lysate containing the protein was poured off. The lysate was added to a Ni-NTA column, washed with buffer A, and the protein was eluted with buffer B (50 mM Tris-HCl [pH 7.5], 200 mM NaCl, 200 mM imidazole, 20 mM zinc acetate, 1 mM DTT, 1 mM PMSF, 0.2% Tween 20). Fractions were run on a 12% acrylamide SDS gel, pooled, and dialyzed into buffer C with 200 Mm NaCl (25 mM Tris-HCl [pH

7.5], 200 mM NaCl, 20 mM zinc acetate, 1 mM DTT, 1 mM PMSF). After dialysis, the protein was further purified using cation exchange chromatography with a gradient from 200 to 600 mM NaCl using buffer C with appropriate salt concentrations. Protein-containing fractions were pooled, concentrated, exchanged into buffer D (10 mM HEPES [pH 7.5], 200 mM NaCl, 10% glycerol, 20 mM zinc acetate, 1 mM DTT, 1 mM PMSF), and then flash frozen for storage at −80°C.

H1.0 overexpression and proviral integration assay.

The H1.0 expression construct [H1.0-HA-pEV833(GFP)] [225] was generously shared by Albert Jordan (Institute of Molecular Biology of Barcelona). HEK293T cells (3.0 × 105 cells per well) were cultured in 6-well plates and transfected with 3 μg of the H1.0 expression construct for 2 to 4 h, and then the cells were washed with fresh media and cultured for another 12 to 24 h. Prior to infection, virus stocks were prepared in DMEM containing Polybrene (6 μg/mL) and either 1 μM RAL or 5 μM EFV were added to the cells where indicated. Infection experiments were carried out by inoculating the H1.0 overexpression and wild-type cells with 1,500 ng (p24) of VSVG-ΔENV(GFP) HIV-1 particles. The infected cells were cultured overnight and then washed with fresh medium the following day. The cells were then cultured for 24 to 48 h at 37°C with 5% CO_2 and harvested for further use. Total DNA from the infected and uninfected control cells was isolated using a Quick-DNA miniprep kit according to the manufacturer-recommended protocol (Zymo Research, Irvine, CA). To measure HIV-1 PVI, a nested PCR method was used that consisted of a first-round endpoint PCR with primers designed to amplify only the integration junctions between human Alu repeats and HIV-1 viral DNA, followed by a second round of qPCR with primers designed to specifically amplify only the viral LTR from the first-round PCR products (143, 146, 147, 192). The first-round PCR contained 100 ng of total DNA, 1× Bestaq PCR buffer (Applied Biological Materials Inc.), a dNTP mix containing 200 mM concentrations of each nucleotide (Promega, Madison, WI), 500 nM primers targeting the Alu repeat sequence (5′-GCCTCCCAAAGTGCTGGGATTACAG-3′) and the HIV-1 Gag sequence (5′-

GTTCCTGCTATGTCACTTCC-3′), and 1.25 U of Bestaq DNA polymerase in a 50-μL final volume. The first-round PCR conditions consisted of the following cycles: 95°C for 5 min, followed by 25 cycles of amplification at 94°C for 30 s, 50°C for 30 s, and 72°C for 4 min and then a final incubation at 72°C for 10 min. The second-round PCR consisted of one-tenth (5 μL) of the product from the first-round PCR as the DNA template, 1× iTaq universal probe Supermix (Bio-Laboratories), 300 nM concentrations (each) the viral Gag-specific primers (F2, 5′-TCAGCCCAGAAGTAATAC-3′) and (R2, 5′-CACTGGATGCAATCTATC-3′), and a 100 nM concentration of TaqMan probe (5′-6-[FAM]-TCAGCATTATCAGAAGGAGCCACC-6-[TAMRA]-3′). The qPCR program consisted of the following amplification conditions: 95°C for 3 min, followed by 39 cycles of amplification and acquisition at 94°C for 15 s, 58°C for 30 s, and 72°C for 30 s. Concurrent with the qPCR of the samples, a standard curve was generated in parallel and under the same conditions using a 10-fold serial dilution of known copy numbers (1 × 100 to 1 × 108) of the HIV-1 molecular clone plasmid (NL4-3). The qPCR experiments were performed in triplicates, and the data were analyzed using CFX Maestro software (Bio-Rad Laboratories). The integrated viral DNA copy numbers were calculated by plotting the qPCR data against the standard curve.

Sequencing analysis of the integration junctions.

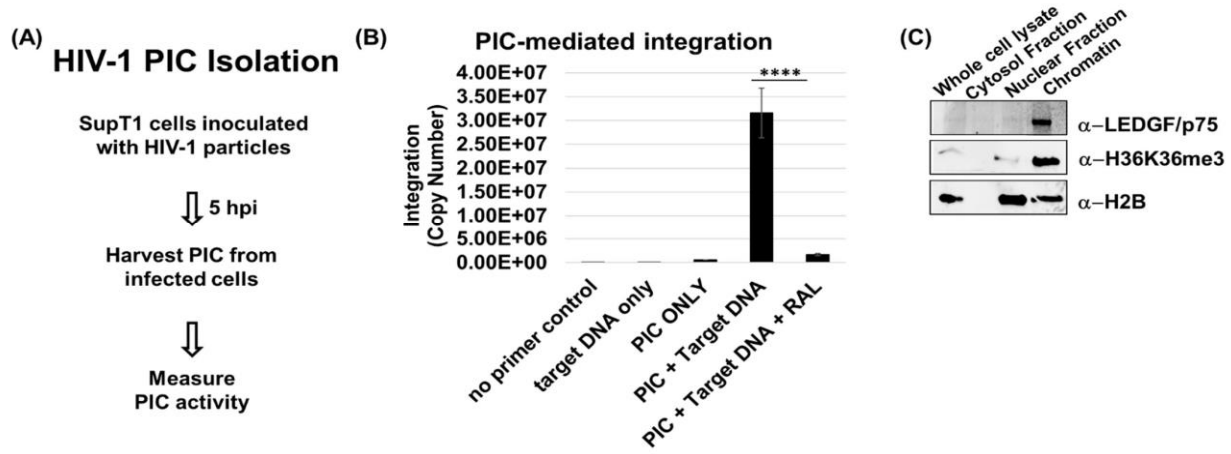
The first-round endpoint PCR product of the PIC-mediated integration assays was subjected to a second endpoint PCR amplification with the same primers to increase the yield of DNA. The second endpoint PCR amplification was subjected to a PCR DNA purification kit (Zymo Research), and the DNA was submitted for sequencing as described by Christensen et al. and performed at the DNA Sequencing Core Facility, University of Utah [155]. In brief, 100-ng samples were prepared using an Illumina TruSeq DNA PCR-Free Library Prep, fragmented to a mean size of 200 to 250 bp, and sequenced on a NovaSeq reagent kit v1.5_150 × 150 flow cell, with a target read depth of 5 mega (M) reads per sample. Due to the relatively short target DNA length (247 bp) and the viral DNA primer binding 84 bp upstream of the LTR (~720 bp into the HIV-1 genome), generating a maximum integration junction of <1 kb, sequencing the integration

junctions presented inherent technical limitations. This sequencing protocol was optimized for samples >1 kb; thus, there were technical challenges for generating the expected thousands to hundreds of thousands of reads with our samples. As a result, we selected the samples that contained the most contiguous integration junction reads for analysis. Therefore, two samples per naked DNA with linker, unmodified nucleosome with linker, and H3K36Cme3 nucleosome with linker, which met these criteria, were analyzed. Next, fastq files were then filtered using SEAL to find reads that contained a 20-base kmer that mapped to the end of the 5′ LTR with at least 17/20 matches. The reads passing this filter were then mapped onto the nucleosome target sequences to generate SAM and BAM files. The Integrated Genomics Viewer (IGV; https://software.broadinstitute.org/software/igv/) was used to visualize the integration junctions [233]. The integration junctions were tabulated by quantifying the continuous junction between the target DNA (in gray) and the viral DNA sequence (red, green, yellow, or blue) and then plotted in Excel.

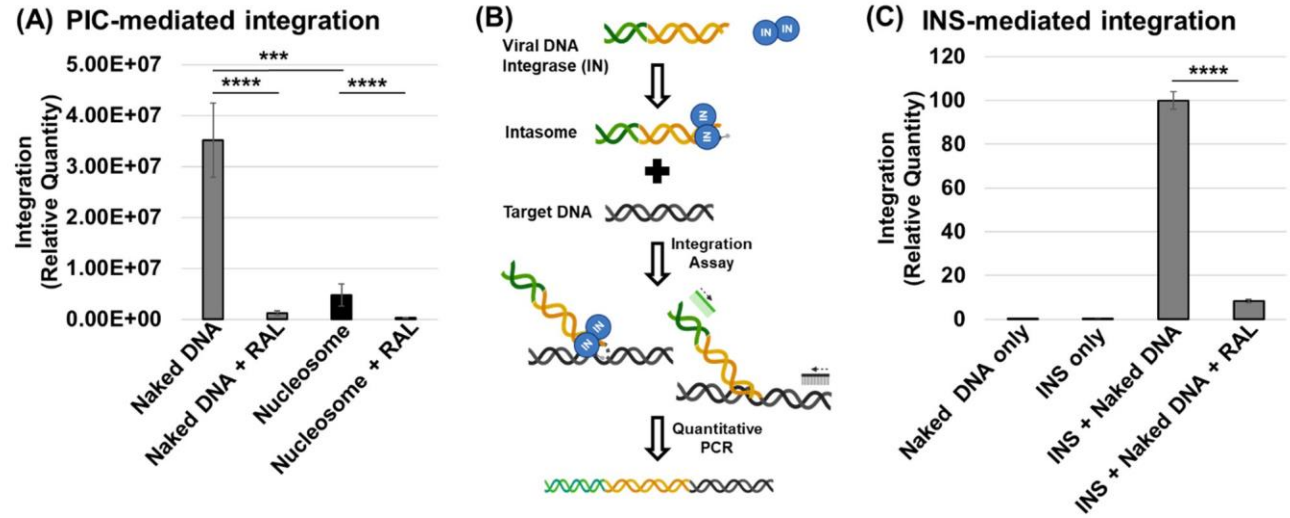
Statistical analysis.

Statistical significance was determined by using a two-sided unpaired Student t test with Prism software (GraphPad; *, $P < 0.05$; **, $P = 0.01$ to 0.05; ***, $P = 0.01$ to 0.001; ****, $P = 0.001$ to 0.0001; *****, $P < 0.0001$).

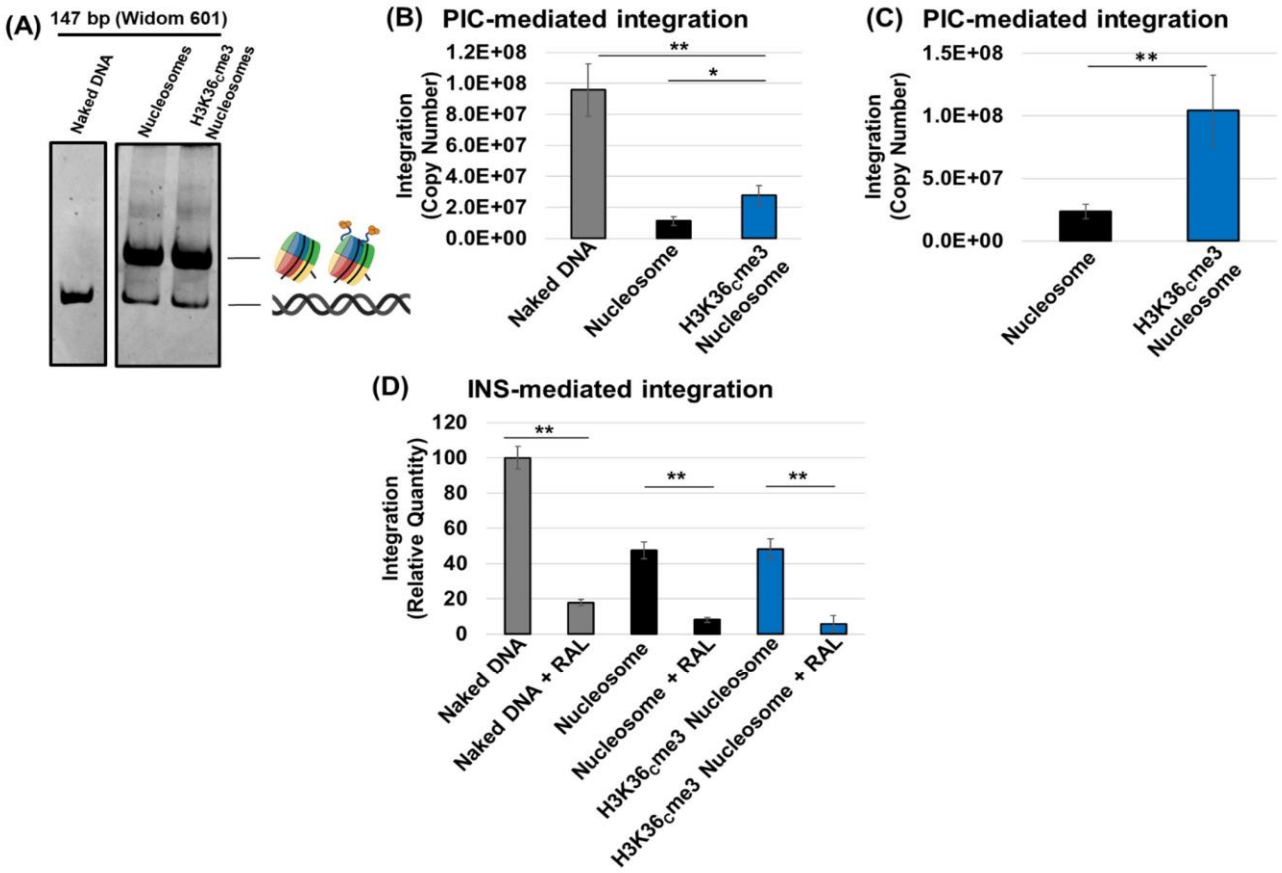
3.6 Supplemental Material



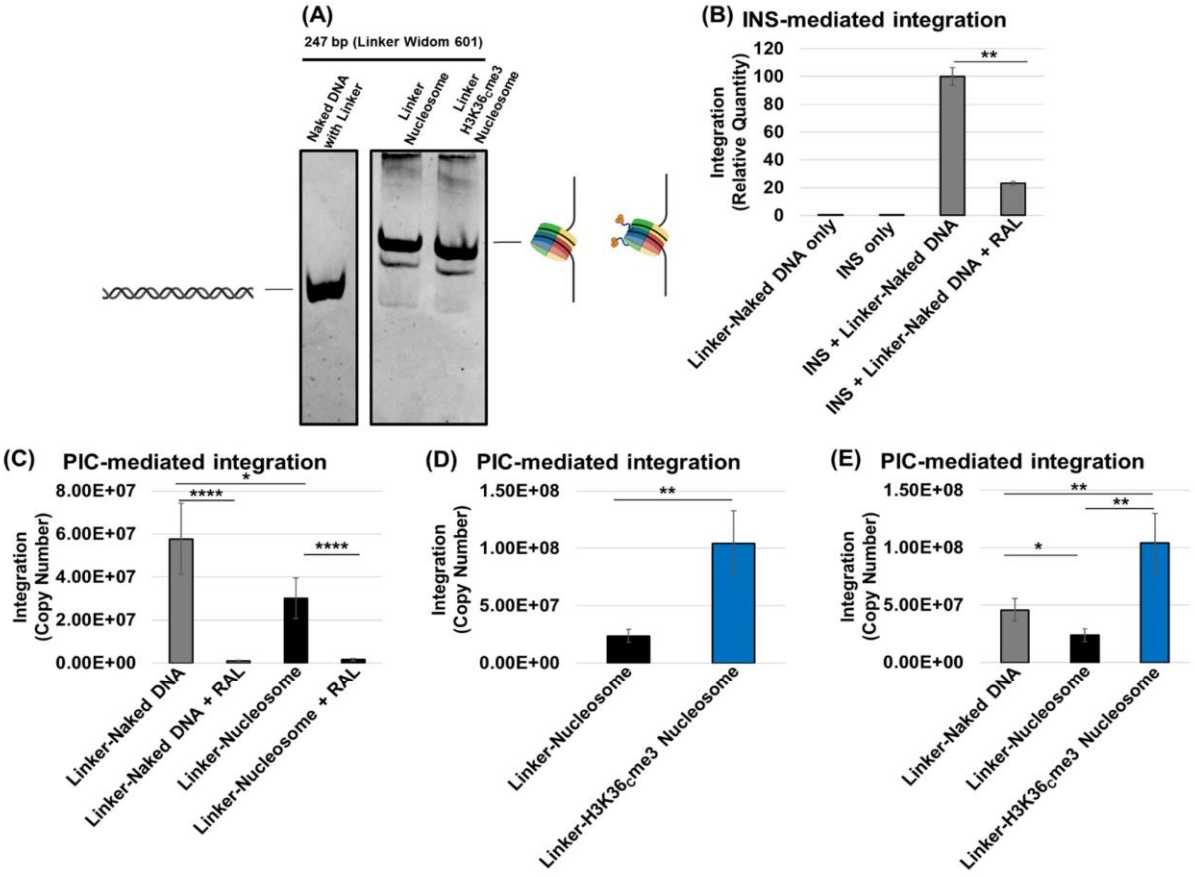
Supplemental Figure S3.1: HIV-1 PIC isolation and Integration Assay. (A) The isolation of HIV-1 PICs is shown as a diagram. First, HIV-1 virions were produced by transfecting HEK293T cells with the pNL4.3 HIV-1 molecular clone, collecting the supernatant 36-48 hours later, filtering the supernatant and treating with DNase I. Then HIV-1 pre-integration complexes (PIC) were isolated from SupT1 cells acutely infected with high tiger HIV-1 virions. The resultant PICs are assessed for viral DNA integration using a nested-PCR assay. (B) HIV-1 PIC-mediated integration is measured by a nested PCR strategy that amplifies the DNA junction formed between the viral DNA and targeted substrate by the stand transfer activity of the HIV-1 integrase. The PIC-mediated integration is specifically inhibited by the integrase strand-transfer inhibitor raltegravir (RAL). The error bars were determined by the SEM and **** P = 0.001 to 0.0001. (C) The chromatin sample that was used as a substrate for PIC-mediated integration was probed by western blot analysis for the nuclear proteins associated with HIV-1 integration.



Supplemental Figure S3.2: The recombinant nucleosome is a barrier to PIC- and INS-mediated viral DNA integration. (A) PIC-mediated integration was measured with the Widom 601 nucleosome positioning sequence as the naked DNA substrate and the nucleosome core particle (nucleosome) substrate, including RAL controls. The data is shown as the mean copy numbers of at least three independent experiments. (B) A schematic of the Ssod7 tagged intasome (INS) assembly. The INS was pre-assembled with the purified integrase-Ssod7 protein and viral DNA sequences from the HIV-1 longterminal repeat (LTR). Once assembled, the INS can carry out a DNA strand transfer reaction with any given target DNA. INS-mediated integration is measured by quantitative PCR using primers designed for the INS donor DNA and the target DNA (acceptor DNA). (C) INS-mediated integration was measured by qPCR with the Widom 601 naked DNA alongside the INS alone, substrate alone, or in the presence of the integrase inhibitor RAL. The error bars were determined by the SEM. The *** P = 0.01 to 0.001 and **** P = 0.001 to 0.0001.

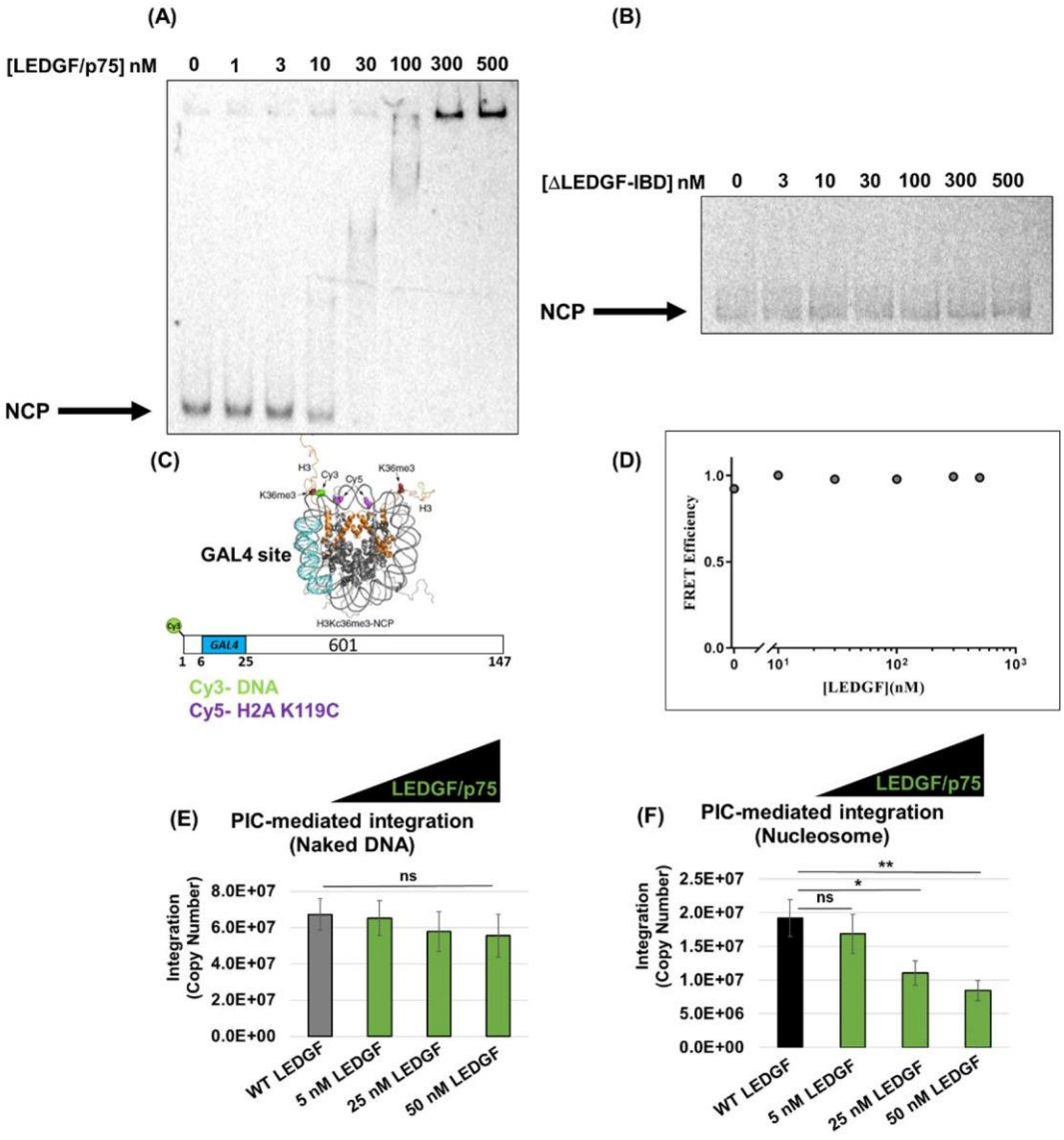


Supplemental Figure S3.3: Comparative analysis of the PIC- and INS- viral DNA integration with the target naked DNA, reconstituted nucleosome, and H3K36Cme3 nucleosome. (A) The Widom 601 DNA alongside the biochemically assembled, nucleosome and H3K36$_C$me3-modified nucleosome were assessed by a 5% polyacrylamide gel electrophoresis (PAGE) and visualized with ethidium bromide staining. (B) Comparative analysis of the PIC-mediated integration with the naked DNA, the nucleosome, and H3K36$_C$me3 nucleosome was measured by nested qPCR, and the data is represented by the mean of viral DNA copy numbers from at least three independent experiments. (C) The direct comparison of the PIC-mediated integration activity with the unmodified nucleosome and the H3K36$_C$me3 nucleosome is shown as the mean of copy numbers from at least three independent experiments. (D) INS-mediated integration was measured with the naked DNA, nucleosome, and H3K36$_C$me3 nucleosome, including RAL controls. The error bars represent the SEM. * Represents $P < 0.05$ and ** $P = 0.01$ to 0.05.



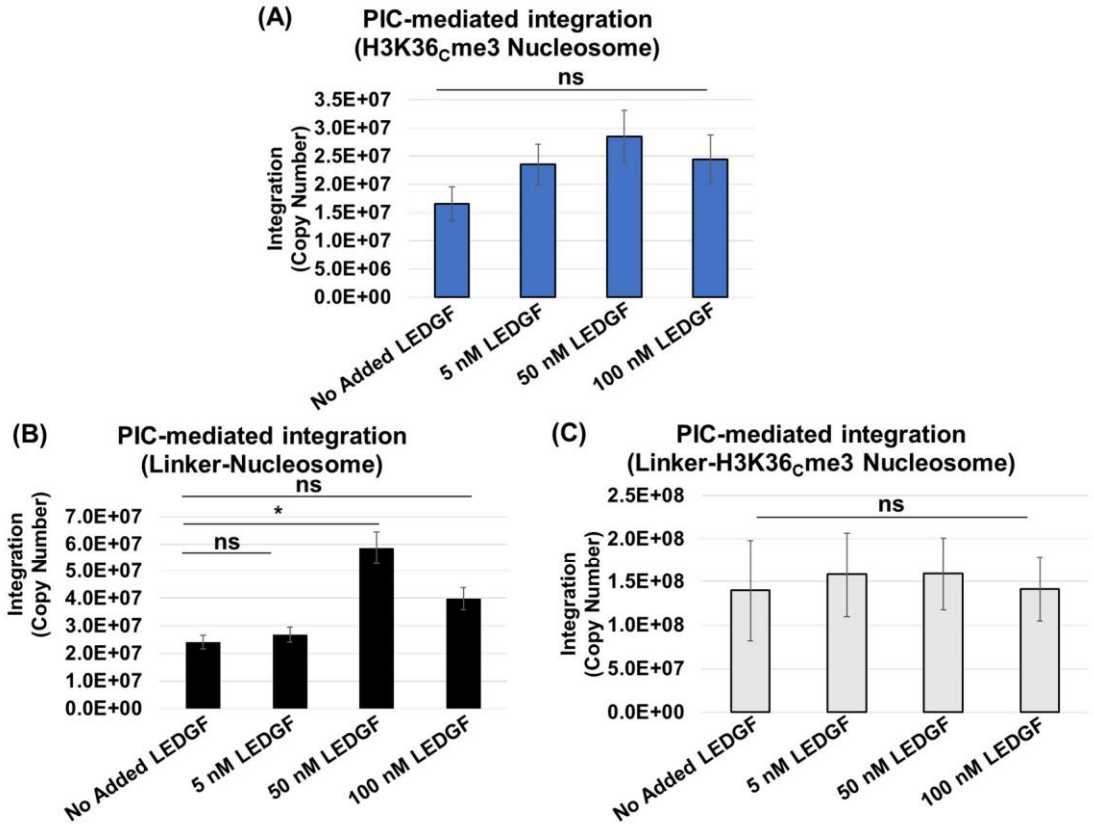
Supplemental Figure S3.4: The H3K36Cme3 nucleosome with linker DNA is an optimal substrate for PIC integration activity, despite the linker containing nucleosome remaining a barrier to HIV-1 integration. (A) The Widom 601 naked DNA with linker DNA, the linker-nucleosome, and the linkerH3K$_C$36me3 nucleosome was visualized on a 5% PAGE stained with

ethidium bromide. (B) INSmediated integration activity was measured with the linker-naked DNA, compared to the intasome only, target DNA only, and RAL integrase inhibitor control. (C) PIC-mediated integration activity was measured with the linker-naked DNA and linker-nucleosome substrate in parallel with RAL control assays. (D) PIC-mediated integration was measured comparing the unmodified nucleosome with linker to the linker-H3K36$_C$me3 nucleosome substrate. (E) The linker-naked DNA, unmodified linkernucleosome, and linker-H3K36$_C$me3 nucleosome were compared as substrates for PIC-mediated integration activity. These results are shown as the mean copy numbers of at least three independent experiments and the error bars represent the SEM. * Represents P < 0.05, ** P = 0.01 to 0.05, and **** P = 0.001 to 0.0001.



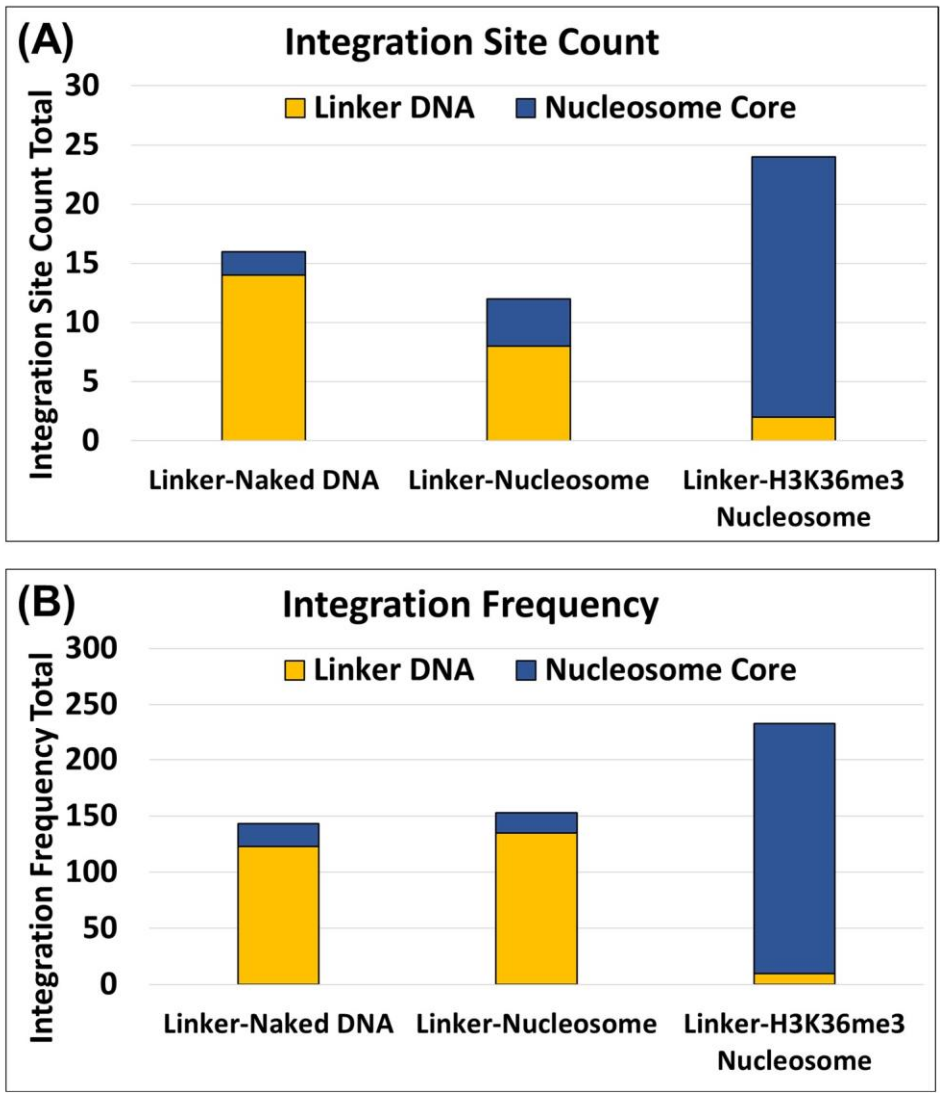
Supplemental Figure S3.5: LEDGF/p75 binds the nucleosome core particle and specifically reduced PIC-mediated integration into unmodified nucleosomes. (A) The interaction between the nucleosome core particle (NCP) and recombinant LEDGF/p75 was assessed by EMSA. (B)

Also, a recombinant truncated mutant of LEDGF/p75, containing only the integrase binding domain (IBD), was assessed for binding to the NCP by EMSA. This truncated LEDGF/p75 does not include the AT-hook motifs or the PWWP domain. (C) A schematic of the modified NCP for the fluorescence resonance energy transfer (FRET) based binding assay is depicted. The ensemble FRET binding assay used NCPs labeled with Cy3 at the end of the nucleosome positioning sequence (NPS) and a Cy5 at the H2AK119C position. A binding site for the transcription factor GAL4 was cloned into the NPS near the H3 tail protrusion site thru the nucleosomal DNA (6-25 bp) was designed to specifically determine the effect of LEDGF/p75 binding to the NCP. (D) The FRET efficiency was measured in the presence of increasing amounts of LEDGF/p75. (E-F) PIC-mediated integration with the naked DNA and the reconstituted nucleosome core particle were measured in the presence of LEDGF/p75 addition. These data are represented as the mean copy numbers of at least three independent experiments and the error bars represent the SEM. * Represents $P < 0.05$ and ** $P = 0.01$ to 0.05.



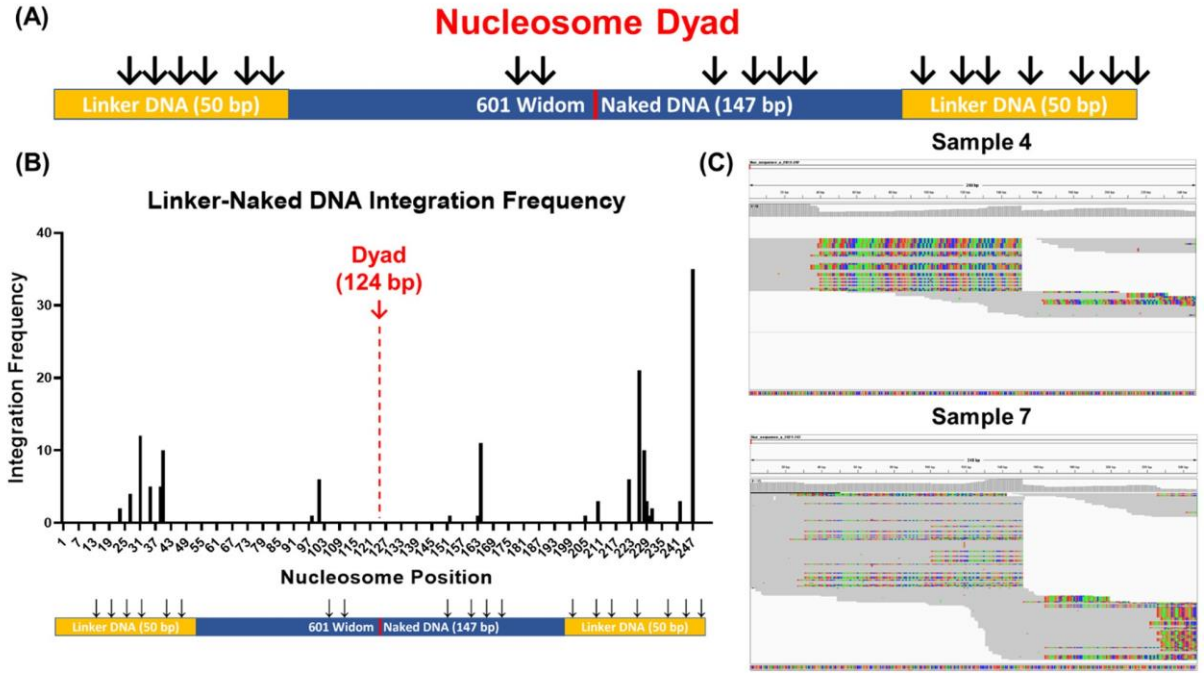
Supplemental Figure S3.6: The LEDGF/p75 mediated inhibition of PIC-mediated integration is relieved by the H3K36Cme3 or the presence of linker DNA. PIC-mediated integration was measured with H3K36$_C$me3 nucleosomes in the presence of LEDGF/p75 addition. (A) PICmediated integration with the H3K36$_C$me3 nucleosome (147 bp) was measured in the presence of LEDGF/p75 addition [5, 50, and 100 nM]. (B-C) PIC-mediated integration was measured with LEDGF/p75 addition to the linker-nucleosome and linker-H3K36$_C$me3 nucleosome substrates. These data are shown as the mean copy numbers of at least

three independent experiments. The error bars were determined from the SEM and * Represents $P < 0.05$.

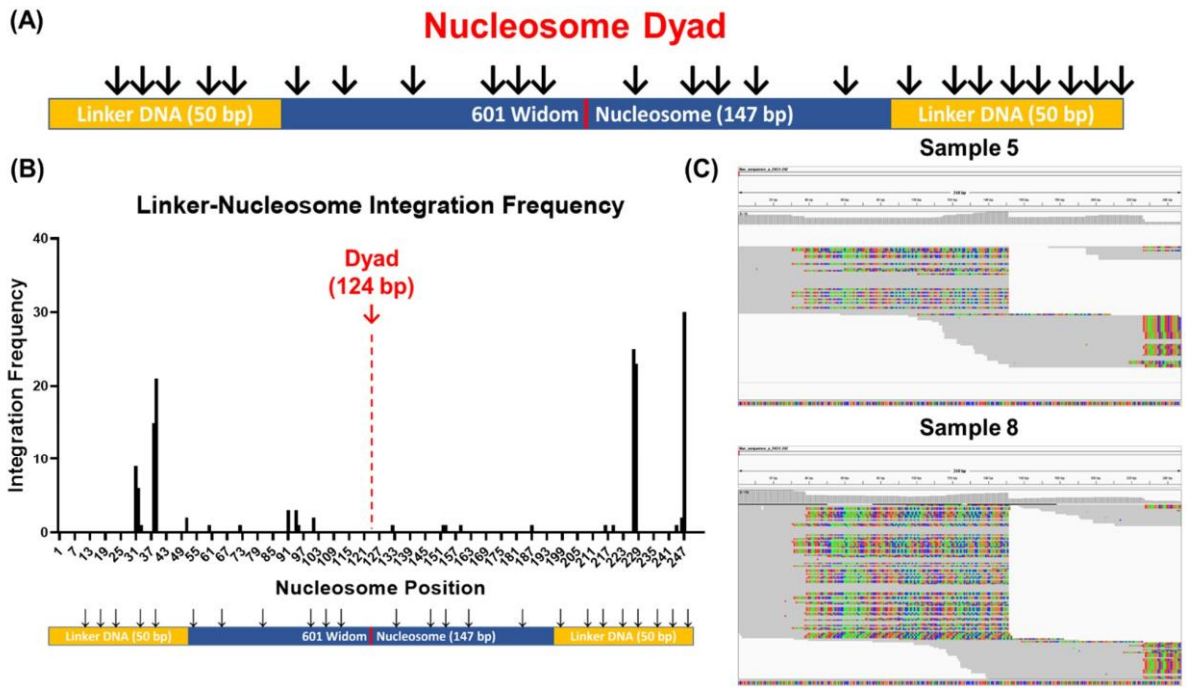


Supplemental Figure S3.7: PIC-mediated integration within the H3K36Cme3 nucleosome with linker DNA is overrepresented in the nucleosome core compared to the naked DNA and unmodified nucleosome substrates. To further study the HIV-1 integration preference within the linker DNA substrates, the DNA from the integration reactions was PCR amplified and

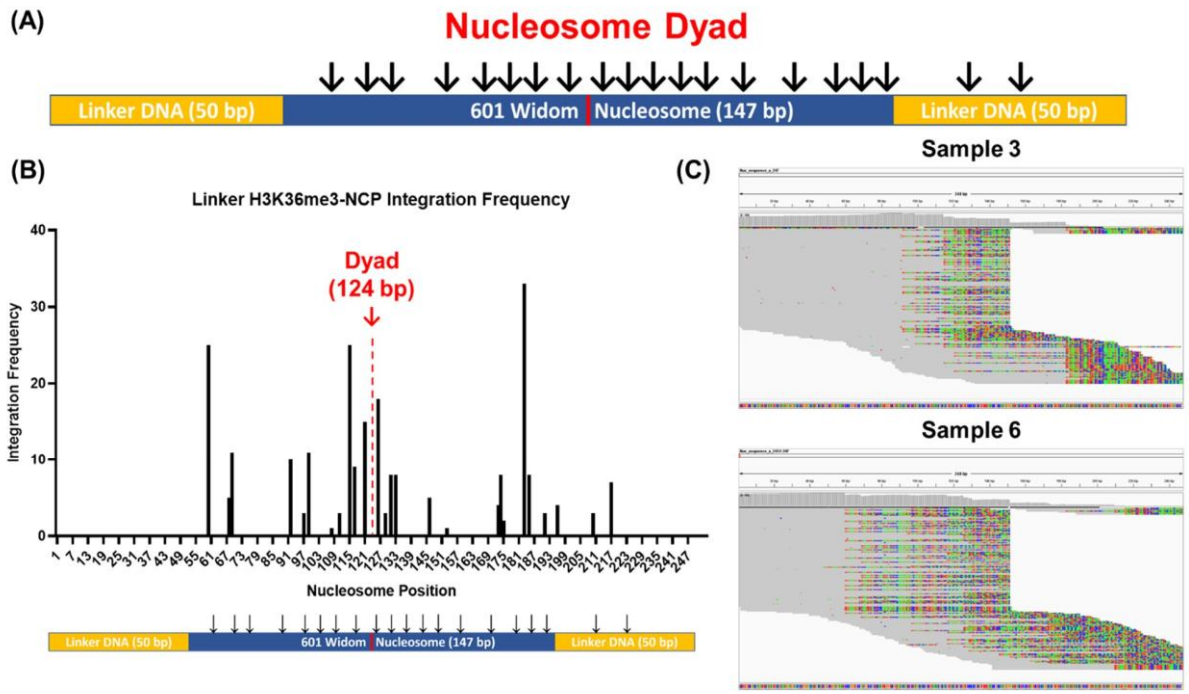
subjected to next-generation sequencing. Integration sites and frequency at those sites within the linker substrates were determined by quantifying the integration junctions. (A) Quantification of the integration sites within the linker substrates shows the number of integration events (sites) within the linker sequences (1-50 bp and 197247 bp) and the nucleosome core sequence (50-197 bp). (B) The frequency of integration junctions at a particular site within the sequence was quantified as the integration frequency in the linker sequences and the nucleosome core sequence.



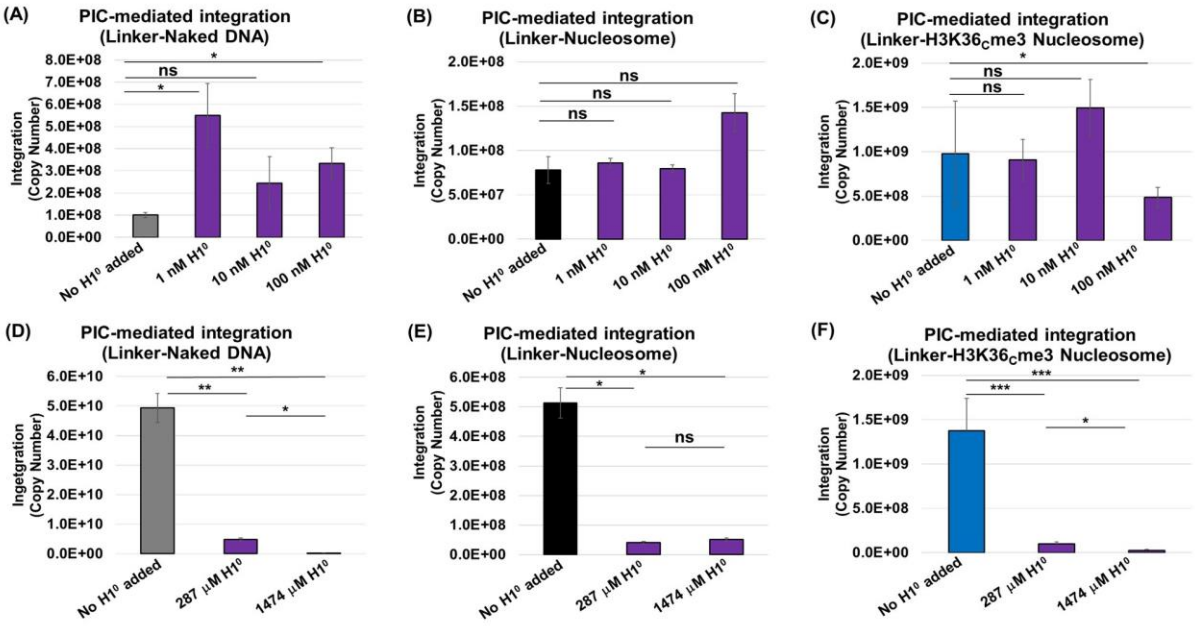
Supplemental Figure S3.8: Distinct Integration sites with the linker-naked DNA. The integration frequency within the linker-naked DNA was determined by quantifying the integration junctions along the substrate sequence. (A) A schematic representation of the integration sites within the linker-naked DNA. (B) The integration frequency is plotted as a histogram for the linker naked DNA, showing the frequency of integration at particular sites of the substrate. (C) The raw IGV data shows the integration sites within the linker-naked DNA substrate.



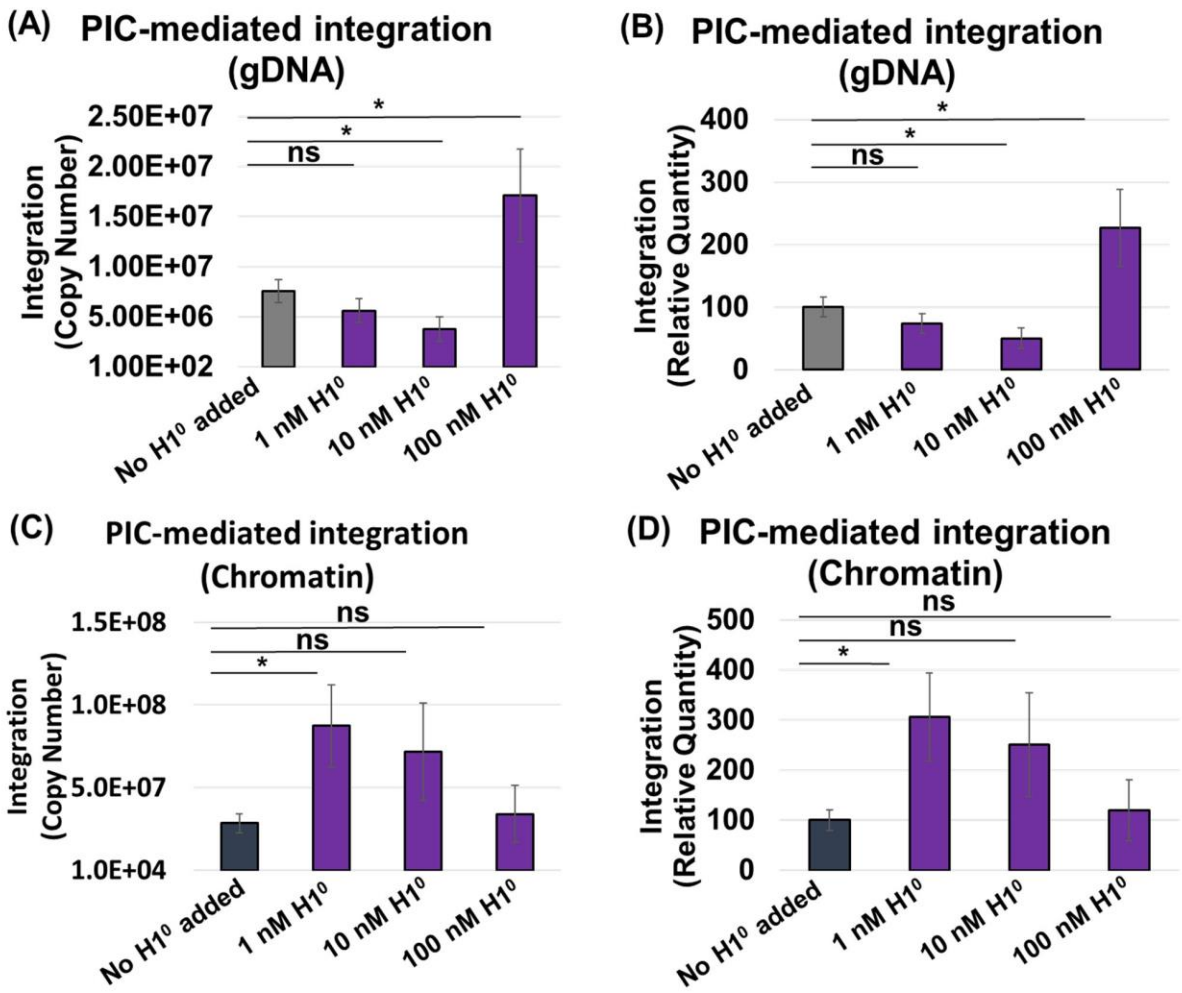
Supplemental Figure S3.9: Distinct Integration sites of the unmodified linker-nucleosome. The integration frequency with the unmodified nucleosome with linker was determined by quantifying the integration junctions along the substrate sequence. (A) A schematic representation of the integration sites within the linker-nucleosome. (B) The integration frequency is plotted as a histogram for the linker-nucleosome, showing the frequency of integration in specific sites of the substrate. (C) The raw IGV data shows the integration sites within the linker-nucleosome substrate.



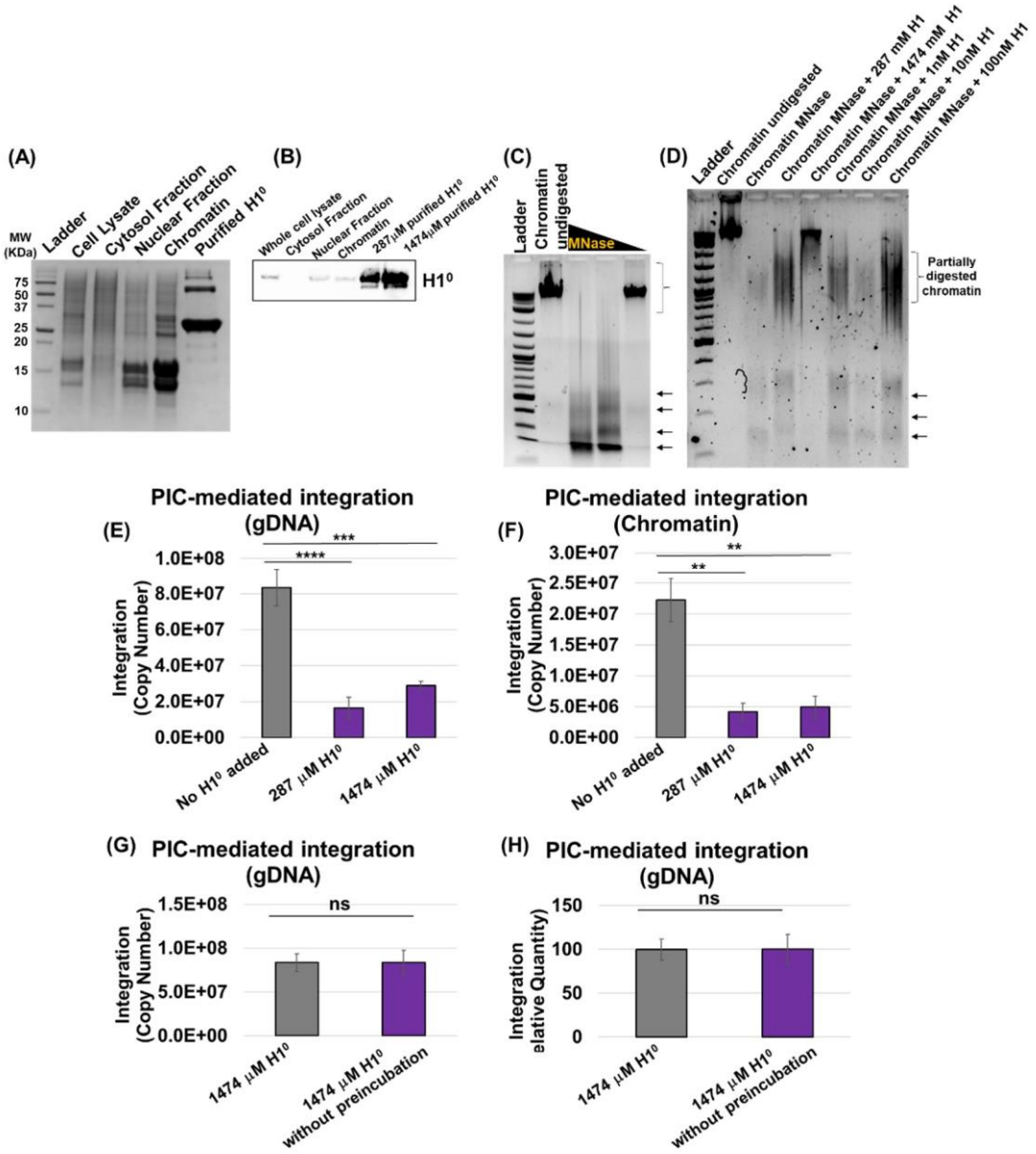
Supplemental Figure S3.10: Distinct Integration sites with the linker-H3K36Cme3 nucleosome. The integration frequency with the linker-H3K36$_C$me3 nucleosome was determined by quantifying the integration junctions along the substrate sequence. (A) A schematic representation of the integration sites within the linker-H3K36$_C$me3 nucleosome. (B) The integration frequency is plotted as a histogram for the linker-H3K36$_C$me3 nucleosome, showing the frequency of integration in specific sites of the substrate. (C) The raw IGV data shows the integration sites within the linker-H3K36$_C$me3 nucleosome.



Supplemental Figure S3.11: Linker histone H1 inhibited PIC-mediated integration at saturating concentrations. PIC-mediated integration with the linker substrates was measured with nM amounts of $H1^0$ addition to the assay condition. (A-C) Non-saturating amounts of $H1^0$ [1, 10, and 100 nM] were added to PIC assay with linker-naked DNA, linker-nucleosome, and linker-H3K36$_C$me3 nucleosome. The data is represented as the mean of the copy number from duplicate experiments. (D-F) The PICmediated integration with saturating amounts of $H1^0$ [247 and 1474 μM] for the linker-naked DNA, the linker-nucleosome, and the linker-H3K36$_C$me3 nucleosome are represented as the mean of viral DNA copy numbers from at least three independent experiments. For all data, the error bars indicate the standard error mean, and * represents $P < 0.05$, ** $P = 0.01$ to 0.05, and *** $P = 0.01$ to 0.001.



Supplemental Figure S3.12: Non-saturating H1 addition to genomic DNA and chromatin differentially impacts PIC-mediated integration. The PIC-mediated integration with the gDNA and chromatin substrates was measured with nM amounts of $H1^0$ addition to the assay. (A-B) PIC-mediated integration with the gDNA in the presence of non-saturating amounts of $H1^0$ is represented as the mean copy numbers of independent replicates and as the relative integration quantity. (C-D) PIC-mediated integration with non-saturating amounts of $H1^0$ incubated with chromatin is shown as the mean copy number and relative integration quantity. The error bars represent the standard error of the mean and * represents $P < 0.05$.



Supplemental Figure S3.13: Chromatin and genomic DNA substrates saturated with H1^0 reduced PIC-viral DNA integration. (A) The purified recombinant H1^0 (NEB) was assessed alongside the chromatin isolation fractions by Coomassie staining. (B) Western blot analysis of the chromatin isolation fractions and the saturating amounts of H1^0 were probed for H1. (C) Chromatin samples isolated from HEK293T cells were assessed for nucleosome integrity by a partial micrococcal nuclease (MNase) digestion at decreasing amounts. The resultant digested DNA was visualized by ethidium bromide staining of a 1.5 % regular agarose gel. (D) Chromatin samples pre-incubated with various amounts of H1^0 were analyzed by a partial MNase digestion and visualized by ethidium bromide staining of a 1.5% agarose gel. (E-F) PIC-mediated integration was measured with genomic DNA and chromatin that were incubated on ice with 287 and 1474 mM of H1^0 (saturating amounts). The concentrations reflect H1^0 amounts that are 1:1 and 1:5 (w/w) of DNA substrate to H1^0 protein. (G-H) An H1^0 time-of-addition experiment with the saturating amount of H1^0 is shown comparing the PIC-mediated

integration with gDNA to an assay with $H1^0$ added after the PIC was exposed to the substrate. PIC-mediated integration data are represented as the mean of viral DNA copy numbers from at least three independent experiments and the error bars represent the standard error of the mean. The ** represents P = 0.01 to 0.05, *** P = 0.01 to 0.001, and **** P = 0.001 to 0.0001.

3.7 LEDGF's effect on transcription factor binding kinetics

We created Cy3-Cy5 FRET ready nucleosomes with a LexA transcription factor binding site built into the nucleosomal DNA and tethered them to a quartz slide surface to perform single molecule FRET imaging. We then monitored the Cy3-Cy5 emission of individual nucleosomes over time after addition of LexA to observe LexA binding events as a sudden change in the nucleosome to a low FRET state (Cy5 emission drops along with a corresponding simultaneous rise in Cy3 emission). LexA bound and unbound times were collected and fit to exponential decays to determine on and off rates of the transcription factor to 601 NPS nucleosomes (Figure 3.10).

LEDGF was found to decrease accessibility of nucleosomes to transcription factors in ensemble FRET, so in order to further study the mechanism of this effect, we decided to probe how LEDGF binding affects the kinetics of transcription factor binding to nucleosomes.

Repeating the single-molecule FRET experiments in the presence of binding saturation concentrations of LEDGF showed that LEDGF binding had a complex effect on the transcription factor kinetics, reducing both the on and off rates of LexA to the nucleosome, but the reduction in on rates was greater (50% reduction in on rates compared to 40% reduction in off rates). This difference in how the rates are changed supports the ensemble FRET experiments showing an overall modest decrease in accessibility.

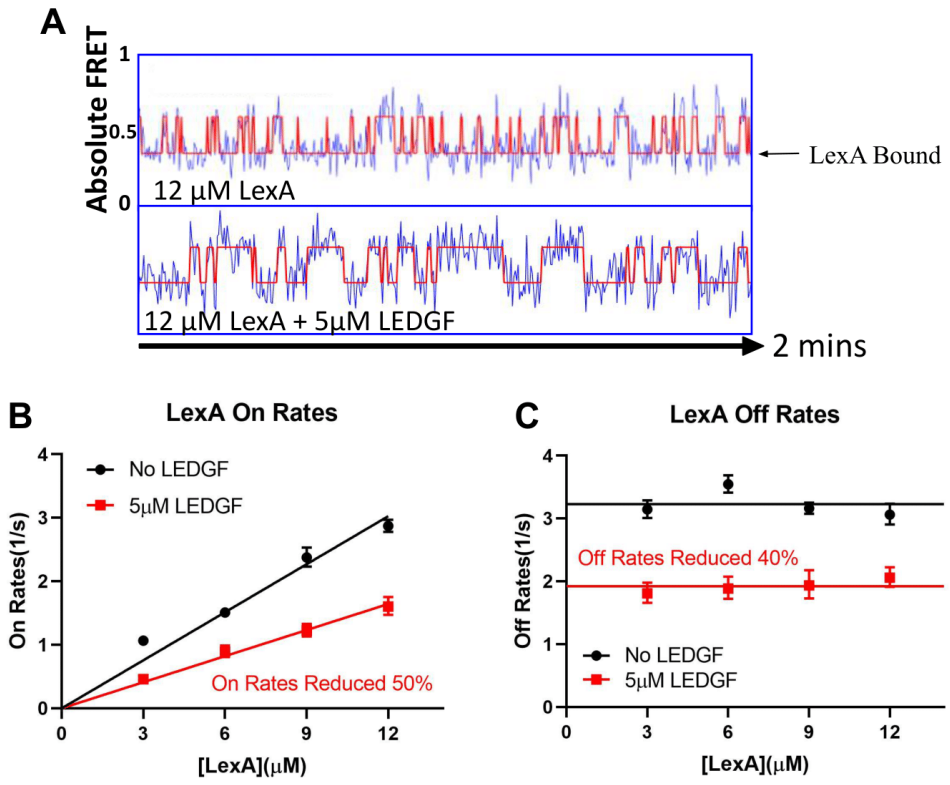


Figure 3.10: LEDGF's impact on transcription factor binding kinetics. Nucleosomes bound with LEDGF had their accessibility to transcription factor binding decreased. (A) FRET time traces of Cy3-Cy5 nucleosomes experiencing binding of LexA transcription factors inducing a low FRET state. Addition of LEDGF affects the LexA binding kinetics. (B) The rate of LexA binding to the nucleosome was reduced by a factor of 50%. (C) The rate of LexA unbinding from the nucleosome was reduced by a factor of 40%. Linear fits to different concentrations of LexA showed concentration dependent on rates and independent off rates, as found previously [37,39].

Chapter 4. Covalent Modifications of Histone H3K9 Promote Binding of CHD3

Reproduced from: Tencer, A. H., Cox, K. L., Di, L., Bridgers, J. B., Lyu, J., Wang, X., … Kutateladze, T. G. (2017). Covalent Modifications of Histone H3K9 Promote Binding of CHD3. *Cell Reports*, 21(2), 455–466.

This chapter covers a study published in Cell Reports that looked into characterizing the PHD domains of the CHD3 chromatin remodeler. I was responsible for the FRET and fluorescence polarization data.

4.1 Abstract

Chromatin remodeling is required for genome function and is facilitated by ATP-dependent complexes, such as nucleosome remodeling and deacetylase (NuRD). Among its core components is the chromodomain helicase DNA binding protein 3 (CHD3) whose functional significance is not well established. Here, we show that CHD3 co-localizes with the other NuRD subunits, including HDAC1, near the H3K9ac-enriched promoters of the NuRD target genes. The tandem PHD fingers of CHD3 bind histone H3 tails and posttranslational modifications that increase hydrophobicity of H3K9-methylation or acetylation (H3K9me3 or H3K9ac)-enhance this interaction. Binding of CHD3 PHDs promotes $H3K9_{C}me3$-nucleosome unwrapping in vitro and perturbs the pericentric heterochromatin structure in vivo. Methylation or acetylation of H3K9 uniquely alleviates the intra-nucleosomal interaction of histone H3 tails, increasing H3K9 accessibility. Collectively, our data suggest that the targeting of covalently modified H3K9 by CHD3 might be essential in diverse functions of NuRD.

4.2 Introduction

Chromatin structure is spatially and temporally regulated throughout the life cycle of the cell and in response to external stimuli. This regulation is driven by two major mechanisms: post-translational modifications (PTMs) of histone proteins and ATP-dependent remodeling of nucleosomes. Histone modifications can directly impact nucleosome structure and dynamics or they can mediate the association and activity of chromatin co-factors [225, 262]. ATP-dependent chromatin remodelers act through disrupting contacts between histones and DNA, promoting nucleosome sliding, nucleosome disassembly and assembly, and eviction of histones [223,248]. In mammals, there are four ATP-dependent chromatin remodelers, CHD (chromodomain helicase DNA binding), ISWI, INO80, and SWI/SNF-like or BAF, with each harboring a conserved SNF2/ SWI2-like catalytic domain and distinguished by varied auxiliary domains. Studies of Drosophila chromatin remodelers reveal that although each remodeler has individual features, ISWI pushes nucleosomes away from its binding site decreasing nucleosome density, whereas other remodelers, including CHD, pull nucleosomes together, increasing nucleosome density at their binding sites [242].

The CHD family of ATPases contains nine members (CHD1–9) that are split into three subfamilies. Subfamily II consists of CHD3, CHD4, and CHD5, which are alternative components of the NuRD (nucleosome remodeling and deacetylase) complex (Figures 4.1A and 4.1B). NuRD alters chromatin structure and DNA accessibility and regulates gene transcription utilizing ATP-dependent remodeling activity of the CHD subunit and histone deacetylase activity of the HDAC1/2 subunit, thus coupling both regulatory mechanisms [262, 264, 270, 274]. The NuRD complex is required for specific gene silencing during differentiation of embryonic stem cells and neurons, reprogramming somatic

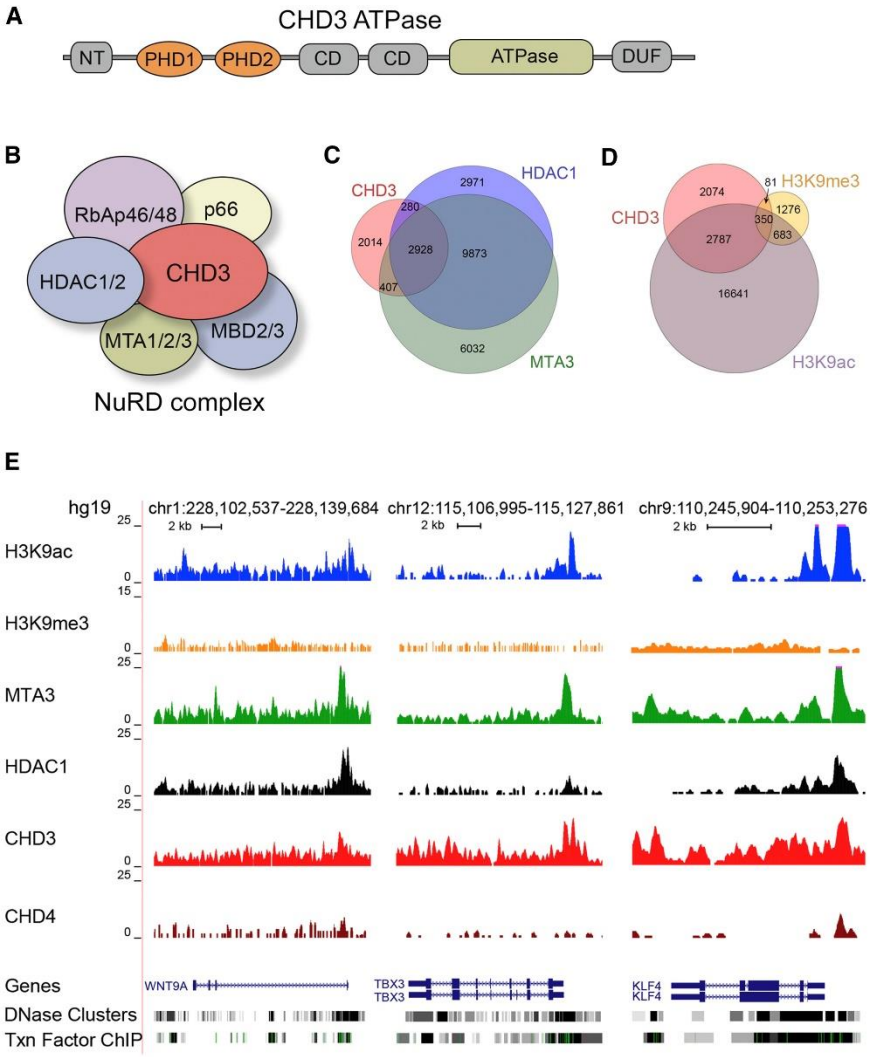


Figure 4.1: CHD3 Co-localizes with HDAC1 and H3K9ac at NuRD Target Genes (A) A diagram of CHD3. (B) The CHD3/NuRD complex subunit composition. (C and D) Venn diagrams show the overlap of the CHD3-, HDAC1-, and MTA3-bound genomic regions (C) and the CHD3-, H3K9ac-, and H3K9me3- bound genomic regions (D). (E) ChIP-seq profiles of the NuRD components (MTA3, HDAC1, and CHD4) and histone marks (H3K9ac and H3K9me3) in K562 cells and of CHD3 in LNCaP cells (SRX1181992) are shown. The three lower tracks correspond to the representative RefSeq genes, as well as DNase I hypersensitivity clusters and transcription factor ChIP-seq binding sites, derived from the ENCODE project and the ENCODE Factorbook repository. See also Figures S1 and S2.

cells, gene repression throughout the cell cycle, and transcriptional regulation of tumor suppressors [220, 233, 255, 256, 267, 271, 272]. Aberrant activity of the NuRD complex is associated with human diseases, including cancer and cardiovascular, neurodegenerative, and developmental disorders [216, 226, 230, 237, 239, 254]. Though CHD3, CHD4, and CHD5 are

highly homologous, they form distinct NuRD complexes, and the combinatorial assembly of the complexes is thought to provide functional specificity. This was recently elegantly demonstrated in the development of the mouse cortex, where incorporation of each CHD was shown to control different sets of genes that mediate non-redundant aspects of development [249]. Differential expression of CHD3, CHD4 and CHD5 is also observed in multiple tissues and disease stages [218, 219, 235, 268]. A longer CHD3 isoform has a SUMO-interacting motif at its C terminus that associates with SUMOylated KAP-1. This interaction is necessary for hetero-chromatin compaction and silencing KAP-1 regulated genes [231, 261].

The CHD3–5 subfamily is characterized by the lack of a conventional C-terminal DNA binding domain and the presence of a tandem of PHD fingers at the N terminus (Figure 5.1A). PHD fingers compose a family of histone binding domains or epigenetic readers that recognize the N-terminal tail of histone H3 and are sensitive to unique patterns of histone PTMs [244]. The PHD fingers of CHD4 and CHD5 have been shown to associate with the H3 tail and display sensitivity to various modifications on residues 3–10 of H3 [228, 240, 246, 251]. The histone-binding activity of PHDs is necessary for CHD4/NuRD-mediated repression [247] and tumor suppressor function of CHD5 [236, 253]; however, the precise role of the PHD fingers of CHD3 remains elusive. Here, we demonstrate that CHD3 co-localizes with other subunits of the NuRD complex and H3K9ac at NuRD target genes. Both PHD fingers of CHD3 associate with histone H3 tails, and this association is augmented through methylation or acetylation of H3K9. Our findings suggest that the ability of CHD3 to sense covalently modified H3K9 may be important for multiple functions of NuRD.

4.3 Results and discussion

4.3.1 CHD3 Co-localizes with HDAC1, MTA3, and H3K9ac at NuRD Target Genes

Growing evidence suggests that CHD3-containing and CHD4-containing NuRD complexes have both overlapping and non-overlapping functions. While the CHD4/NuRD complex has been the focus of many studies, much less is known regarding the CHD3/NuRD complex. To determine if

CHD3 co-localizes in vivo with other subunits of the NuRD complex, such as HDAC1 and MTA3, we analyzed chromatin immunoprecipitation sequencing (ChIP-seq) datasets generated by the ENCODE consortium in K562 cells and in LNCaP cells for CHD3 [224]. We found that CHD3 is recruited to >5,600 genomic regions (Figure 4.1C). Approximately 57% (3,208 out of 5,629) of these regions overlap with the regions occupied by HDAC1, and 52% of the CHD3-bound regions overlap with the HDAC1- MTA3-bound regions. These results suggest a strong co-occupancy of CHD3 with other components of the NuRD complex genome-wide.

Subsequent examination of histone PTMs revealed that ~60% (3,137 out of 5,292) of the CHD3 binding sites overlap with acetylated K9 of histone H3 (H3K9ac), a mark associated with active gene transcription, and only 8% (431 out of 5,292) overlap with trimethylated K9 of histone H3 (H3K9me3), a mark associated with condensed heterochromatin (Figure 4.1D). The NuRD complex has been shown to localize to promoter-proximal regions of transcriptionally active genes to largely downregulate expression [257, 258, 269]. We chose a set of reported NuRD targets [257, 272] to assess CHD3 binding. Variable amounts of CHD3 were detected near promoters of many of the genes examined (data not shown), including KLF4, TBX3, and WNT9A (Figure 4.1E); however, there were a number of NuRD targets lacking CHD3. For instance, CHD3 was below detectable level at CPNE7, DCTN3, and KLF5 (Figure S4.1). CHD3 binding correlated tightly with H3K9ac and was most evident near the transcription start sites (TSS) of the genes, although broader distribution throughout gene bodies was also observed. Other components of the NuRD complex, specifically HDAC1 and MTA3, occupied the same H3K9ac-enriched regions as CHD3, suggesting formation of the CHD3-containing complex at these binding loci. We next examined whether CHD3 and CHD4 are mutually exclusive NuRD subunits or they can be recruited to the same genomic regions. As shown in Figure 4.1E, CHD3 and CHD4 co-occupied the WNT9a and KLF4 promoters; however, CHD4 was not detected at TBX3. Conversely, CHD4 bound to the KLF5 promoter, whereas CHD3 did not (Figure S4.1). The varied degree co-localization of CHD3 and CHD4 supports the idea that the CHD3/4-NuRD complexes can possess both distinct and redundant functions.

Interestingly, CHD3, CHD4, HDAC1, and MTA3 co-localized at promoters of the NuRD components, most notably HDAC1, RBBP4, MTA2, and MBD3 genes (Figure S4.2). This co-

localization suggests a feed-forward role of the CHD3/NuRD complex in mediating expression of its own subunits [241, 259].

4.3.2 PTMs on H3K9 Enhance Binding of the CHD3 PHD Fingers to Histone H3

Depletion of the Drosophila homolog of CHD3/4, MI2, results in a severe loss of the histone H3 ChIP signal at NuRD binding sites, suggesting that nucleosome occupancy depends on its ATPase subunit [242]. As CHD3 contains two sets of reader candidates, the PHD fingers (PHD1 and PHD2) and two chromodomains, we investigated binding of the former to histones. Each PHD finger of CHD3 was individually cloned, expressed as ^{15}N-labeled protein, and examined by nuclear magnetic resonance (NMR) (Figure 4.2). ^{1}H,^{15}N heteronuclear single quantum coherence (HSQC) spectra of PHD1 and PHD2 showed well-dispersed resonances, implying that both domains are independently folded (Figures 4.2A and 4.2D). Gradual addition of the unmodified (un) H3 peptide corresponding to the residues 1–12 of histone H3 tail to the NMR samples of PHD1 or PHD2 caused substantial chemical shift perturbations (CSPs), indicating direct binding between PHDs and H3un (Figures 4.2A and 4.2D, left panels).

To assess the effect of PTMs commonly found in histone H3 on binding of CHD3 PHDs, the H3K9me3, H3K9ac and H3K4me3 peptides were probed in the ^{1}H,^{15}N HSQC titration and tryptophan fluorescence experiments (Figure 4.2). Analysis of NMR CSPs revealed that the K9me3 modification enhances binding of both domains to H3 (compare crosspeaks in the left and middle panels at the protein:peptide ratio of 1:0.5 [red] in Figures 4.2A and 4.2D), whereas the K4me3 modification had strong inhibitory effects. The dissociation constants (KD) for the interactions of PHD1 and PHD2 with H3K9me3, H3K9ac, and H3un were measured by tryptophan fluorescence and found to be 37, 58, and 75 mM for PHD1 and 11, 21, and 33 mM for PHD2, respectively (Figures 4.2B, 4.2C, 4.2E, and 4.2F). These values fall in the range of binding affinities of many other epigenetic readers [217, 245, 252, 263].

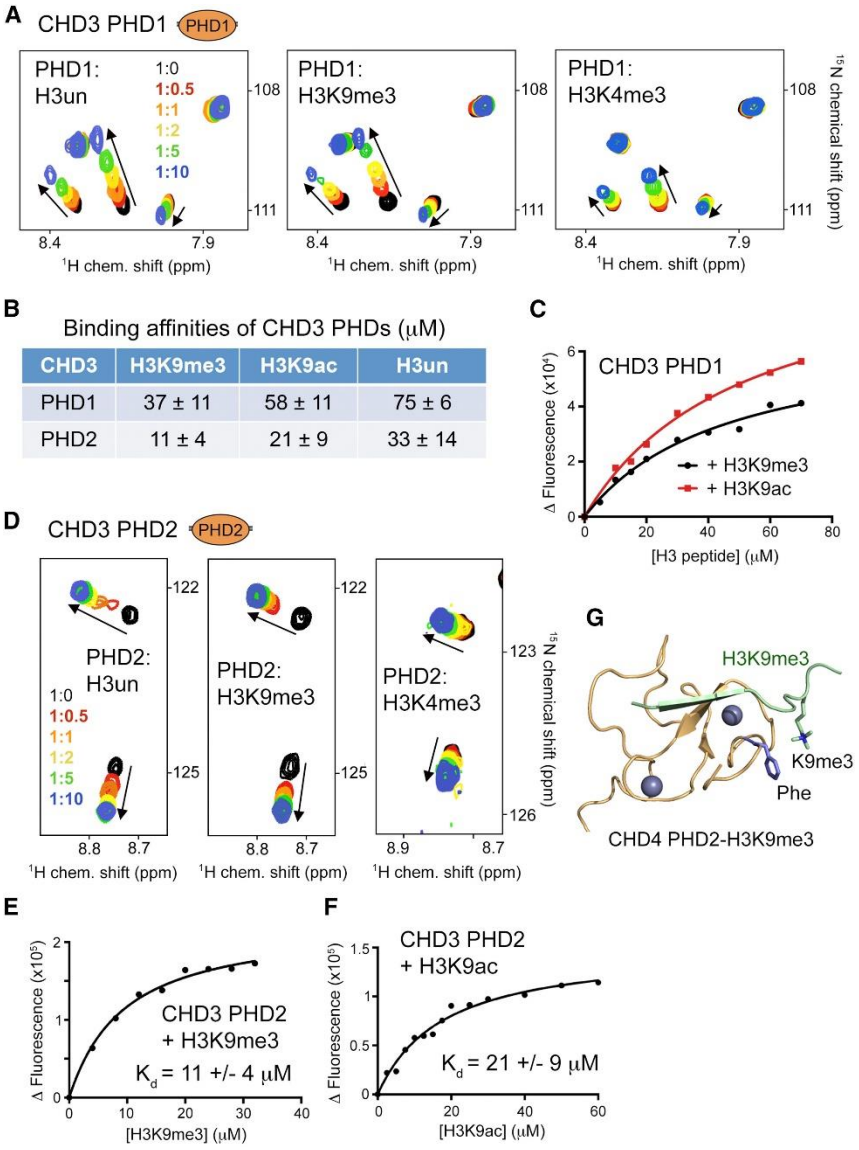


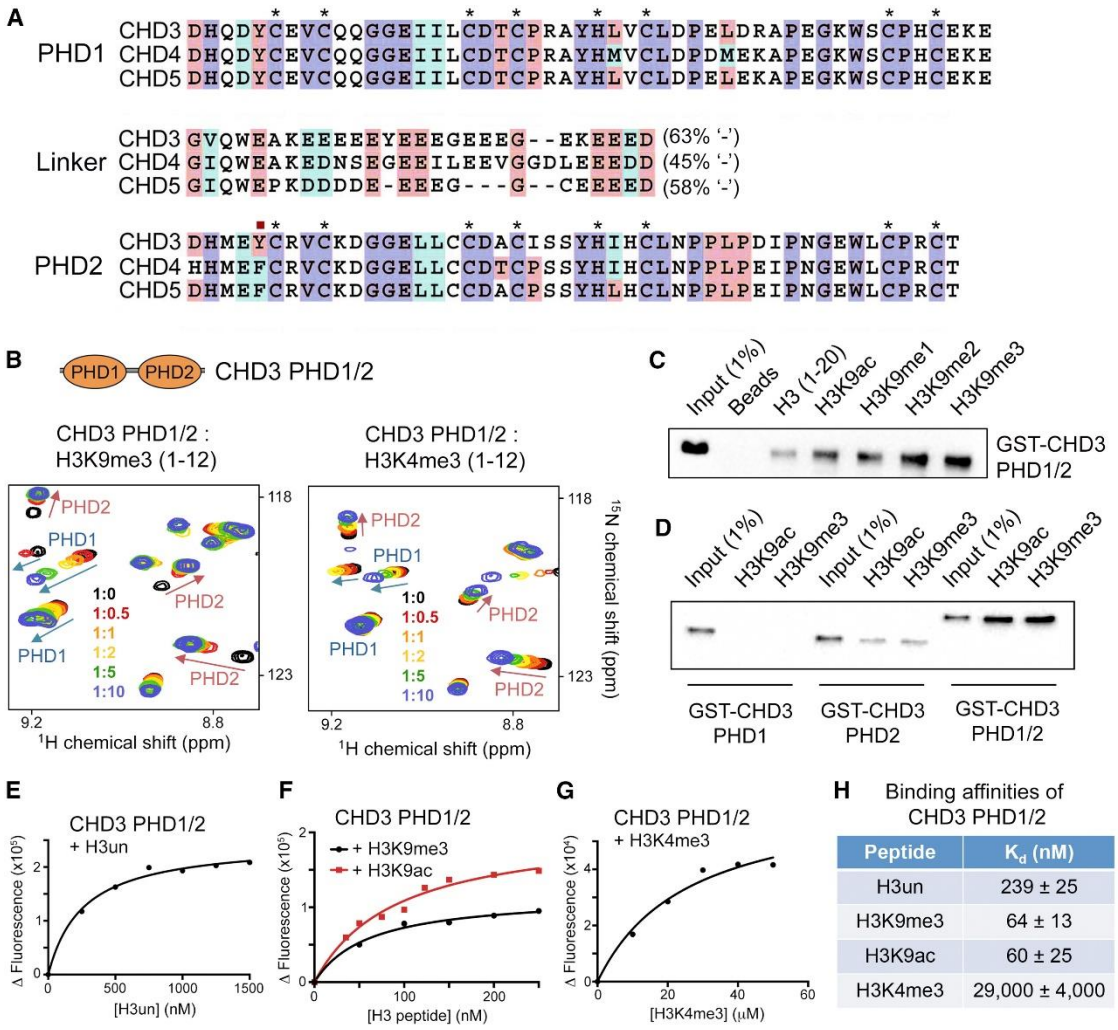
Figure 4.2: PTMs on H3K9 Enhance Binding of the CHD3 PHD Fingers to the H3 Tail (A and D) Superimposed $^{1}H,^{15}N$ HSQC spectra of CHD3 PHD1 (A) and PHD2 (D) collected upon titration with indicated histone H3 peptides (aa 1–12). Spectra are color-coded according to the protein:peptide molar ratio. (B) Binding affinities of the CHD3 PHD fingers to indicated histone peptides as measured by tryptophan fluorescence. Error bars represent SD based on two separate experiments. (C, E, and F) Representative binding curves used to determine the KD values for PHD1 (C) and PHD2 (E) and (F) by tryptophan fluorescence. (G) The ribbon diagram of the CHD4 PHD2-H3K9me3 complex (PDB: 2L75). The H3K9me3 peptide and two zinc ions are shown as a light green ribbon and gray spheres, respectively.

Although trimethylation of H3K9 had a greater effect, facilitating binding of PHD2 by ~3-fold and binding of PHD1 by ~2-fold, we also observed an increase in binding affinities of both PHD1 and

PHD2 to the acetylated H3K9 peptide compared to the unmodified H3 peptide. In support, the enhancement of binding to acetylated or methylated H3K9 has also been reported for the PHD2 finger of homologous CHD4 [246]. The structure of the CHD4 PHD2 finger in complex with H3K9me3 provides an explanation for such an enhancement (Figure 4.2G) [240]. In the complex, K9me3 is engaged in hydrophobic and cation-π interactions with a phenylalanine of the protein [240]. Methylation and acetylation of lysine are both known to augment interactions with aromatic residues due to the increased hydrophobicity. The phenylalanine of CHD4 PHD2 is replaced with a tyrosine in CHD3 PHD2 and a tyrosine is present at the same position in PHD1 of CHD3 (Figure 4.3A, indicated by red dot). Therefore it is possible that PHDs of CHD3 utilize similar mechanisms as PHD2 of CHD4 does to favor methylation or acetylation of H3K9 [246]. Clearly, both PTMs would have a more profound positive effect on the histone-binding activity of CHD3 if PHDs were capable of making simultaneous contacts with histone tails, which, in turn, are restrained by the nucleosome structure. In agreement, full-length CHD3 has been shown to prefer H3K9me3-enriched nucleosomes to unmodified nucleosomes isolated from HeLa cells [234].

4.3.3 The Linked PHD Fingers (PHD1/2) in CHD3 Associate with Individual Histone Peptides

A short linker between the PHD fingers is the least conserved region in the homologous CHD proteins (Figure 4.3A). The CHD3 linker consists of 27 residues compared to 29 residues and 23 residues in CHD4 and CHD5, respectively, and is the most acidic of the three, containing 63% negatively charged amino acids. Overlay of $^{1}H,^{15}N$ HSQC spectra for the apo-states of PHD1, PHD2, and natively linked PHD1/2 of CHD3 demonstrates that resonances corresponding to the linker residues are degenerate in the proton dimension and have high intensity compared to resonances of the folded domains, implying that the linker in CHD3 is



Peptide	K_d (nM)
H3un	239 ± 25
H3K9me3	64 ± 13
H3K9ac	60 ± 25
H3K4me3	29,000 ± 4,000

Figure 4.3: Histone-Binding Activity Is Conserved in the Linked CHD3 PHD1/2 (A) Alignment of the CHD3, CHD4, and CHD5 sequences: absolutely, moderately, and weakly conserved residues are colored purple, pink, and cyan, respectively. The phenylalanine in the PHD2 finger of CHD4 involved in the interaction with K9me3 is indicated by a red dot. Asterisks indicate the zinc-coordinating cysteine residues. The percentage of the negatively charged residues in each linker is shown. (B) Overlays of ^{1}H,^{15}N HSQC spectra of CHD3 PHD1/2 collected as indicated histone H3 peptides (aa 1–12) were added stepwise. (C and D) Western blot analysis of pull-downs using indicated GST-CHD3 constructs and indicated biotinylated histone H3 peptides. (E–G) Representative binding curves used to determine the KD values for PHD1/2 by tryptophan fluorescence. (H) Binding affinities of CHD3 PHD1/2. Error bars represent SD based on two separate experiments (three for H3K9me3). See also Figure S3.

unstructured and flexible (Figure S4.3). Although several resonances corresponding to the PHD fingers themselves display small CSPs compared to the unlinked domains, a considerable spectral

overlap suggested that the overall fold of both domains remains largely unchanged when they are connected.

Next, we tested if the histone binding activity is preserved in the linked CHD3 PHDs. Titration of the H3K9me3 peptide into ^{15}N-labeled PHD1/2 led to large CSPs, indicative of the interaction (Figure 4.3B, left). A close examination of the pattern of CSPs revealed two distinct binding events. One subset of resonances exhibited changes consistent with tighter binding, saturating at a ~2-fold excess of H3K9me3, and another subset exhibited changes consistent with weaker binding, having not fully saturated at a ~10-fold excess of the peptide. Superimposition with NMR spectra of the individual domains confirmed that resonances associated with the tighter binding event all correspond to residues in PHD2, whereas resonances associated with the weaker binding event all correspond to residues in PHD1. Altogether these data suggest that each CHD3 PHD finger in the linked PHD1/2 construct recognizes a distinct histone peptide and that PHD2 binds with higher affinity than PHD1. Titration of the H3K4me3 peptide into the PHD1/2 NMR sample induced CSPs consistent with a much weaker association (Figure 4.3B, right). Thus, methylation of K4 in histone H3 inhibits binding of the CHD3 PHD fingers either separated or linked.

To compare the effect of PTMs in more detail, we tested glutathione S-transferase (GST)-fused CHD3 PHDs in solution-based peptide pull-down assays. The GST-PHD1/2, GST-PHD1, and GST-PHD2 proteins were incubated with biotinylated histone peptides (residues 1–20 of H3) and streptavidin-coated magnetic beads. The histone peptide-bound proteins were detected by western blot (Figures 4.3C and 4.3D). In agreement with NMR data, methylation or acetylation of H3K9 enhanced binding of the linked PHD1/2 fingers (Figure 4.3C). The binding of PHD1/2 was progressively increased from mono- to di- and trimethylated H3K9: the higher methylation state of H3K9 led to the tighter interaction. When the GST-PHD1/2, GST-PHD1, and GST-PHD2 blots were overexposed to the same degree, at which binding of PHD1/2 was evident, association of individual PHD1 and PHD2 with the same histone peptides was barely detectable, indicating that the linked PHD1/2 is a more robust reader (Figure 4.3D). Indeed, apparent KD measured by tryptophan fluorescence showed that the linked PHD1/2 fingers bind to either peptide nearly 10^2-fold stronger than the unlinked individual domains bind (Figures 4.2B and 4.3E–4.3H).

Furthermore, a ~3- to 4-fold enhancement in binding to H3K9me3 and H3K9ac over binding to unmodified H3 peptide was conserved in the linked PHD1/2 fingers.

4.3.4 Natively Linked PHDs Engage Histone H3 Tails in Nucleosomes

To determine whether the linked PHD1/2 in CHD3 concurrently bind histone tails in nucleosomes, we generated nucleosome core particles (NCPs) and used them in ^{1}H,^{15}N transverse relaxation optimized spectroscopy (TROSY) experiments (Figure 4.4A). Upon titrating NCPs into ^{15}N-labeled PHD1/2, we observed a continual decrease in intensity of amide resonances in PHD1/2 and only small CSPs. The loss of signal intensities was likely due to the formation of a large, >220 kDa NCP/CHD3-PHDs complex, characterized by slow correlation time. Much like in the NMR experiments with peptides, two distinct binding events were observed: PHD2 resonances broadened at lower molar ratios of the nucleosome than did PHD1 resonances, indicating that PHD2 has a higher affinity and PHD1 has a lower affinity (Figure 4.4A). Nevertheless, resonances of both PHDs were perturbed, implying that both bind to H3 tails.

The apparent dissociation constant for the interaction of CHD3 PHD1/2 with H3K9Cme3-NCP, the nucleosome carrying a methyllysine analog at position 9 of H3, was measured by fluorescence polarization (KD = 1.4 mM) (Figure 4.4B) and substantiated by electrophoretic mobility shift assay (EMSA) (Figure 4.4C). The CHD3 PHD1/2 was incubated with H3K9Cme3-NCP at various molar ratios, and the reaction mixtures were resolved on a 5% native polyacrylamide gel (Figure 4.4C). A gradual increase in amount of added CHD3 PHD1/2

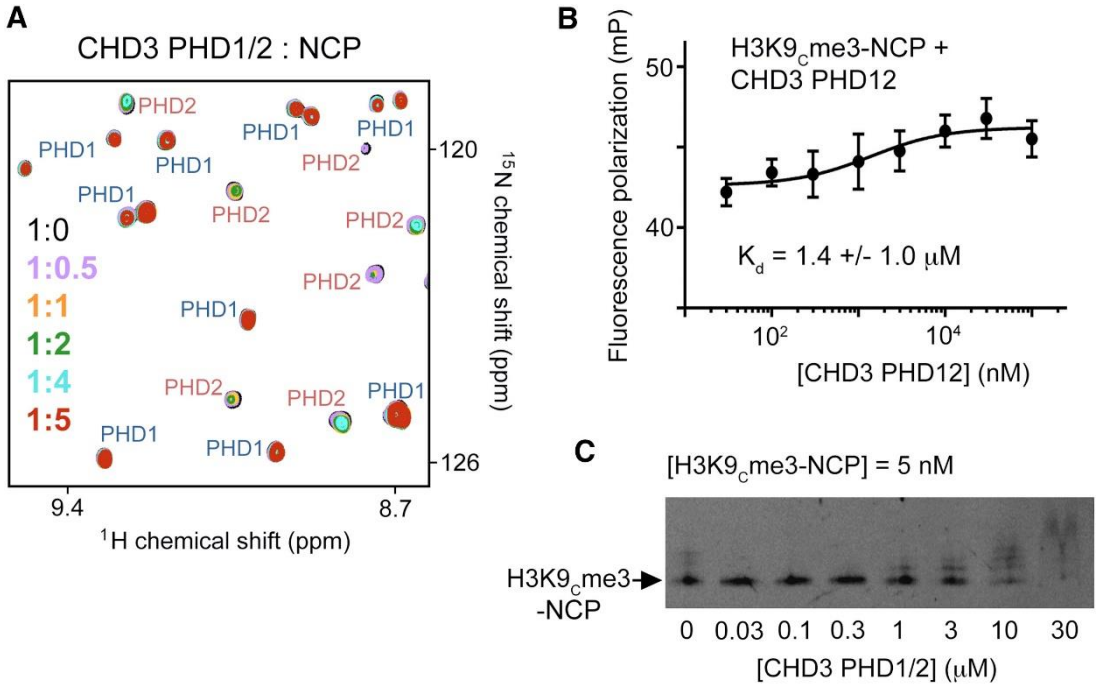


Figure 4.4: Natively Linked PHDs Engage Histone H3 Tails in Nucleosomes (A) Superimposed $^{1}H,^{15}N$ TROSY-HSQC spectra of CHD3 PHD1/2 recorded upon titration with nucleosome. (B) Binding affinity of the CHD3 PHD1/2 fingers to H3K9Cme3-NCP as measured by fluorescence polarization. Error bars represent SD based on three separate experiments. (C) EMSA with H3K9Cme3-NCP in the presence of indicated amounts of CHD3 PHD1/2.

resulted in a shift of the H3K9Cme3-NCP band visible at 3 mM, indicating formation of the H3K9Cme3-NCP-CHD3-PHD1/2 complex. Additional bands observed at 10 and 30 mM corresponded to slower migrating species and suggested that CHD3 PHD1/2 might be involved in multivalent interactions with H3K9Cme3-NCP.

4.3.5 Interaction of CHD3 PHD1/2 with H3 Promotes Destabilization of H3K9Cme3-NCP

The consequence of the interaction of CHD3 PHD1/2 with H3 tails on the nucleosome stability and dynamics was investigated by Forster resonance energy transfer (FRET). We reconstituted NCPs using the 147-bp Widom 601 DNA, in which bases 8-27 were replaced with the transcription factor LexA (LA) binding site (Figure 4.5A). In addition, the Cy3 donor fluorophore

was attached to the 5' end of the 601 DNA adjacent to the LA site, and the Cy5 acceptor fluorophore was attached to histone H2A(K119C) [243]. We anticipated that in the intact conformation of the nucleosome, the FRET signal should be maximal due to the close proximity of fluorophores, whereas destabilization of the nucleosome due to DNA unwrapping would reduce FRET. A substantial decrease in FRET was observed upon titrating LA into wild-type Cy3-Cy5-NCP, and in agreement with previous studies [238, 250], the LA concentration at which FRET efficiency is reduced by half ($S_{1/2}$) was measured to be 1.9 ± 0.4 mM (Figure 4.5B). A similar value of LA $S_{1/2}$ (1.1 ± 0.1 mM) was obtained in the reaction with Cy3-Cy5-H3K9Cme3-NCP, indicating that H3K9Cme3 itself does not significantly alter nucleosome accessibility.

To examine whether CHD3 PHD1/2 impacts the nucleosome stability, we first titrated Cy3-Cy5-H3K9Cme3-NCP with CHD3 PHD1/2 up to 300 mM and detected no measurable change in the FRET efficiency (Figure S4.4). However when PHD1/2 was titrated into Cy3-Cy5-H3K9Cme3-NCP in the presence of LA at a concentration equivalent to the LA $S_{1/2}$, the FRET efficiency was reduced substantially, pointing to an increased accessibility of DNA to LA and yielding a $S_{1/2}$ = 60 mM for CHD3 PHD1/2 (Figure 4.5C). These results indicate that binding of CHD3 PHD1/2 to H3K9Cme3-NCP further promotes nucleosome unwrapping and that both PHD1 and PHD2 of CHD3 must be fully engaged with histone tails to alter the H3K9Cme3-NCP dynamics.

4.3.6 CHD3 PHD1/2 Perturb the Heterochromatin Structure

To test the ability of CHD3 PHD1/2 to affect the nucleosome dynamics in vivo, we transfected HEK293T cells with a GFP-fusion construct of PHD1/2 (GFP-CHD3-PHD1/2) and examined changes in pericentric heterochromatin structure (Figure 4.5D). GFP+ cells were sorted 48 hr posttransfection and the subcellular localization of GFP was assessed by fluorescence

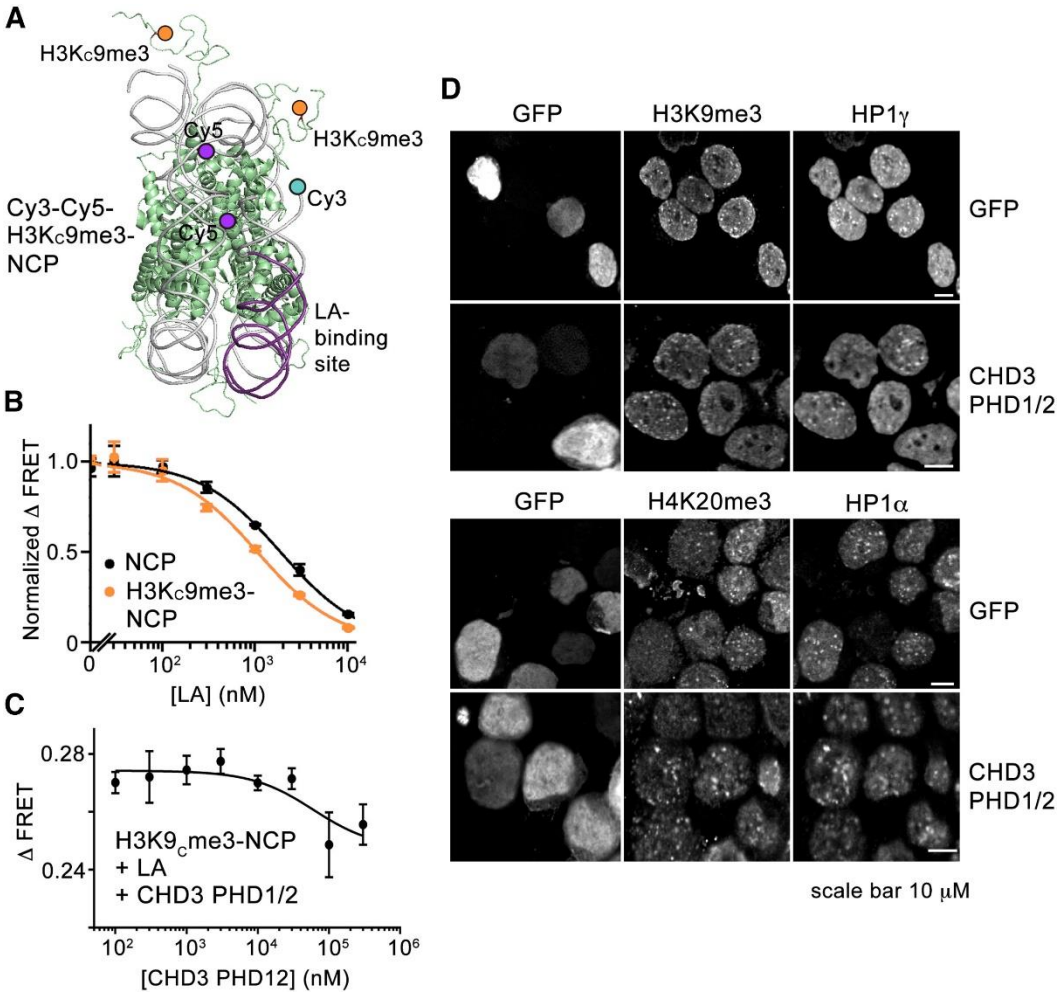


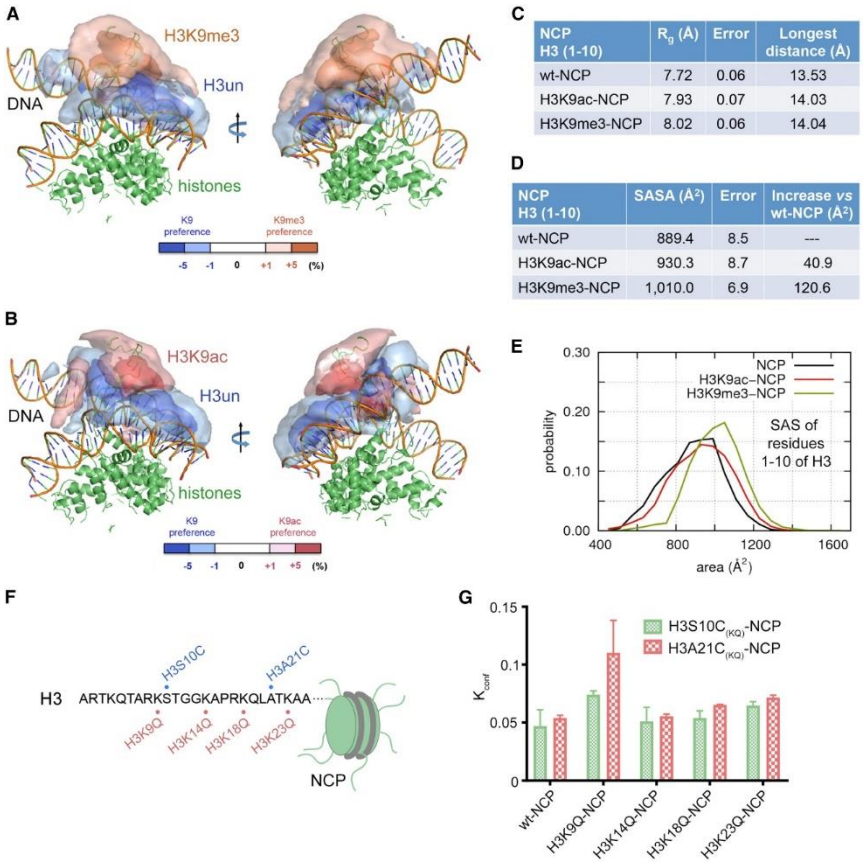
Figure 4.5: CHD3 PHDs Decrease H3K9Cme3- NCP Stability and Alter the Heterochromatin Structure (A) Crystal structure of the NCP (PDB: 1KX5) with histones (green ribbon), H3K9Cme3 (orange circle), Cy5 at H2AK119C (purple circle), Cy3 at the 5' end of 601L (cyan circle), and the LA target site (purple) is shown and labeled. (B) Normalized change in FRET efficiency of the Cy3-Cy5-labeled WT-NCP and H3K9Cme3-NCP due to titration of LA. Error bars represent SD based on three separate experiments. (C) Normalized change in FRET efficiency of the Cy3-Cy5-labeled H3K9Cme3-NCP upon addition of CHD3 PHD1/2 in the presence of 1.3 mM of LA. Error bars represent SD based on three separate experiments. (D) CHD3 PHD1/2 induce changes in pericentric heterochromatin. Representative images of immunofluorescence performed in sorted GFP and GFP-CHD3 PHD1/2 cells. Cells were spotted onto slides and stained for H3K9me3, H4K20me3, HP1g, and HP1a. Scale bar, 10 mm. See also Figure S4.4.

microscopy. As anticipated, in control cells expressing only the GFP tag, we observed the focal accumulation of pericentric heterochromatin markers, H3K9me3 and HP1g (Figure 4.5D, top panel). However in cells transfected with GFP-CHD3-PHD1/2, both H3K9me3 and HP1g were redistributed (Figure 4.5D). In contrast, no changes were detected in localization of H4K20me3

and HP1a markers, indicating that the effect of CHD3 is specific to H3K9me3 and HP1g (Figure 4.5D, two bottom panels). HP1g is known to associate with spatially proximal nucleosomes via binding to H3K9me3 and dimerization and play a critical role in heterochromatin formation and spreading [221]. It has also been shown that histone-binding functions of both PHD fingers of homologous CHD4 is necessary to disrupt the pericentric heterochromatin assembly [247]. Our data suggest that similar to CHD4, CHD3 PHD1/2 fingers could compete with HP1g for H3K9me3-containing tightly packed nucleosomes in heterochromatin regions, resulting in displacement of HP1g and redistribution of H3K9me3. The changes in pericentric heterochromatin structure induced by CHD3 PHD1/2 further support the idea that this region of CHD3/4 is essential in chromatin targeting and remodeling activities of NuRD.

4.3.7 PTMs on H3K9 Facilitate Dissociation of the H3 Tail from NCP

Considering the H3K9me3- and H3K9ac-dependent enhancement in binding of CHD3 PHDs, we examined the effect of these PTMs on the position of H3 tail and its accessibility within the intact nucleosome using molecular dynamics (MD) simulations. Consistent with our recent observations, we found that the un-modified H3 tail is positioned in close proximity to and aligns with nucleosomal DNA in the intact NCP (Figures 4.6A and 4.6B, blue color). However, in the presence of trimethylated K9 or acetylated K9, the H3 tail partially dissociates from the NCP and therefore is more accessible (Figures 4.6A and 4.6B, orange and red color). The atomic resolution structures of canonical PHD fingers, including that of PHD2 of CHD4, show an extended PHD-H3 interface with residues from 1 to up to 10 of the H3 peptide, forming direct contacts with the



NCP H3 (1-10)	R_g (Å)	Error	Longest distance (Å)
wt-NCP	7.72	0.06	13.53
H3K9ac-NCP	7.93	0.07	14.03
H3K9me3-NCP	8.02	0.06	14.04

NCP H3 (1-10)	SASA (Å²)	Error	Increase *vs* wt-NCP (Å²)
wt-NCP	889.4	8.5	---
H3K9ac-NCP	930.3	8.7	40.9
H3K9me3-NCP	1,010.0	6.9	120.6

Figure 4.6: An Increase in Hydrophobicity of H3K9 Facilitates Dissociation of Histone H3 Tail from NCP (A) Differences in the spatial distributions of H3K9ac (orange) and unmodified H3 (blue) tails in the corresponding NCP models. (B) Differences in the spatial distributions of H3K9me3 (pink) and unmodified H3 (blue) tails in the corresponding NCP models. (C) R_g distribution values for H3, H3K9ac, and H3K9me3 tails (all aa 1–10) in the respective NCPs. The error was calculated using 256 independent MD simulations. (D and E) Solvent-accessible surface areas calculated for heavy atoms in H3 (aa 1–10) in WT-NCP, H3K9ac-NCP, and H3K9me3-NCP. The error bars represent the standard deviation calculated using 256 independent MD simulations. (F) Schematic shows the amino acid sequence and mutated positions in the histone H3 tail of NCPs. (G) Conformational equilibrium constant (K_{conf}) determined for each indicated WT and modified NCP. Error bars represent SD based on three separate experiments.

PHD finger [240, 244]. We therefore evaluated the accessibility of residues 1–10 of histone H3 ($H3_{1–10}$) through calculating radii of gyration (R_g) for the modified and un-modified H3 tail in the NCP. Compared to the R_g value for the un-modified $H3_{1–10}$ tail in the wild-type (WT)-NCP, R_g values for the corresponding tails in H3K9ac-NCP and H3K9me3-NCP were significantly higher, pointing to an increase in overall accessibility of these residues (Figure 4.6C). In further support, a substantial increase in the solvent-accessible surface (SAS) areas for the $H3_{1–10}$ tails in H3K9ac-

NCP and H3K9me3-NCP compared to WT NCP (41 angstroms and 121 angstroms, respectively) indicated that H3K9me3 and H3K9ac tails in NCP are more freely available to the CHD3 PHDs than the unmodified H3 tail (Figures 4.6D and 4.6E).

To validate the MDs data, we probed intra-nucleosomal binding of the H3 tail by measuring reaction rates in modified NCPs. Four mutants of H3C110A were generated via replacing K9, K14, K18, and K23 individually with the acetyllysine mimetic, glutamine. In addition, either S10 or A21 in each resultant mutant of H3C110A was substituted with a cysteine residue (Figure 4.6F). Reconstituted NCPs containing the mutated H3 proteins and WT histones H2A, H2B, and H4 (and similarly mutated DNA-free H3/H4 tetramers) were then reacted with fluorescein-5-maleimide, and reaction rates and conformational equilibrium constants ($K_{conf,}$) were determined (Figure 4.6G). Analysis of the conformational equilibrium constants revealed that the unmodified H3 tail binds tightly to the nucleosome surface (K_{conf} of ~0.05). However, replacing K9 with an acetyllysine mimetic resulted in a ~two-fold increase in the average K_{conf} value, whereas the effect of other K-to-Q substitutions tested was less pronounced. Collectively, MDs and chemical reactivity results point to a unique feature of PTMs on H3K9 that notably weaken intra-nucleosomal H3 tail binding and boost the H3K9 accessibility to readers.

4.3.8 Conclusions

Chromatin remodeling is required for genome function and stability and is facilitated by a number of ATP-dependent remodeling complexes. Critical to orchestrating formation of proper chromatin structure and execution of transcriptional programs is the ability of these complexes to interpret the local chromatin landscape. This is mediated in part through the action of histone readers within the complex, which can sense the histone modification states and regulate, recruit, or stabilize the complex accordingly. CHD3 is an ATPase and component of the NuRD complex, a chromatin remodeler and transcriptional regulator [264, 265, 270, 274]. In this study we report that the CHD3 subunit of the NuRD complex is not redundant to the homologous CHD4 subunit and that the tandem PHDs of CHD3 are histone readers that engage distinct histone tails selecting for methylated or acetylated H3K9.

It is interesting that the CHD3 PHD fingers are capable of sensing two histone PTMs that are often linked to opposite biological outcomes, however, this ability can be essential for distinguishing a particular epigenetic state. It also implicates CHD3 PHDs in multiple and dynamic functions of the NuRD complex which operates on H3K9ac-containing and H3K9me3-containing chromatin. One of the H3K9ac-related activities of NuRD is the transcriptional regulation. NuRD localizes to promoter-proximal regions of its target genes and deacetylates histones via the HDAC1/2 subunit, promoting silencing of these genes [257, 258, 269]. Previous observations and data presented in this study demonstrate that CHD3 co-localizes with other subunits of NuRD and H3K9ac near promoters and therefore binding of PHDs to H3K9ac would help in stabilizing the complex.

NuRD is implicated in a wide range of nuclear events that require sliding of nucleosomes, and it can facilitate both closing and opening of chromatin. The nucleosome sliding and chromatin remodeling function of NuRD requires the catalytic activity of the ATPase subunit but is independent of deacetylation by HDAC1/2. The importance of NuRD for heterochromatin maintenance and assembly is particularly well established and is coupled to methylation of H3K9. Thus interaction of the CHD3 PHD fingers with H3K9me3 could aid in both heterochromatin targeting and remodeling as our data show that PHDs promote H3K9Cme3-nucleosome destabilization in vitro and lead to redistribution of pericentric heterochromatin markers, H3K9me3 and HP1g in vivo. Furthermore, MDs and nucleosome-reaction rates assays suggest that PTMs that increase hydrophobic character of H3K9 uniquely alleviate the intra- nucleosomal contacts of the histone H3 tail, freeing the lysine side chain for binding of PHDs.

We note though that to fully understand the multifaceted biological roles of CHD3, characterization of the entire CHD3/NuRD complex is required. Further studies are also needed to elucidate an intricate crosstalk involving not only the gene regulatory and nucleosome remodeling activities of NuRD but also the associated coregulator-mediated events, including demethylation of H3K4 by LSD1 [267], histone H2A.Z deposition [272], a co-factor control [273], and deacetylation of H3K27 followed by its methylation by PRC2 complex [258].

4.4 Experimental Procedures

DNA Cloning and Protein Purification

The following constructs were cloned into pGex6p1: PHD1 (373–427), PHD2 (454–508), and PHD1/2 (373–508) of CHD3. The constructs were expressed in Escherichia coli BL21 DE3 pLysS cells grown in Luria broth or 15NH4Cl minimal media supplemented with 50 mM ZnCl2. Bacteria were harvested by centrifugation after induction with IPTG (isopropyl b-D-1-thiogalactopyranoside) (0.5–1.0 mM) and lysed by sonication. The unlabeled and ^{15}N-labeled GST-fusion proteins were purified on glutathione Sepharose 4B beads, and the GST tag was cleaved with Prescission protease. The proteins were concentrated into 20 mM Tris (pH 6.8), in the presence of 150 mM NaCl and 3 mM DTT.

Genomic Data Analysis

Raw ChIP-seq data of H3K9ac, H3K9me3, MTA3, HDAC1, and CHD4 in K562 cells were downloaded from the ENCODE Project (https://www. encodeproject.org)[224]. Analysis of the CHD3-bound regions was performed using the GEO ChIP-seq dataset SRX1181992 in LNCaP cells. The raw reads were mapped to human reference genome NCBI 36 (hg19) by Bowtie 1, allowing up to 2 mismatches. The ChIP-seq profiles were generated using model-based analysis of ChIP-seq (MACS) (v.1.4.2) with only unique mapped reads. Clonal reads were automatically removed by MACS. The ChIP-seq profiles were drawn using the UCSC Custom tracks utility. Peaks within a 1-kbp neighborhood were merged across multiple samples by peak overlap analysis. Venn diagrams were prepared using the VennDiagram R package.

Nucleosome Preparation

6 x His-tagged or untagged H. sapiens histones H2A, H2B, and H3.2 were expressed in Escherichia coli BL21 DE3 RIL cells and H. sapiens histone H4 in E. coli BL21 DE3 pLysS, isolated from inclusion bodies and purified using Ni-NTA beads (QIAGEN, 30250) or ion exchange. To reconstitute the octamer, appropriate molar ratios of all four 6 x His-tagged or

untagged histones were mixed in denaturing conditions and dialyzed against 20 mM Tris-HCl (pH 7.5), 2 M NaCl, 1 mM EDTA, and 5 mM DTT. 6 x His-tag was cleaved off from the histones octamer overnight by adding the Tobacco etch virus (TEV) and PreScission proteases. The octamer was further purified by size exclusion chromatography. Thirty-two copies of 147-bp 601 Widom DNA were cloned in pJ201 with EcoRV cut sites on either end. The plasmid was purified as previously described (Dyer et al., 2004), and individual copies were released with EcoRV digestion. Purification away from parent plasmid was done by polyethylene glycol (PEG) precipitation. The unmodified NCPs were reconstituted as described previously [227]. Briefly, 601 Widom DNA and purified octamer were mixed at the appropriate molar ratio in 2MKCl/NaCl, 10mM Tris (pH 7.5), and 1 mM DTT and then were reconstituted by slow desalting dialysis into the final solution of 150mM KCl/NaCl, 10mM Tris (pH 7.5), and 1mM DTT. NCPs were further purified from free DNA by sucrose gradient.

NMR Spectroscopy

NMR experiments were performed on a Varian 600 and 500 MHz spectrometers at the University of Colorado School of Medicine NMR core facility or on a Bruker 800 MHz spectrometer at the University of Iowa. ^{1}H,^{15}NHSQCor TROSY-HSQC spectra of 0.05–0.1 mM uniformly ^{15}N-labeled CHD3 PHDs were collected at 298 K or 310 K (nucleosome titration). The binding was characterized by monitoring chemical shift changes in ^{1}H,^{15}N HSQC or TROSY spectra upon addition of modified and unmodified histone H3 peptides (amino acids [aa] 1–12, synthesized by Synpeptide) and nucleosomes.

Fluorescence Spectroscopy

Spectra were recorded at 25ºC on a Fluoromax-3 spectrofluorometer (HORIBA). The samples containing 0.5–1 mM CHD3 PHDs in 20mM phosphate (or 25 mM Tris) (pH 7.5), 150 mM NaCl, and 3 mM DTT buffer, and progressively increasing concentrations of the histone peptides (aa 1–12 of H3) were excited at 295 nm. Emission spectra were recorded between 320 and 360 nm with a 0.5-nm step size and a 0.5-s integration time and averaged over 3 scans. The KD values were determined using a nonlinear least-squares analysis and the equation:

$$\Delta I = \Delta I_{max} \frac{(([L] + [P] + K_d) - \sqrt{([L] + P] + K_d)^2 - 4[P][L]))}{2[P]}$$

where [L] is the concentration of the histone peptide, [P] is the concentration of the protein, ΔI is the observed change of signal intensity, and ΔI_{max} is the difference in signal intensity of the free and bound states of the protein, or $\Delta I = \Delta I_{max}$[L]/(K_D +[L]). The K_D values were averaged over two or three separate experiments, with error calculated as the SD between the runs.

Pull-Down Assays

The GST-CHD3-PHD1/2, GST-CHD3-PHD1, and GST-CHD3-PHD2 proteins (50 pmol per reaction) were incubated with 500 pmol of each biotinylated histone peptide for 2 hr at 4ºC in peptide binding buffer (50 mM Tris [pH 8.0], 300 mM NaCl, 0.1% NP-40, and 1 mM $ZnCl_2$) supplemented with 0.5% (w/v) BSA (Sigma). The protein-peptide mixtures were then incubated with pre-equilibrated streptavidin-coated magnetic beads (Pierce) for 1 additional hour at 4ºC. The beads were washed 3 x with peptide binding buffer. The protein-peptide complexes were eluted with 1 x SDS loading buffer, resolved on SDS-PAGE, and transferred onto a polyvinylidene fluoride (PVDF) membrane. Bound proteins were detected using an anti-GST antibody (EpiCypher; cat. no. 13-0022) diluted 1:2,000 in Tris-buffered saline (TBS)-T supplemented with 5% (w/v) milk.

Methyllysine Analog Generation

A histone H3(C110A)$K9_C$ point mutant was generated by site-directed mutagenesis using the Stratagene QuickChange XL Kit and purified as described above. The histone H3$K9_C$me3 was generated by alkylation of H3$K9_C$ with (2-bromoethyl) trimethylammonium bromide following the protocol [260]. After desalting, the protein was dialyzed into water and re-lyophilized. Labeling was confirmed by MALDI-TOF mass spectrometry.

Fluorescence Polarization

H3K9$_C$me3-NCPs were prepared similarly to those used for FRET assays, except that the nucleosomal DNA was labeled with fluorescein instead of Cy3 and nucleosomes contained unlabeled WT H2A. Fluorescence polarization measurements were carried out with increasing amounts of CHD3 PHD1/2 and 5 nM nucleosomes in 15 mM Tris (pH 7.5) buffer supplemented with 75 mM NaCl, 0.00625% Tween20, and 3 mM dithiothreitol. The samples were loaded into a Corning round bottom polystyrene plate, and the anisotropy measurement was acquired with a Tecan infinite M1000Pro plate reader by exciting at 470 nm and measuring the polarized emission at 519 nm with 5 nm excitation and emission bandwidths. The fluorescence polarization was calculated from the emission polarized parallel and perpendicular to the polarized excitation light essentially as previously described (Canzio et al., 2013). The data were fitted to the binding isotherm, $(C_F\text{-}C_0) / (1+S_{1/2}/ [CHD3]) + C_0$, to determine the $S_{1/2}$. [CHD3] is the total concentration of CHD3 PHD1/2, and C_F and C_0 are the final and initial polarization values, respectively.

EMSA

EMSAs were carried out using 5 nM Cy3-Cy5-labeled H3K9$_C$me3-NCPs in 8.3 mM Tris-HCl (pH 7.5), 62.5 mM NaCl, and 1.25 mM dithiotheritol. Each sample containing 2.5% Ficoll 400 was loaded onto a 5% native polyacrylamide gel, and electrophoresis was performed in 0.33 Tris-borate-EDTA (TBE) at 300 V for 60 min. Cy5 fluorescence images were acquired with a Typhoon Phosphor Imager (GE Healthcare).

FRET Assays

Nucleosomal DNA, 601L, was prepared by PCR from a plasmid containing the LA binding site at bases 8–27 with the Cy3-labeled oligonucleotide, Cy3-CTGGAGATACTGTATGAGCATACAGTACAATTGGTC, and the unlabeled oligonucleotide, ACAGGATGTATATATCTGACACGTGCCTGGAGACTA. The Cy3-labeled oligonucleotide was labeled using a Cy3-NHS ester (GE Health- care) at a 5’ amino group and purified by RP-HPLC on a 218TPTM C18 (Grace/Vydac) column. Expression and purification of human histones and LA, the labeling of recombinant H2AK119C-H2B heterodimer with Cy5-maleimide, and

reconstitution and purification of Cy3-Cy5-labeled WT-NCP and H3K9$_C$me3-NCP was performed essentially as previously described [229].

FRET efficiency measurement experiments were carried out on a Horiba Scientific Fluoromax 4. Samples were excited at 510 and 610 nm, and the photoluminescence spectra were measured from 530 to 750 nm and 630 to 750 nm for donor and acceptor excitations, respectively. Each wavelength was integrated for 1 s, and the excitation and emission slit width were set to 5 nm with 2-nm emission wavelength steps. FRET measurements were computed through the (ratio)A method. LA titrations were carried out with 5 nM nucleosomes in 15 mM Tris buffer (pH 7.5), supplemented with 75 mM NaCl, 0.00625% Tween20, and 10% glycerol. CHD3 PHD1/2 titrations were carried out with 75 nM H3K9$_C$me3-NCP in the presence of 1.1 mM LA. Titrations were fit to $E = E_0 + (E_f - E_0)/(1 + (S_{1/2}/C))$, where E is the FRET efficiency at concentration C of the titrant, E_0 the efficiency in the absence of the titrant, E_f the efficiency at high titrant concentration, and $S_{1/2}$ is the inflection point. Errors represent SD based on three experiments.

Cell Culture, Immunofluorescence, and Imaging

HEK293T cells were purchased from ATCC and maintained in DMEM/F12 medium (GIBCO) supplemented with 10% fetal bovine serum (FBS). Cells were transfected using Lipofectamine 2000 according to the manufacturer's protocol (Invitrogen). Transfected 293T cells were fluorescence-activated cell sorting (FACS) sorted based on GFP intensity (BD FACSAria II), and immuno-fluorescence was performed as previously described [247]. Images were collected on a Zeiss Axiovert 200 imaging system equipped with an Axiocam MR digital camera controlled by AxioVision software.

MD Simulations

The MD simulations in the presence of NCP were carried out using an adaptive lambda square dynamics (ALSD) enhanced sampling method essentially as previously described [232]. The initial structure of the nucleosome with a 10-bp linker was modeled using NCP structures (1kx5

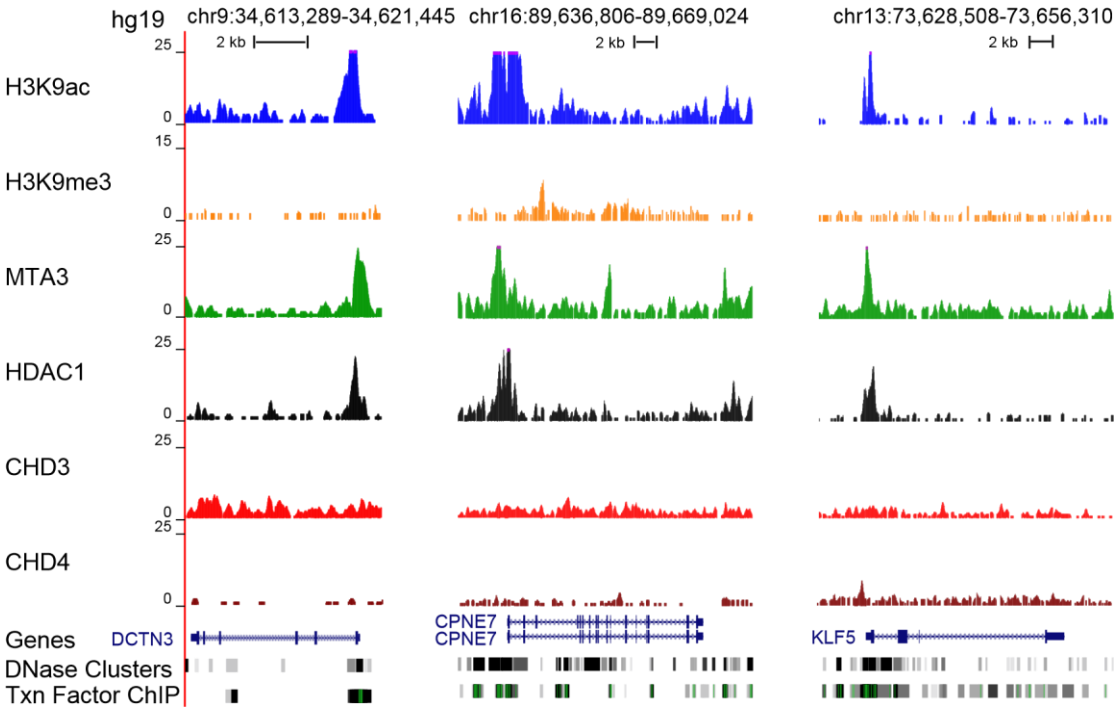
and 1zbb). In the calculation, only atoms within a sphere of 54 Å radius at the root of the H3 tail (a nitrogen atom in the 40th residue) were considered. The simulations were carried out with an explicit solvent model at a salt concentration of 150 mM. The force fields used are an assisted model building with energy refinement (AMBER)-based hybrid force field, AMBER bsc0, TIP3P, and Joung-Cheatham for proteins, DNA, water molecules, and ions, respectively (see references in [232]). The point charge parameters for acetylated and tri-methylated lysines were taken from http://pc164.materials.uoi.gr/dpapageo/amberparams.php.

To speed up the conformational sampling, 256 independent simulations were conducted with different initial conformations. For each of the systems, we carried out the ALSD production run for 7.68 μs (= 30ns x 256 runs) in total after ALSD iterative runs (8.704 μs [34 ns x 256 runs] for K9me3 and 14.848 μs [= 58 ns x 256 runs] for K9ac) to realize a random walk on the l axis. The conformational ensembles obtained by the production run were analyzed after being re-weighed at l = 1 based on a re-weighing scheme, which corresponds to conformations in the physiologically relevant condition (300 K and 1 atm).

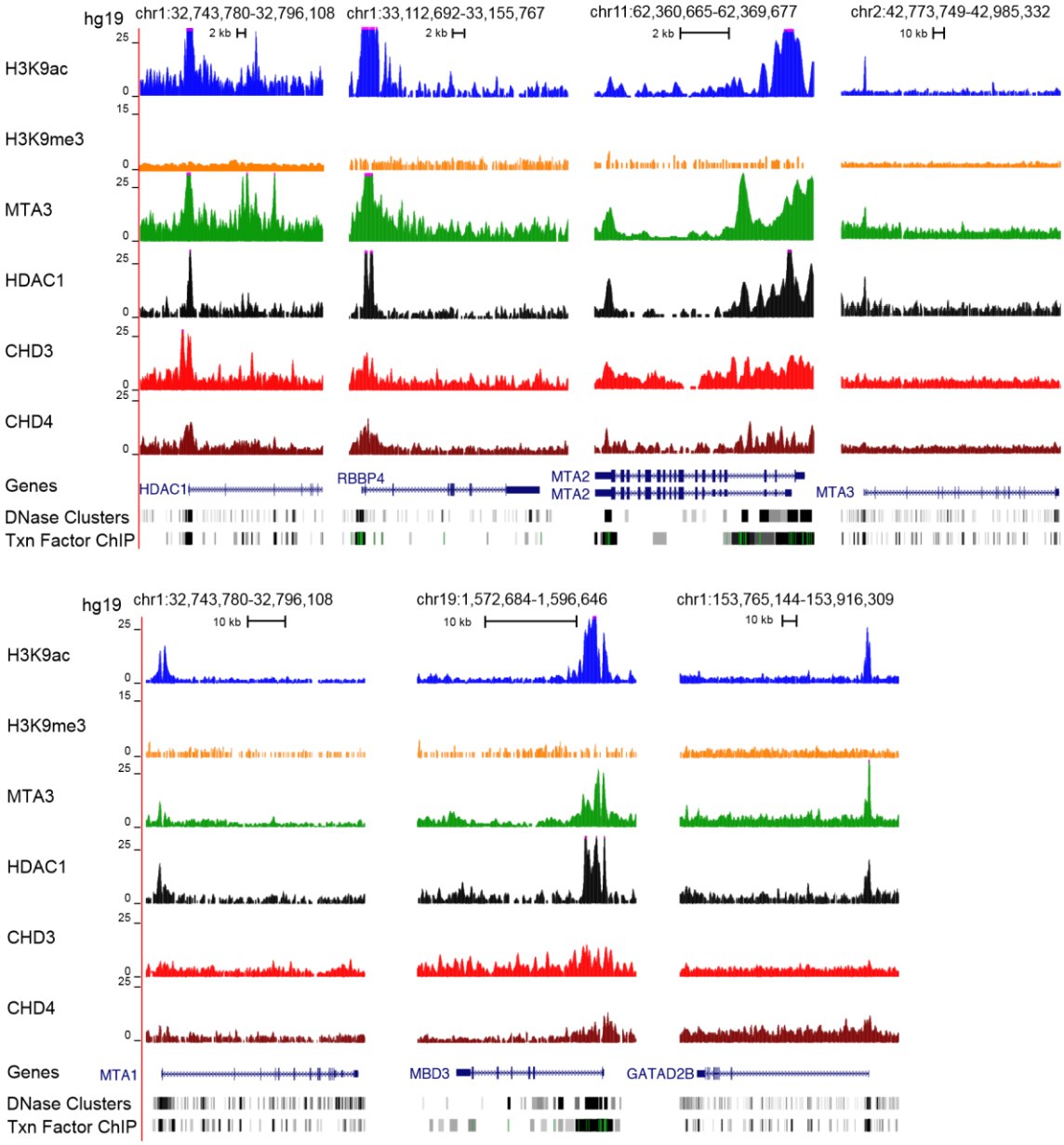
NCP Reaction Rates Measurements

Recombinant WT Xenopus histones H2A, H2B, H4, and H3 (C110A) with cysteine and lysine-to-glutamine mutations in the H3 tail as indicated in Figure 6F, and corresponding NCPs were generated as previously described (Wang and Hayes, 2007). NCPs were reduced in 10 mM DTT for 2 hr at 25ºC and stored at -80ºC after removing DTT by exchanging buffer to 10 mM Tris-HCl (pH 8.0) and 10% glycerol. Samples of 0.8 mM NCPs were reacted with 5.6 mM fluorescein maleimide (FM) at the salt concentration of 150 mM, and the reactions were stopped by adding DTT to a final concentration of 5 mM. Reactions of free H3/H4 tetramers with FM were performed using quenched-flow (Kintec), as previously described [266]. The extent of FM conjugation with H3 was analyzed by running samples on a 15% SDS-PAGE gel. Band intensities in fluorographs were quantified using ImageQuant (MDs). Reaction rate constants and global fits were determined using the standard single-phase exponential equation, $At = A_0(1-e^{-kt})$in GraphPad Prism. K_{conf} was calculated from the free protein and nucleosome reaction rates as previously described [266].

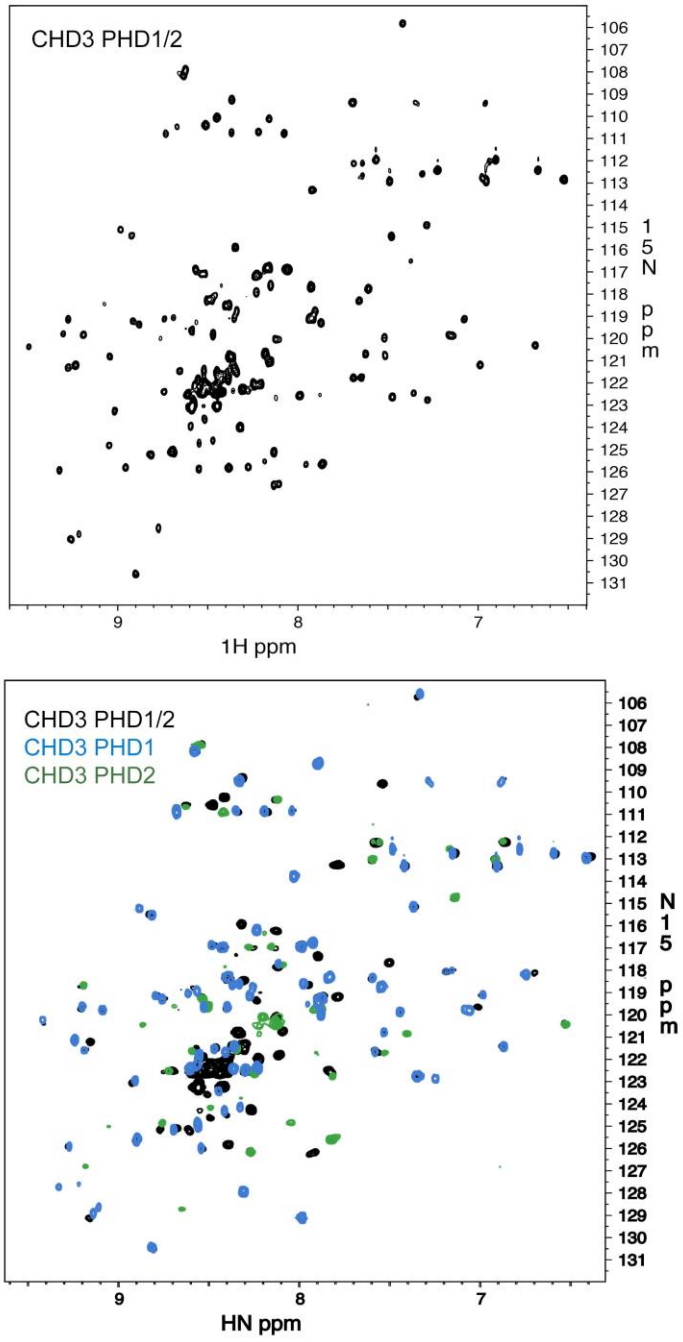
4.5 Supplemental Figures



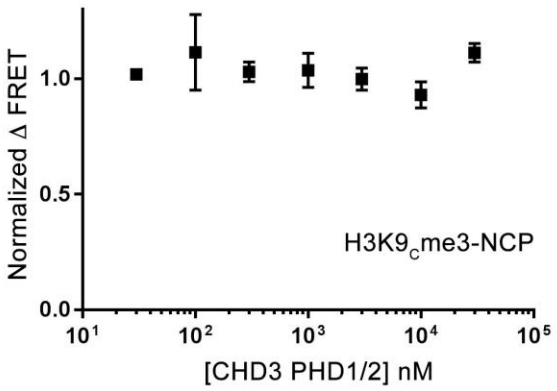
Supplemental Figure S4.1. CHD3 is not detected at promoters of the indicated NuRD-target genes. ChIP-seq profiles of the NuRD components (MTA3, HDAC1, and CHD4) and histone marks (H3K9ac and H3K9me3) in K562 cells, and of CHD3 in LNCaP cells (SRX1181992) are shown. The three lower tracks correspond to the representative RefSeq genes, as well as DNase I Hypersensitivity clusters and Transcription Factor ChIP-seq binding sites, derived from the ENCODE project and the ENCODE Factorbook repository. Related to Figure 4.1.



Supplemental Figure S4.2: Co-localization of CHD3 and other NuRD subunits at promoters of the indicated genes. ChIP-seq profiles of the NuRD components (MTA3, HDAC1, and CHD4) and histone marks (H3K9ac and H3K9me3) in K562 cells, and of CHD3 in LNCaP cells (SRX1181992) are shown. The three lower tracks correspond to the representative RefSeq genes, as well as DNase I Hypersensitivity clusters and Transcription Factor ChIP-seq binding sites, derived from the ENCODE project and the ENCODE Factorbook repository. Related to Figure 4.1.



Supplemental Figure S4.3: $^{1}H,^{15}N$ TROSY spectrum of CHD3 PHD1/2 recorded on the 800 MHz spectrometer (top). Overlays of the $^{1}H,^{15}N$ HSQC spectra of CHD3 PHD1/2, PHD1 and PHD2 in apo-states (bottom). Related to Figure 4.3.



Supplemental Figure S4.4: Normalized change in FRET efficiency of the Cy3-Cy5 labeled H3K9$_C$me3-NCP upon titration of CHD3 PHD1/2. Error bars represent a standard deviation based on three separate experiments. Related to Figure 4.5.

Chapter 5. Molecular Basis for the PZP Domain of BRPF1 Association with Chromatin

Reproduced from: Klein, B. J., Cox, K. L., Jang, S. M., Côté, J., Poirier, M. G., & Kutateladze, T. G. (2020). Molecular Basis for the PZP Domain of BRPF1 Association with Chromatin. Structure, 28(1), 105-110.e3.

This was a study done to characterize the interaction between the BRPF1 plant homeodomain-zinc knuckle-plant homeodomain (PZP) region and chromatin as mononucleosomes. I contributed the fluorescence polarization data.

5.1 Abstract

The assembly of human histone acetyltransferase MOZ/MORF complexes relies on the scaffolding bromodomain plant homeodomain (PHD) finger 1 (BRPF1) subunit. The PHD-zinc-knuckle-PHD module of BRPF1 ($BRPF1_{PZP}$) has been shown to associate with the histone H3 tail and DNA; however, the molecular mechanism underlying recognition of H3 and the relationship between the histone and DNA-binding activities remain unclear. In this study, we report the crystal structure of $BRPF1_{PZP}$ bound to the H3 tail and characterize the role of the bipartite interaction in the engagement of $BRPF1_{PZP}$ with the nucleosome core particle (NCP). We find that although both interactions of $BRPF1_{PZP}$ with the H3 tail and DNA are required for tight binding to NCP and for acetyl-transferase function of the BRPF1-MORF-ING5-MEAF6 complex, binding to extra-nucleosomal DNA dominates. Our findings suggest that functionally active $BRPF1_{PZP}$ might be important in stabilization of the MOZ/MORF complexes at chromatin with accessible DNA.

5.2 Introduction

Bromodomain and PHD finger-containing protein 1 (BRPF1) is a core subunit of the native monocytic leukemic zinc-finger (MOZ) and MOZ-related factor (MORF) acetyltransferase complexes that acetylate histones, particularly lysine 23 of histone H3 (H3K23), and mediate transcriptional programs [288,290, 309]. Genetic studies identified BRPF1 as being essential in the development of brain, axial skeleton, and the hematopoietic system, and a large number of mutations and truncations in human BRPF1 have been linked to intellectual disability, congenital abnormalities, and leukemias [280, 287, 296, 297, 300, 302, 308, 310-312]. In the complexes, BRPF1 functions as an adaptor protein that interacts with all other subunits, including the catalytic MOZ/MORF subunit, inhibitor of growth 5 (ING5), and MYST/Esa1-associated factor 6 (MEAF6), and therefore is required for the tetrameric assembly, catalytic activity, and substrate specificity of these complexes [291, 309].

BRPF1 is a large multi-modular protein containing the MOZ/MORF-binding domain (I), a short motif implicated in the interaction with ING5 and MEAF6 (II) [294, 305], and three histone-binding modules, also known as epigenetic readers characterized by different specificities toward post-translational modifications (PTMs) in histone proteins (Figure 5.1A). The BRPF1 bromodomain (BD) associates with various mono- and poly-acetyllysine-containing sequences in histones H4 and H3 (H4/H3Kac), whereas the PWWP domain recognizes trimethylated K36 of histone H3 (H3K36me3) [296, 303, 307]. The central region of BRPF1 harbors two PHD fingers closely linked through a single zinc knuckle, termed the PZP (PHD-zinc-knuckle-PHD) domain ($BRPF1_{PZP}$). Recent biochemical studies have shown that $BRPF1_{PZP}$ binds to the unmodified histone H3 tail and is also capable of associating with DNA [292]. The atomic-resolution structures of the BD and PWWP domains of BRPF1 in complex with their histone ligands provide information that is vital to our understanding of how BRPF1 and the MOZ/MORF complexes localize to genomic regions enriched in H4/H3Kac and H3K36me3 [296, 303, 307]. The structures have also been instrumental in the development of numerous small-molecule inhibitors for BD of BRPF1, which emerged as one of the promising therapeutic targets in leukemias [277, 289, 313, 314]; however, the structural basis underlying recognition of H3 by

BRPF1$_{PZP}$ and the relationship between its histone- and DNA-binding functions remain not well characterized.

Here, we report the molecular mechanism by which BRPF1$_{PZP}$ targets the histone H3 tail and DNA and assess contributions of the two binding events to the association of BRPF1$_{PZP}$ with the nucleosome core particle (NCP). We find that while both interactions of BRPF1$_{PZP}$ with the H3 tail and DNA are required for tight binding to NCP and for acetyltransferase function of the BRPF1-MORF-ING5-MEAF6 complex, the interaction with extra-nucleosomal DNA predominates.

5.3 Results and Discussion

To determine the molecular basis for the histone H3 recognition by BRPF1$_{PZP}$, we generated a chimeric construct that contains residues 1–12 of H3 fused to residues 271–454 of BRPF1 through a short GSGSS linker. The ^{1}H,^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the uniformly ^{15}N-labeled fused H3-PZP construct overlaid well with the spectrum of isolated BRPF1$_{PZP}$ collected in the presence of a 5-fold excess of the H3 (residues 1–12) peptide, confirming that the linked and unlinked complexes adopt similar structures in solution (Figure S5.1). The fusion protein was crystallized, and the structure of the H3-bound BRPF1$_{PZP}$ was determined to a 2.2-Å resolution (Figure 5.1B). The structure shows a saddle-like globular fold composed of five zinc-binding clusters. The Ala1-Thr3 fragment of the H3 tail is bound in a shallow groove of the first PHD finger (PHD1) of BRPF1$_{PZP}$. The N-terminal amino group of Ala1 of H3 is restrained by two hydrogen bonds with the backbone carbonyl groups of P311 and G313 of the protein, whereas the methyl group of Ala1 fits in a hydrophobic cavity formed by L291, I310, and W315 (Figures 5.1B and 5.1C). The guanidino moiety of Arg2 is constrained via two hydrogen bonds, one to the side-chain carboxyl oxygen of D294 and another to the side-chain amide oxygen of N297. The backbone amide nitrogen of F292 donates a hydrogen bond to the backbone carbonyl oxygen of Arg2, and the methyl group of Thr3 occupies the same hydrophobic cavity as the methyl group of Ala1 of H3.

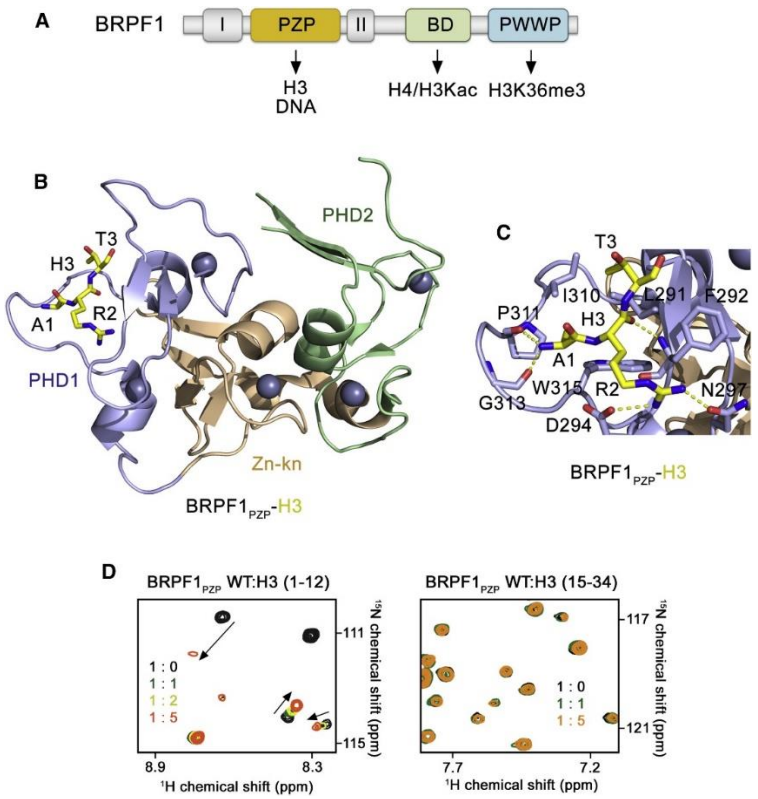


Figure 5.1: BRPF1$_{PZP}$ recognizes the N-terminal part of H3 tail**.** (a) BRPF1 domain architecture. (b) The crystal structure of the H3-bound BRPF1$_{PZP}$ is depicted in a ribbon diagram with PHD1, Zn-kn and PHD2 colored blue, wheat, and green, respectively. The Ala1-Thr3 fragment of the H3 tail is shown as yellow sticks, and the zinc ions are grey spheres. (c) Close up view of the H3 binding pocket. Hydrogen bonds between residues of H3 tail and PHD1 indicated by yellow dash lines. (d) Superimposed ^{1}H,^{15}N HSQC spectra of BRPF1$_{PZP}$ collected in the presence of increasing amounts of H3 peptide (residues 1–12 on the left, and residues 15–34 on the right). The spectra are color-coded according to the protein-peptide ratio (inset). See also Figures S5.1 and S5.2.

Recent studies of AF10, a co-factor of the H3K79-specific methyltransferase DOT1L, show that the PZP domain of AF10 recognizes a middle part of the histone H3 tail, specifically residues Ala21–Lys27 of H3 (Chen et al., 2015). To test whether this binding is conserved in BRPF1, we carried out ^{1}H,^{15}N HSQC titration experiments using ^{15}N-labeled BRPF1$_{PZP}$. As expected, the H3 peptide (residues 1–12 of H3) induced large chemical-shift perturbations (CSPs) in the BRPF1$_{PZP}$ spectrum, indicating binding. However, no CSPs were observed upon titration of the H3 peptide (residues 15–34 of H3), implying that BRPF1$_{PZP}$ does not bind to this part of H3 (Figure 5.1D). Furthermore, the long H3 peptide (residues 1–31 of H3) caused CSPs in BRPF1$_{PZP}$ almost

identical to CSPs caused by the short H3 peptide (residues 1–12), supporting the finding that $BRPF1_{PZP}$ recognizes the far N-terminal but not middle region of the H3 tail (Figure S5.2).

In addition to recognizing H3, $BRPF1_{PZP}$ was also shown to bind DNA; however, the relationship between these functions is not well understood. Electrostatic surface potential of the $BRPF1_{PZP}$:H3 complex reveals that the positively charged residues of PZP involved in DNA binding, particularly K383, K390, and R392, are clustered in the second PHD finger (PHD2) and a zinc knuckle, whereas the negatively charged residues in the first PHD finger (PHD1) form the binding site for H3 (Figure 5.2A). Such separation of the binding sites suggests that the two interactions are likely independent of each other. We generated NCP using a 207-bp DNA (NCP_{207}) in which 147-bp Widom 601 DNA is flanked by a 30-bp linker DNA on either side and internally labeled with fluorescein 27 bp in from the 5' end, and tested binding of $BRPF1_{PZP}$ to NCP_{207} by electrophoretic mobility shift assay (EMSA) and fluorescence polarization. Both wild-type (WT) $BRPF1_{PZP}$ and mutants impaired in either H3 binding or DNA binding, including D294K that lost its ability to interact with H3 peptide (Figure 5.2B) and K383E/K390E/R392E (KKR) mutant incapable of binding to DNA [292], were assayed. NCP_{207} was incubated with increasing amounts of $BRPF1_{PZP}$, and the reaction mixtures were resolved on a 5% native polyacrylamide gel (Figure 5.2C). A gradual increase in amount of added WT $BRPF1_{PZP}$ resulted in a shift of the NCP_{207} band, indicative of formation of the $BRPF1_{PZP}$:NCP_{207} complex. However, this shift was delayed when $BRPF1_{PZP}$ D294K mutant was used, implying that interaction with H3 is essential for high-affinity binding. In contrast, no

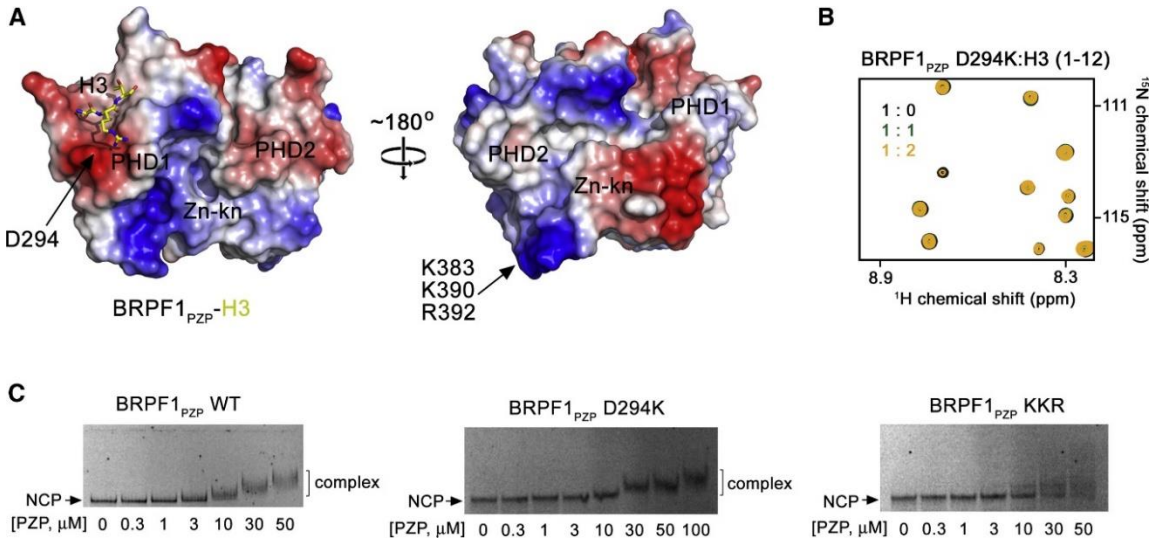


Figure 5.2: BRPF1$_{PZP}$ binds to nucleosomes in a bivalent manner. (a) Electrostatic surface potential of BRPF1$_{PZP}$ in the complex with blue and red colors representing positive and negative charges, respectively. The H3 tail is shown in yellow sticks. (b) Superimposed ^{1}H,^{15}N HSQC spectra of BRPF1$_{PZP}$ D294K mutant recorded in the presence of increasing amount of H3 peptide (residues 1–12). The spectra are color-coded according to the protein-peptide ratio (inset). (c) EMSA with NCP$_{207}$ incubated with increasing amounts of the indicated BRPF1$_{PZP}$ proteins. Amount of each protein mixed with 5 nM NCP$_{207}$ is shown below each gel image.

stable complex with NCP$_{207}$ was formed by BRPF1$_{PZP}$ KKR mutant, indicating that binding to DNA is required for the association with NCP$_{207}$. These results were substantiated by measuring $S_{1/2}$ in fluorescence anisotropy assays. Titration of WT BRPF1$_{PZP}$ against NCP$_{207}$ yielded $S_{1/2}$ of 50 mM for the BRPF1$_{PZP}$:NCP$_{207}$ complex formation (Figure 5.3A). Binding of the BRPF1$_{PZP}$ D294K mutant to NCP$_{207}$ was 2-fold weaker ($S_{1/2}$ = 100 mM), and binding of the BRPF1$_{PZP}$ KKR mutant was undetectable (Figures 5.3B and 5.3C). The association of WT BRPF1$_{PZP}$ with the nucleosome containing a 147-bp Widom 601 DNA sequence (NCP$_{147}$) was also undetectable, confirming that BRPF1$_{PZP}$ prefers extranucleosomal DNA (Figures 5.3D and 5.3E). This preference for the linker DNA might have a significant implication and suggests that functional BRPF1$_{PZP}$ could stabilize the MORF complexes at chromatin regions with accessible DNA, such as euchromatin.

To assess the contribution of binding to H3 and DNA by BRPF1$_{PZP}$ to catalytic function of the MORF complex, we produced the MORF complexes by co-transfecting full-length HA-BRPF1 with FLAG-MORFN1-716, FLAG-ING5, and FLAG-MEAF6 in 293T cells and purified

the complexes by immunoprecipitation. The HAT activity of the MORF complexes, containing HA-BRPF1, WT, or D294A and BRPF1$_{PZP}$D359-450 (amino acids [aa] 359–450, including K383, K390, and R392 of BRPF1 are deleted) mutants, were measured on human free histones and short oligonucleosomes. The MORF complex containing WT BRPF1 subunit showed a strong HAT activity (Figures 5.3F and S5.3). This activity was decreased 3-fold for the BRPF1 D294A mutant defective in H3 binding and 5-fold for the BRPF1$_{PZP}$D359-450 mutant defective in DNA binding. Together, these data support the notion that concurrent binding of BRPF1$_{PZP}$ to H3 and DNA is required for proper enzymatic activity of the BRPF1-MORF complex, with the latter contributing to a greater degree (Figure 5.3E).

BRPF1$_{PZP}$ has expanded a subset of epigenetic readers capable of binding histones and DNA. The dual engagement increases affinity of these readers in the context of chromatin owing to the avidity effect, but relative contribution of the contacts with histones and DNA vary among the readers. For example, interaction of the PWWP domain of LEDGF/PSIP1 with DNA enhances a markedly weak (millimolar-range) binding of this reader to histone H3K36me3 by four orders of magnitude [283, 306], whereas a relatively weak DNA binding of Tudor, YEATS, and bromodomain augments association with NCPs by a few-fold [286, 298, 301]. In this study, we show that engagement of BRPF1$_{PZP}$ with the histone H3 tail and DNA, especially extra-nucleosomal DNA, is required for tight binding to and acetylation of NCP, with the DNA-binding activity being more essential. These in vitro data are in agreement with previously reported findings that while the HAT activity of the MOZ complex requires both BRPF1$_{PZP}$ functions, the DNA binding rather

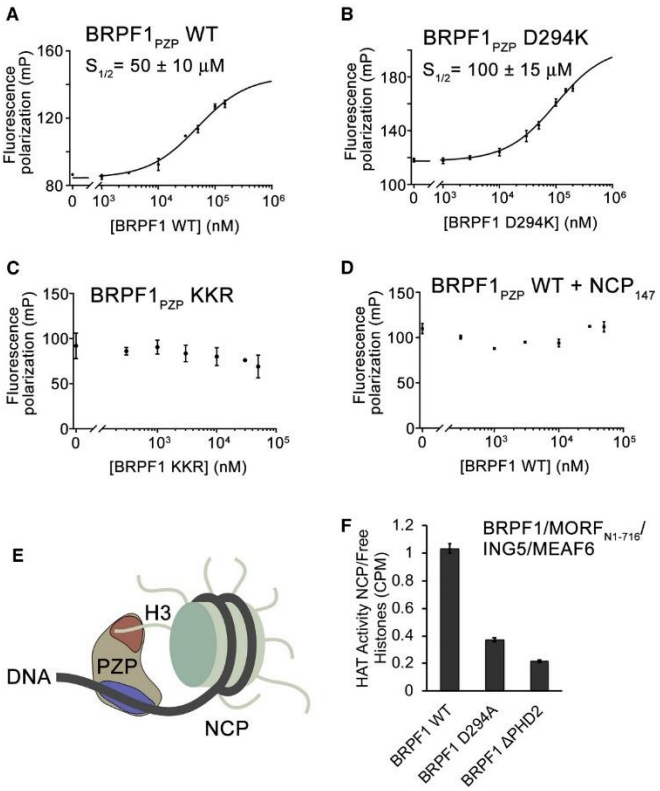


Figure 5.3: Binding of BRPF1$_{PZP}$ to DNA predominates. (a-d) Binding curves obtained for the interactions of indicated BRPF1$_{PZP}$ proteins with NCP$_{207}$ (a-c) or NCP$_{147}$ (d) in fluorescence polarization assays. Error bars are SD based on three separate experiments. (e) A model of the bivalent interaction of BRPF1$_{PZP}$ with histone H3 of the nucleosome and extra-nucleosomal DNA. (f) Ratio of HAT activities of the WT and mutant BRPF1-MORF-ING5-MEAF6 complexes on nucleosomes versus free histones. HA-BRPF1, wild type (WT) or mutants (D294A and BRPF1$_{PZP}$Δ359–450), FLAG-MORFN1–716 (aa 1–716 of MORF, containing the catalytic MYST domain), FLAG-ING5 and FLAG-MEAF6 were used to transfect 293T cells. Error bars indicate the range from duplicate samples. Background counts obtained with fractions from mock transfections were subtracted (~150 cpms). See also Figure S5.3.

than histone binding is critical for the association of the complex with chromatin in vivo [292, 294]. The lesser contribution of histone binding could be due to a competitive interaction of the histone tail with DNA and, therefore, tail inaccessibility. The intra-nucleosomal histone H3 tail-DNA contacts have been shown to reduce the accessibility of unmodified H3 tails compared with free H3 peptides by up to a factor of ~10 in physiologically relevant conditions [285], but the strength of these contacts can be modulated by PTMs in histones [304]. It will be interesting in future studies to explore the effect of PTMs, particularly acetylation of distal lysine residues in the H3 tail, on bipartite association of BRPF1$_{PZP}$ with NCPs.

5.4 Materials and Methods

DNA Cloning and Protein Purification

The PZP domain of BRPF1 (aa 271–454) was cloned into a pDEST15 vector with the N-terminal GST tag and TEV cleavage site. The mutant $BRPF1_{PZP}$ (D294K and K383E/K390E/R392E) constructs were generated using the Stratagene QuickChange XL Site Directed Mutagenesis kit. The sequences were confirmed by DNA sequencing. All proteins were expressed in Escherichia coli Rosetta-2 (DE3) pLysS cells grown in minimal media supplemented with $^{15}NH_4Cl$ (Sigma) or NH_4Cl (for unlabeled proteins) and 75 M $ZnCl_2$. Protein production was induced with 0.5 – 1.0 mM IPTG for 18 h at 16°C. Bacteria were harvested by centrifugation and lysed by sonication in buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.05% (v/v) Nonident P 40, 5 mM dithiothreitol (DTT), 50 M $ZnCl_2$, 5 mM $MgCl_2$, and DNase). GST-fusion proteins were purified on glutathione agarose 4B beads (Thermo Fisher Sci) and the GST-tag was cleaved with tobacco etch virus (TEV) protease. Proteins were further purified by size exclusion chromatography (SEC) and concentrated in Millipore concentrators (Millipore).

X-Ray Crystallography

For structural studies the H3-GSGSS-$BRPF1_{PZP}$ construct (aa 1–12 of histone H3, a GSGSS linker, and aa 271–454 of BRPF1) was cloned into a pDEST15 vector with the N-terminal GST tag and TEV cleavage site. The linked protein was produced as above. Following cleavage with TEV protease and further purification by SEC, the linked H3-PZP protein was concentrated to ~ 6 mg/mL in buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT) for crystallization. Crystals were grown using sitting-drop diffusion method at 18°C by mixing 600 nL of protein with 600 nL of well solution composed of 0.2 M Lithium sulfate, 0.1 M Tris-HCl pH 8.5, 40% (v/v) PEG400, and 0.01M Praseodymium(III) acetate hydrate. Crystals were cryoprotected with 30% (v/v) glycerol. X-ray diffraction data were collected from a single crystal on the Anschutz Medical Campus X-ray crystallography core facility Rigaku Micromax 007 high-frequency microfocus X-ray generator equipped with a Pilatus 200K 2D area detector. Indexing and scaling were completed using HKL3000 (Minor et al., 2006). The phase solution was solved with Phenix.phaser

[275] using molecular replacement and the $BRPF1_{PZP}$ (PDB ID: 5ERC) structure as the model. Model building was carried out with Coot [284], and refinement was performed with Phenix.refine. The final structure was validated with MolProbity [279]. Crystallographic statistics for the H3-bound $BRPF1_{PZP}$ structure are shown in Table 1.

NMR experiments

Nuclear magnetic resonance (NMR) experiments were performed at 298 K on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe. The $^{1}H,^{15}N$ HSQC spectra of 0.1–0.2 mM uniformly 15N-labeled WT or mutant $BRPF1_{PZP}$ in 50 mM Tris-HCl pH 7.5 buffer, supplemented with 150 mM NaCl, 5 mM dithiothreitol (DTT), and 8% D2O were collected in the presence of increasing amount of H3 (aa 1–12, 15–34, or 1–31) peptides (synthesized by Synpeptide). NMR data were processed and analyzed with NMRPipe and NMRDraw as previously described [293].

Nucleosome assembly

Human H2A, H2B, H3.2, and H4 histone proteins were expressed in Escherichia coli BL21 (DE3) pLysS cells, separated from inclusion bodies and purified using size exclusion and ion exchange chromatography, as described previously [284]. Histones were then mixed together in 7 M guanidine HCL, 20 mM Tris-HCl pH 7.5, and 10 mM dithiothreitol in appropriate molar ratios and refolded into octamer by slow dialysis into 2 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, and 2 mM β-mercaptoethanol. The octamer was purified from tetramer and dimer by SEC. Octamer was then mixed with 75% excess of DNA in 2 M NaCl, 5 mM Tris pH 8.0 and 0.5 mM EDTA, and NCPs were reconstituted from octamer plus DNA by slow desalting dialysis into 5 mM Tris pH 8.0 and 0.5 mM EDTA. Finally, the NCPs were separated from free DNA via sucrose gradient purification. DNAs used for the NCP assembly were 147 bp 601 Widom DNA fluorescein-labeled on the 5' end (for NCP_{147}), and 207 bp DNA (147 bp 601 DNA flanked with 30 bp linker DNA on either side and internally labeled with fluorescein 27 bp in from the 5' end) (for NCP_{207}).

Fluorescence polarization

Fluorescence polarization measurements were carried out by mixing increasing amounts of $BRPF1_{PZP}$, WT or mutants, with 5 nM NCP_{207} or NCP_{147} in 75 mM NaCl, 25 mM Tris-HCl pH 7.5, 0.00625% Tween20, and 5 mM dithiothreitol in a 30 µL reaction volume. The samples were loaded into a Corning round bottom polystyrene plate and allowed to incubate at 4°C for 30 min. The polarization measurements were acquired with a Tecan infinite M1000Pro plate reader by exciting at 470 nm and measuring polarized emission at 519 nm with 5 nm excitation and emission bandwidths. The fluorescence polarization was calculated from the emission polarized parallel and perpendicular to the polarized excitation light as described previously [304]. The data were then fit to a non-cooperative binding isotherm to determine $S_{1/2}$. The $S_{1/2}$ values were averaged over three separate experiments with error calculated as the standard deviation between the runs.

Electrophoretic Mobility Shift Assay

EMSAs were performed by mixing increasing amounts of $BRPF1_{PZP}$, WT or mutants, with 5 nM NCP_{207} in 75 mM NaCl, 25 mM Tris-HCl pH 7.5 buffer, supplemented with 0.00625% Tween20, 10% glycerol and 5 mM dithiothreitol in a 12 µL reaction volume. Each sample was incubated at 4°C for 15 min and then loaded onto a 5% native polyacrylamide gel. Electrophoresis was performed in 0.3x Tris-borate-EDTA (TBE) at 300 V for 90 min. Fluorescein fluorescence images were acquired with a Typhoon Phosphor Imager.

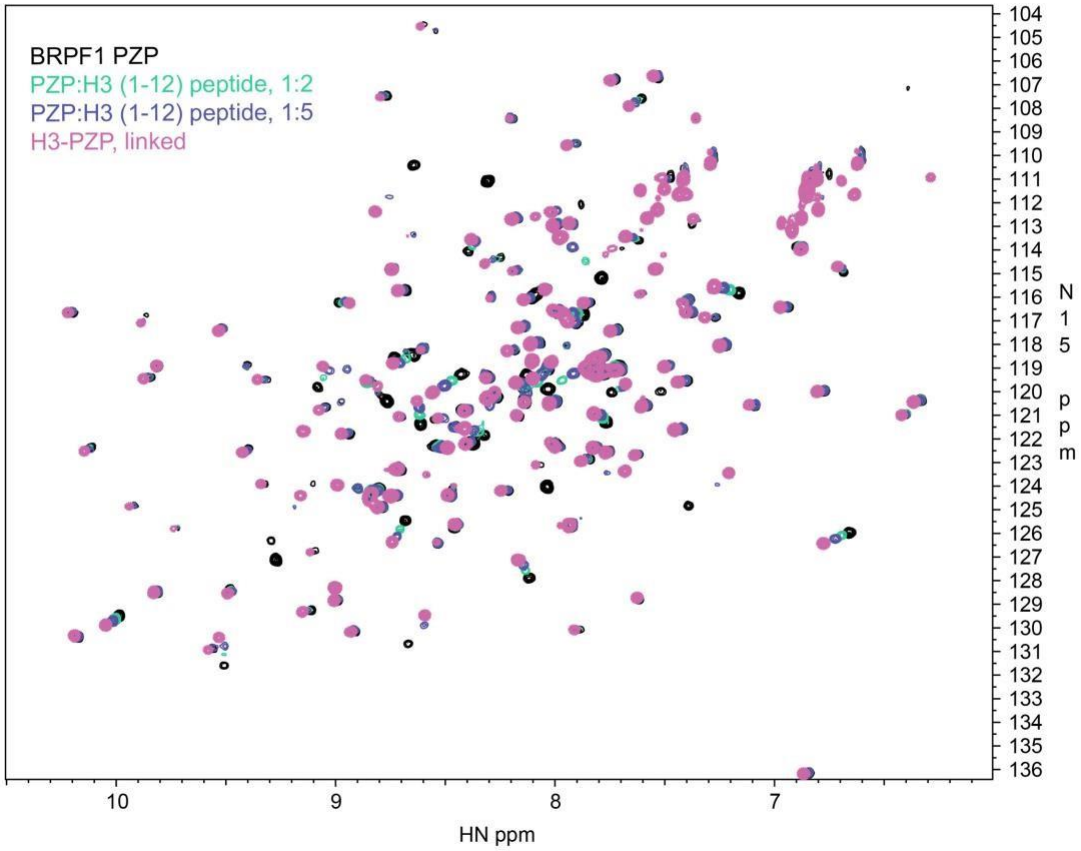
Purification of wt and mutant BRPF1-MORF-ING5-MEAF6 complexes and HAT assays

Plasmids of full length WT HA-BRPF1, HA-BRPF1 D294A mutant or HA-$BRPF1_{PZP}$Δ359–450 mutant, together with FLAG-MORFN1–716, FLAG-ING5 and FLAG-MEAF6 [292] were used to transfect 293T cells by the calcium phosphate method. Cells were harvested 70 h post-cotransfection and nuclear extracts were prepared as previously described [281]. Purification of HA-BRPF1 complexes after co-transfection was performed essentially as previously described

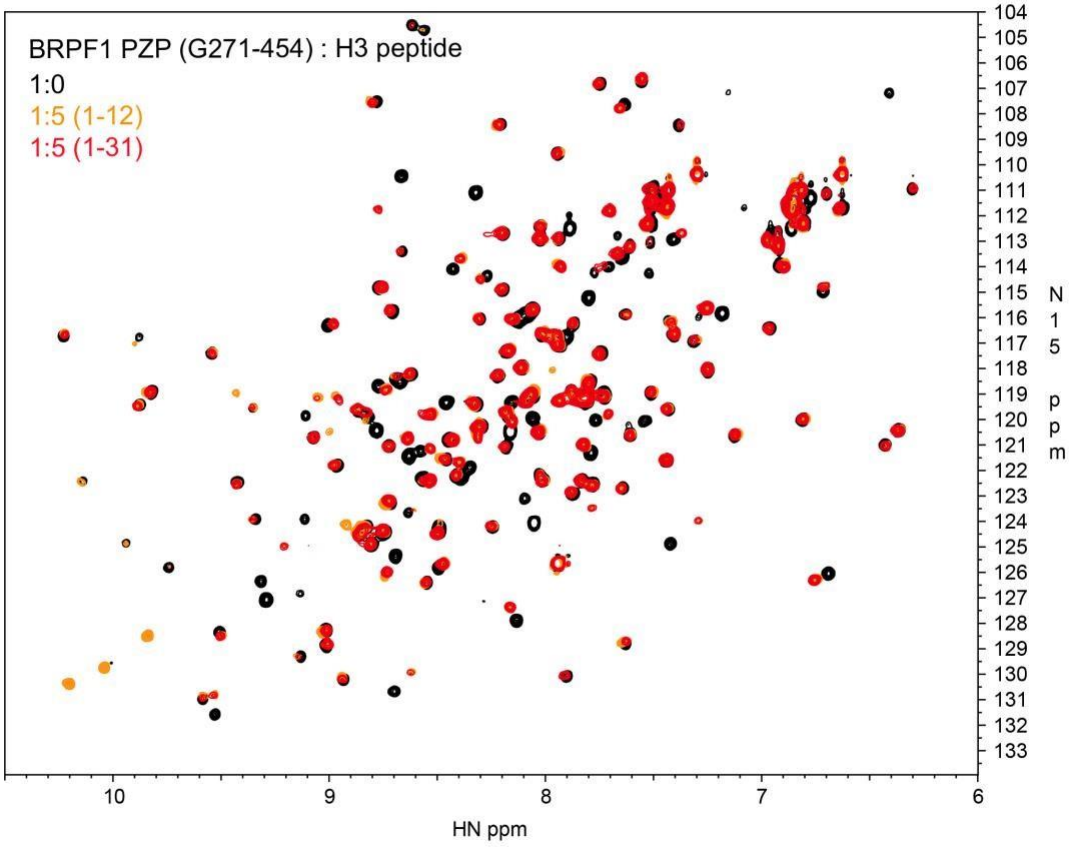
[276, 295]. Briefly, anti-HA immunoprecipitation/elution using anti-HA agarose beads (Roche ref:118 150 16001) and 3xHA peptides was followed by anti-Flag immunoprecipitation/elution using FLAG M2 agarose beads (Sigma) and 3xFLAG peptides.

Acetyltransferase activity of the purified complexes was measured with 0.125 Ci of ^{3}H labeled Ac-CoA (2.1 Ci/mmol; PerkinElmer Life Sciences). The HAT reactions were performed in a volume of 15 l using 0.5 g of free histones and short oligonucleosomes (purified from HeLa cells) as substrates, in HAT buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 10 mM sodium butyrate, 5% glycerol, 0.1 mM EDTA, 1 mM dithiothreitol for 30 minutes at 30°C. The reactions were then captured on P81 filter paper, the free ^{3}H-labeled Ac-CoA was washed away, and the paper was analyzed using Liquid Scintillation.

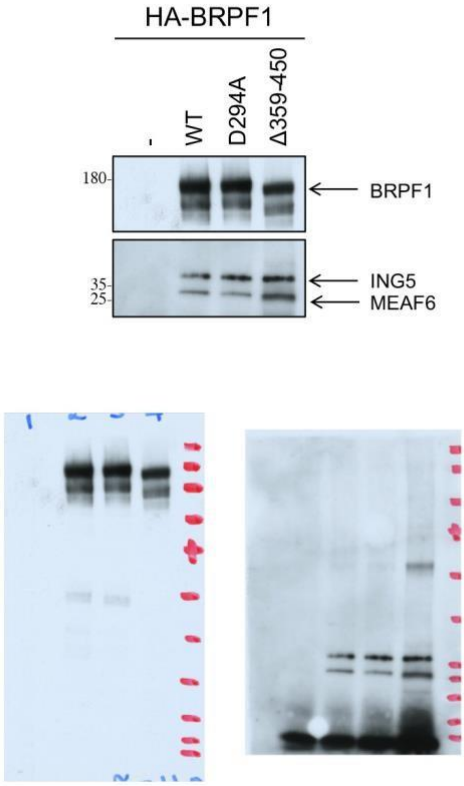
5.5 Supplementary Figures



Supplementary Figure S5.1: Superimposed $^{1}H,^{15}N$ HSQC spectra of the linked H3-PZP construct (magenta) and the isolated $BRPF1_{PZP}$ domain in the apo-state (black), or in the presence of a 2-fold excess of H3 (1-12) peptide (green) or 5-fold excess of H3 (1-12) peptide (blue). Related to Figure 5.1.



Supplementary Figure S5.2: Superimposed $^{1}H,^{15}N$ HSQC spectra of $BRPF1_{PZP}$ without (black) and with H3 peptides (residues 1-12, orange and 1-31, red). Related to Figure 5.1.



Supplementary Figure S5.3: BRPF1 complexes used in HAT assays. BRPF1 complexes purified from HEK293T cells after transient co-transfection of HA-BRPF1 with FLAG-MORF$_{N1-716}$, FLAGING5 and FLAG-MEAF6. Western blot analysis was performed with HA and FLAG antibodies and represents their normalization for HAT assays in Figure 3F. The first lane is from a mock transfection without HA-BRPF1. Anti-FLAG M2 conjugated to horseradish peroxydase (HRP; Sigma) was used at a 1:10,000 dilution, and the immunoblots were visualized using a Western Lightning plus-ECL reagent (Perkin-Elmer). Anti-HA-HRP (clone 3F10, Roche) was used at a 1:2,000 dilution. Uncropped images are shown below. Related to Figure 5.3.

Chapter 6. Molecular Mechanism of the MORC4 ATPase Activation

Reproduced from: Tencer, A. H., Cox, K. L., Wright, G. M., Zhang, Y., Petell, C. J., Klein, B. J., … Kutateladze, T. G. (2020). Molecular mechanism of the MORC4 ATPase activation. *Nature Communications,* 11(1), 1–13.

This study was done on MORC4 meant to shed light on the cellular functions of MORC4, the least well-characterized member of the MORC family of chromatin-binding proteins at the time. I was responsible for the FRET and fluorescence polarization data.

6.1 Abstract

Human Microrchidia 4 (MORC4) is associated with acute and chronic pancreatitis, inflammatory disorders and cancer but it remains largely uncharacterized. Here, we describe the structure–function relationship of MORC4 and define the molecular mechanism for MORC4 activation. Enzymatic and binding assays reveal that MORC4 has ATPase activity, which is dependent on DNA-binding functions of both the ATPase domain and CW domain of MORC4. The crystal structure of the ATPaseCW cassette of MORC4 and mutagenesis studies show that the DNA-binding site and the histone/ATPase binding site of CW are located on the opposite sides of the domain. The ATPase and CW domains cooperate in binding of MORC4 to the nucleosome core particle (NCP), enhancing the DNA wrapping around the histone core and impeding binding of DNA-associated proteins, such as transcription factors, to the NCP. In cells, MORC4 mediates formation of nuclear bodies in the nucleus and has a role in the progression of S-phase of the cell cycle, and both these functions require CW and catalytic activity of MORC4. Our findings highlight the mechanism for MORC4 activation, which is distinctly different from the mechanisms of action observed in other MORC family members.

6.2 Introduction

Microrchidia 4 (MORC4) is a poorly characterized member of the new family of CW-type zinc finger nuclear proteins. MORC4 has been associated with acute and chronic pancreatitis and inflammatory bowel disorders, including Crohn's disease and ulcerative colitis, and more recently was found to be overly expressed in breast cancer cells and in diffuse large B cell lymphoma [315-322]. Analysis of breast cancer tissues shows that MORC4 is negatively regulated by microRNAs, including miR-193b-3p, and elevated MORC4 levels are linked to poor survival [323]. Despite growing evidence suggesting an essential role of MORC4 in normal physiological and pathological processes, how this protein functions at molecular and cellular level is not well understood.

Members of the MORC family are characterized by their gyrase, Hsp90, histidine kinase, and MutL (GHKL)-type ATPase domain, which is also present in several chromatin-modifying enzymes [325,325]. Although the ability of MORC4 to hydrolyze ATP has not been investigated, the homologous MORC proteins, such as MORC2 and MORC3 are enzymatically active. MORC3 has been shown to exist in an autoinhibited state and is activated through binding to methylated histone mark H3K4me3, associated with transcriptionally active chromatin [326-330]. Conversely, MORC2 does not recognize H3K4me3 and is necessary for the human silencing hub (HUSH)-dependent silencing of transgenes integrated at chromatin loci with the methylated histone mark H3K9me3 [331,332]. The ATPase domain of MORCs is followed by a CW-type zinc finger and one or more coiled-coil regions (Fig. 6.1a). Recognition of H3K4me3 by MORC3 is mediated by the CW domain, and this interaction regulates the catalytic activity of MORC3 and is required for the MORC3 recruitment to chromatin and formation of liquid–liquid phase separation (LLPS) droplets in the nucleus [327,329,333,344]. While we have begun understanding the markedly different mechanisms of action and some biological roles of the individual MORC2 and MORC3 proteins, virtually nothing is known about the function of MORC4.

Here, we describe the unique molecular mechanism for MORC4 activation. We show that MORC4 is an ATPase with an intrinsic DNA-dependent but histone-independent enzymatic activity. The ATPase and CW domains of MORC4 cooperate in binding of MORC4 to the NCP,

impeding the association of transcription factors with the NCP. Our cell data reveal two biological functions of MORC4: it regulates the formation of nuclear bodies (NBs) in the nucleus and plays a role in the cell cycle S-phase progression with the ATPase activity and chromatin binding being necessary for both functions.

6.3 Results

6.3.1 MORC4 is an ATPase

To determine whether MORC4 is catalytically active, we produced the ATPase domain of human MORC4 and examined its ability to hydrolyze ATP in enzymatic assays (Fig. 6.1a–c). Following incubation of the ATPase domain with ATP, 2-amino-6-mercapto-7-methylpurine ribonucleoside and nucleoside phosphorylase, the release of inorganic phosphate generated by the hydrolysis of ATP to ADP was monitored by measuring an increase in absorbance at 360 nm. We found that the MORC4 ATPase domain itself possesses little ATP hydrolyzing activity, however addition of 147 base pair (bp) 601 Widom DNA at a 1:1 molar ratio led to a ~4-fold increase in the rate of ATP hydrolysis, indicating that the catalytic activity of the ATPase domain is DNA-dependent (Fig. 6.1b). A similar increase in the ATP hydrolysis rate was observed upon adding 601 DNA to the ATPase domain linked to the CW domain of MORC4 in the ATPaseCW cassette (Fig. 6.1a, c). Neither methylated histone peptide H3K4me3 nor unmodified H3 peptide (both aa 1–12 of H3) stimulated a further increase in the rate of ATP hydrolysis, implying that the enzymatic activity is largely histone independent (we discuss binding of CW to H3 below). The nucleosome core particle (NCP) also stimulated the catalytic activity of the ATPaseCW cassette (Supplementary Fig. S6.1). Together, these data demonstrate that the ATPase activity is a function of MORC4, which requires the presence of DNA.

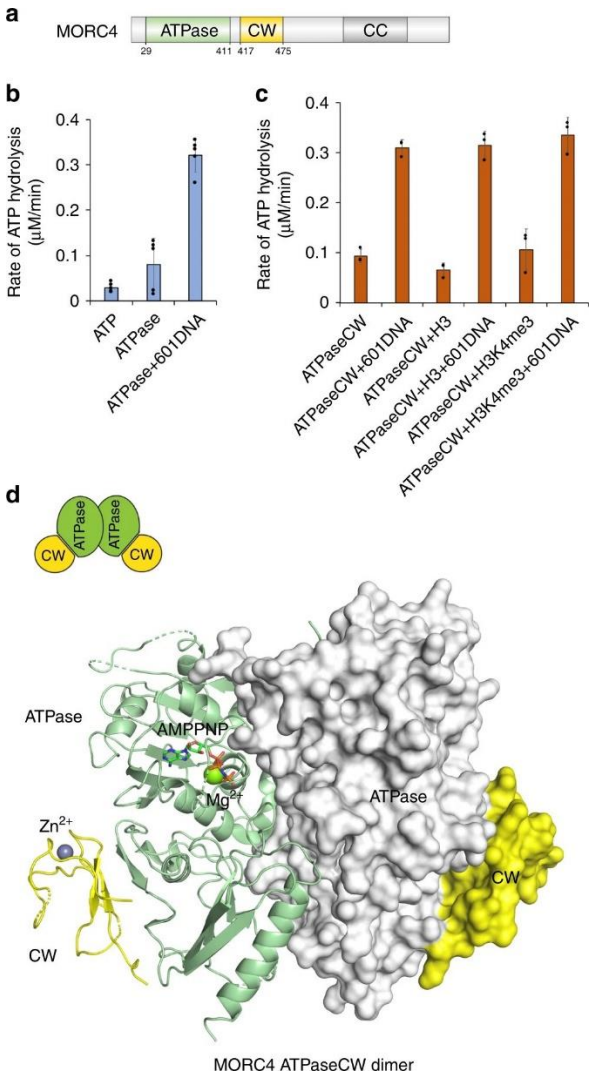


Figure 6.1: MORC4 is an ATPase. (a) MORC4 domain architecture. The ATPase and CW domains are colored green and yellow, respectively. The ATPaseCW cassette construct contains aa 29–486 of MORC4 (Q8TE76). (b, c) Rates of ATP hydrolysis by the ATPase domain of MORC4 (b) and the ATPase-CW cassette of MORC4 (c). Data are represented as mean values +/– S.D. of at least three independent experiments (n ≥ 3). Source data are provided in a Source Data file. (d) The crystal structure of the dimeric MORC4 ATPase-CW/AMPPNP complex. In protomer A, the ATPase domain and the CW domain are shown in a ribbon diagram and colored green and yellow, respectively. In protomer B, the surface representation of the ATPase domain (white) and the CW domain (yellow) is shown. The magnesium (green) and zinc (gray) atoms are shown as spheres. The AMPPNP molecule is in stick representation. Dashed lines represent residues and loops that are not defined by electron density.

6.3.2 Structure of the MORC4 ATPaseCW cassette

To gain mechanistic insight into the MORC4 activation process, we crystallized ATPaseCW (aa 29–486 of human MORC4) in complex with the non-hydrolyzable ATP analog, adenylyl-imidodiphosphate (AMPPNP), collected and analyzed X-ray crystallographic data, and refined the structure to 2.9 Å resolution (Fig. 6.1d and Supplementary Table 1). The structure reveals a symmetric parallel homodimer with each ATPaseCW protomer coordinating one molecule of AMPPNP and one Mg2+ ion. The ATPase module folds into the GHKL-type domain, consisting of a mixture of α-helices and β-strands, and is involved in the formation of a large dimer interface, which was also observed in the respective structures of MORC3 and MORC2 [328,329,331]. Two zinc-binding CW domains tightly pack against the ATPase domains and are positioned on the sides of the ATPase domains opposite to the dimer interface and are far apart (colored yellow in Fig. 6.1d).

Each CW domain interacts mainly with the loops between α8 and β10 of the ATPase domain (Supplementary Fig. S6.2). The residues P279, K289, and K297 of the ATPase domain form a number of polar and electrostatic contacts with the Q424, W426, Q428, D430, W435, and A449 residues of the CW domain. In addition, the hydrophobic side chain of M294 is caged between the aromatic rings of W426 and W435 of CW, and the side chain of R69 of the ATPase domain is restrained by hydrogen bonds with the backbone carbonyl groups of D430 and E431 and the carboxyl group of E431 of CW.

Strikingly, while the MORC4 ATPaseCW cassette structure superimposes well with the structure of ATPaseCW in the homologous protein MORC3[15] (RMSD of 1.5 Å) and shares high sequence similarity with MORC3 (Supplementary Fig. S6.3), the mechanisms of action of the two proteins differ. The ATPaseCW cassette of MORC4 is capable of binding to DNA and therefore is active, whereas the ATPaseCW cassette of MORC3 represents an autoinhibited state, which does not inherently bind DNA. Binding of the CW domain of MORC3 to histone H3K4me3, which leads to the disruption of the CW:ATPase complex [328] (the same surface of CW is involved in binding to the ATPase domain and H3K4me3 [327,328,329]), is required for the release of autoinhibition, thus allowing interaction with DNA and resulting in MORC3 activation [327,329]. Why is the

catalytic activity of MORC4 independent of histone binding in contrast to the MORC3 catalytic activity and what does then stimulate MORC4 activation? To explain this, we considered three possibilities, i.e., MORC4 does not bind histone tails, DNA binding occludes the histone-binding site, or DNA binding promotes the catalytic activity of MORC4 independently of histone binding.

6.3.3 The CW domain of MORC4 selects for histone H3K4me3

A set of zinc fingers, including the CW domain of MORC3[13], has been shown to recognize histone sequences, therefore we tested whether the MORC4 CW domain either isolated (in this section) or linked to the ATPase domain in the ATPaseCW cassette (below) is capable of binding to histone tails using high-throughput histone peptide microarrays, pull-down assays and NMR (Fig. 6.2). In microarrays, GST-tagged CW was incubated with a library of over ~250 synthetic histone peptides containing single posttranslational modifications (PTMs) or combinations of PTMs, including acetylated and methylated lysine residues, phosphorylated serine and threonine residues and methylated arginine residues known to be present in the core and variant histone proteins (Supplementary Data 1). We found that the MORC4 CW domain binds to the N-terminal histone H3 peptides, selecting for methylated H3K4 sequences and does not recognize histone H4, H2A, and H2B (Fig. 6.2a, b). These results were corroborated by pull-down assays, in which GST-tagged CW domain of MORC4 was incubated with biotinylated unmodified and modified histone H3 peptides. As shown in Fig. 6.2c, CW associated with methylated H3K4 peptides, preferring the higher (tri)methylation state, H3K4me3. While methylation of the neighboring sites H3R2 and H3R8 appeared to not affect binding of CW to H3K4me3, phosphorylation of H3T3 or H3T6 disrupted interaction with the doubly modified H3T3phK4me3 and H3K4me3T6ph peptides.

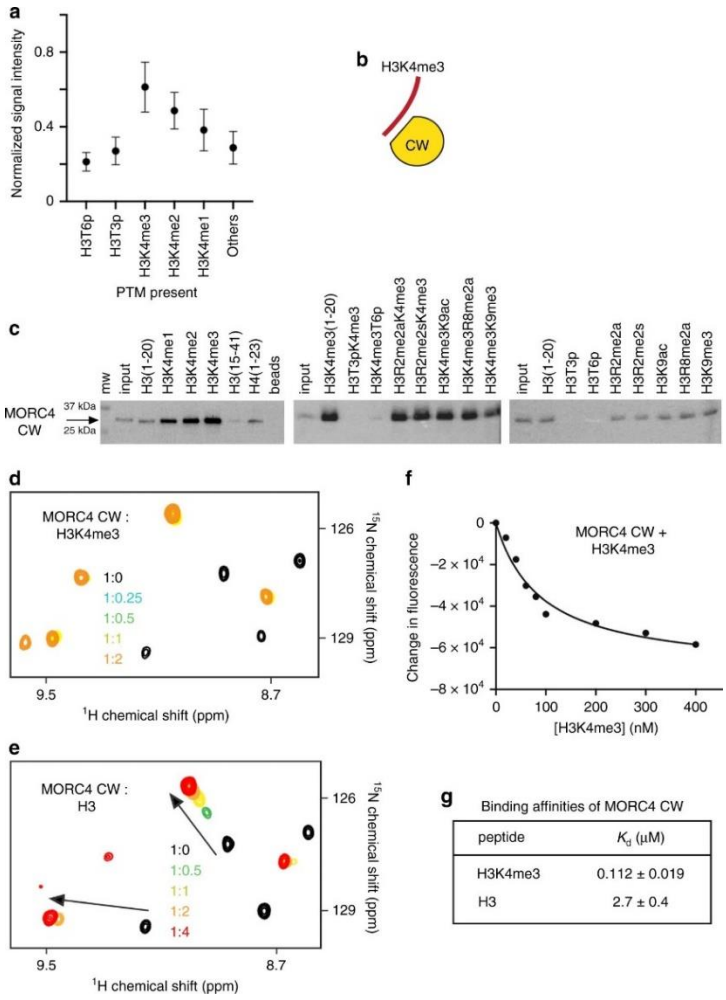


Figure 6.2: The CW domain of MORC4 recognizes H3K4me3. (a) Normalized average microarray signal intensities detected upon binding of the GST-tagged MORC4 CW domain to the peptides carrying indicated PTMs. Data are represented as mean values +/− S.D. from $n \geq 3$ independent arrays. See also supplementary data. (b) Cartoon representation of CW in complex with histone H3K4me3 tail. (c) Peptide pull-down assays of the GST-tagged MORC4 CW domain using the indicated histone H3 peptides. Images are representative of three independent experiments with similar results (n = 3). Source data are provided in a Source Data file. (d, e) Superimposed $^{1}H,^{15}N$ HSQC spectra of the ^{15}N-labeled WT MORC4 CW domain collected upon titration with the histone H3K4me3 (d) and unmodified H3 (e) peptides. Spectra are color coded according to the protein-to-peptide molar ratio. (f) Representative binding curve used to determine K_d values by tryptophan fluorescence. (g) Binding affinities of MORC4 CW for the indicated histone H3 peptides measured by tryptophan fluorescence. Data are represented as mean values +/− S.D. from three independent experiments (n = 3). Source data are provided in a Source Data file.

To further compare binding of the MORC4 CW domain to methylated and unmodified H3K4, we produced ^{15}N-labeled CW and examined it in $^{1}H,^{15}N$ heteronuclear single quantum coherence

(HSQC) experiments (Fig. 6.2d, e). Titration of the H3K4me3 peptide (aa 1–12 of H3) resulted in substantial chemical shift perturbations (CSPs) in the MORC4 CW domain, indicating direct interaction. Throughout the gradual addition of the peptide, a number of crosspeaks corresponding to the free state of the protein disappeared, and simultaneously, another set of resonances, corresponding to the bound state, appeared. This pattern of CSPs indicates a tight binding in the slow exchange regime on the NMR time scale. However, interaction of the MORC4 CW domain with the unmodified H3 peptide was weaker judging by CSPs in the intermediate exchange regime. In support, dissociation constants (K_d's) for the complexes of the MORC4 CW domain with the H3K4me3 and H3K4 peptides were found to be ~0.1 and 2.7 μM, respectively, as measured by tryptophan fluorescence (Fig. 6.2f, g and Supplementary Fig. S6.4). These results demonstrate that the MORC4 CW domain is not only a selective reader of H3K4me3 but also one of the strongest effectors of H3K4me3, which exhibits ~7–10 fold higher binding affinity to this PTM than other readers [335,336].

Furthermore, consecutive NMR titration experiments showed that the two ligands of the MORC4 CW domain—the ATPase domain and histone H3—compete for binding to CW. Gradual addition of the isolated unlabeled ATPase domain to the ^{15}N-labeled MORC4 CW domain led to a decrease in intensities of amide resonances, which suggested formation of the large ATPase:CW complex (Fig. 6.3a). Titration of the ATPase domain into the CW:H3 complex caused CSPs that were indicative of the disruption of the CW:H3 complex and formation of the complex between CW and the ATPase domain (Fig. 6.3b). We concluded that MORC4 CW is capable of binding to either H3 or the ATPase domain and that, similarly to MORC3 [327-329], the ATPase-binding site and the histone-binding site of CW overlap.

6.3.4 The MORC4 CW domain binds DNA

Considering the ability of the MORC4 ATPaseCW cassette to be activated by DNA, we investigated whether the CW domain contributes to the DNA binding by testing it in electrophoretic mobility shift assay (EMSA) (Fig. 6.3c). 601 DNA was incubated with increasing concentrations of CW,

and the reaction mixtures were resolved on a 5% native polyacrylamide gel. A gradual increase in the amount of added CW resulted in the shift and disappearance of the free 601 DNA band and appearance of a band smear corresponding to the complex of CW with 601 DNA, indicating the direct interaction (Fig. 6.3c, e). Unlike the CW domain of MORC4, the CW domain of homologous MORC3 was largely incapable of binding to 601 DNA, as no DNA band shift was observed in EMSA with MORC3 CW (Fig. 6.3d).

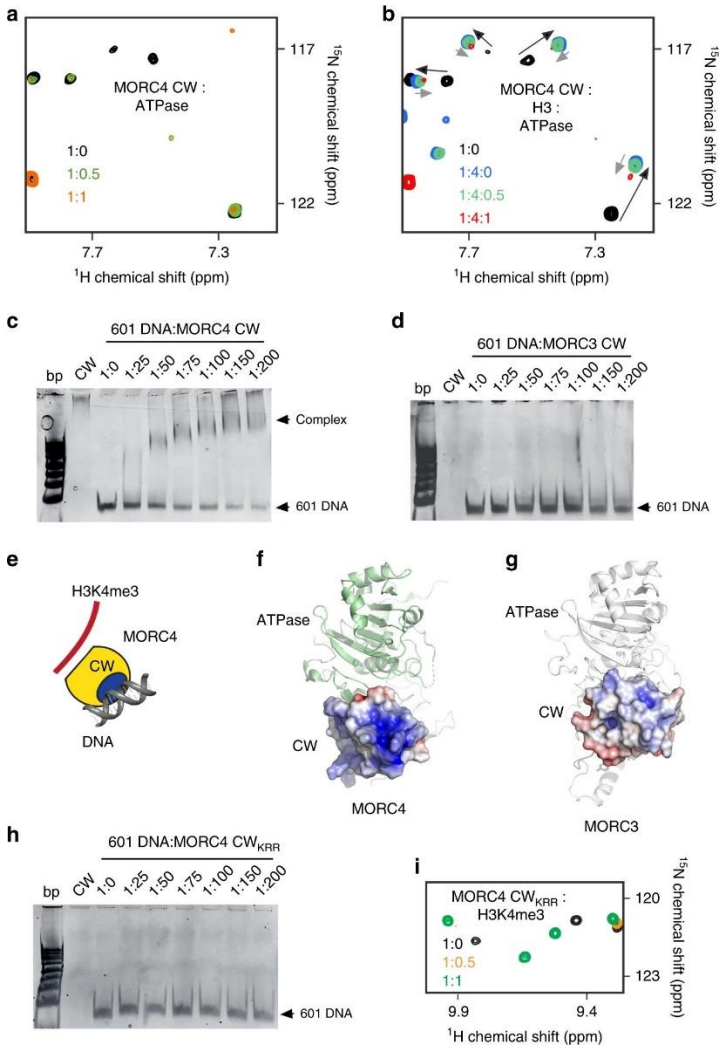


Figure 6.3: MORC4 CW binds to DNA. (a) Superimposed ^{1}H,^{15}N HSQC spectra of the ^{15}N-labeled MORC4 CW domain collected upon titration with the unlabeled ATPase domain of MORC4. (b) Superimposed ^{1}H,^{15}N HSQC spectra of the ^{15}N-labeled MORC4 CW domain (in complex with H3 peptide) collected upon titration with the unlabeled ATPase domain. Spectra in (a, b) are color coded according to the protein-to-ligands molar ratio. (c, d) EMSA with 601 DNA in the presence of increasing amounts of MORC4 CW (c) and MORC3 CW (d). (e) Cartoon representation of MORC4 CW in complex with histone H3K4me3 tail and DNA. (f, g) Electrostatic

surface potential of the CW domain of MORC4 (f) and MORC3 (g) (within the ATPase-CW cassette) was generated using APBS in Pymol with a range of –5/5 kT/e and colored blue and red for positive and negative charges, respectively. The ATPase domain is shown as ribbon and colored green and white in MORC4 and MORC3, respectively. The missing loops (residues K441 and R463) in the model of MORC4 CW were completed for this figure. (h) EMSA with 601 DNA in the presence of increasing amounts of mutant MORC4 CWKRR. (i) Superimposed ^{1}H,^{15}N HSQC spectra of the ^{15}N-labeled MORC4 CW_{KRR} domain collected upon titration with the histone H3K4me3 peptide. Spectra are color coded according to the protein-to-peptide molar ratio.

To identify the DNA-binding site of the MORC4 CW domain, we examined its electrostatic surface potential (ESP) and compared it to that of the MORC3 CW domain (Fig. 6.3f, g). Analysis of ESP of MORC4 CW revealed that particularly one side of CW, which is opposite to the H3K4me3- and ATPase-binding sites, is highly positively charged. Specifically, K460, R462 and R463 form a well-defined positively charged cluster on the protein surface, which suggested to us that these residues might be involved in binding to the negatively charged DNA. We generated the triple mutant of the MORC4 CW domain, K460A/R462A/R463A (CW_{KRR}), and tested its interaction with 601 DNA by EMSA. We did not observe the 601 DNA band shift even at the DNA to CW_{KRR} molar ratio of 1 to 200 (Fig. 6.3h), whereas H3K4me3 binding activity of this mutant appeared to be unaffected based on the pattern of CSPs seeing in NMR titration experiment with the H3K4me3 peptide (Fig. 6.3i). Collectively, these data suggest that the K460, R462, and R463 residues of the CW domain of MORC4 are required for the strong interaction with DNA and that the DNA-binding site and the histone/ATPase-binding sites are located on the opposed sides of the MORC4 CW domain.

6.3.5 DNA-binding activity of the MORC4 CW domain is essential

Can binding of the ATPaseCW cassette to DNA occlude the histone-binding site? We tested this possibility by measuring binding affinities of the ATPaseCW cassette for the H3 peptides in the absence and presence of 601 DNA by tryptophan fluorescence (Fig. 6.4a–c). We found that the ATPaseCW cassette binds to H3K4me3 only slightly (and not significantly) weaker as compared to binding of the isolated CW domain to the same H3K4me3 peptide (K_d's of 180 nM and 112 nM,

respectively) (Figs. 6.2f, g and 6.4a, c and Supplementary Fig. S6.4). However, interaction of the ATPaseCW:601 DNA complex with H3K4me3 peptide was ~4.5-fold stronger (K_d of 40 nM), implying that DNA binding enhances rather than impedes binding to H3K4me3 (Fig. 6.4d). These results led us to the current model for the MORC4 activation, in which DNA binding by the ATPaseCW cassette (both ATPase and CW domains are involved in this interaction) promotes the catalytic activity and binding of CW to H3K4me3 (Fig. 6.4d). This model was further corroborated by EMSA experiments (Fig. 6.4e–h). First, the presence of H3K4me3 peptide slightly increased binding of either wild-type ATPaseCW or mutated ATPaseCWKRR to 601 DNA, and second, the DNA-binding function of the mutated ATPaseCWKRR cassette was notably reduced, demonstrating the role of CW in the association with DNA. The importance of the DNA-binding activity of CW in activation of MORC4 was underscored by the fact that the mutant ATPaseCWKRR cassette failed to be stimulated to hydrolyze ATP in enzymatic assays (Fig. 6.4i, j)

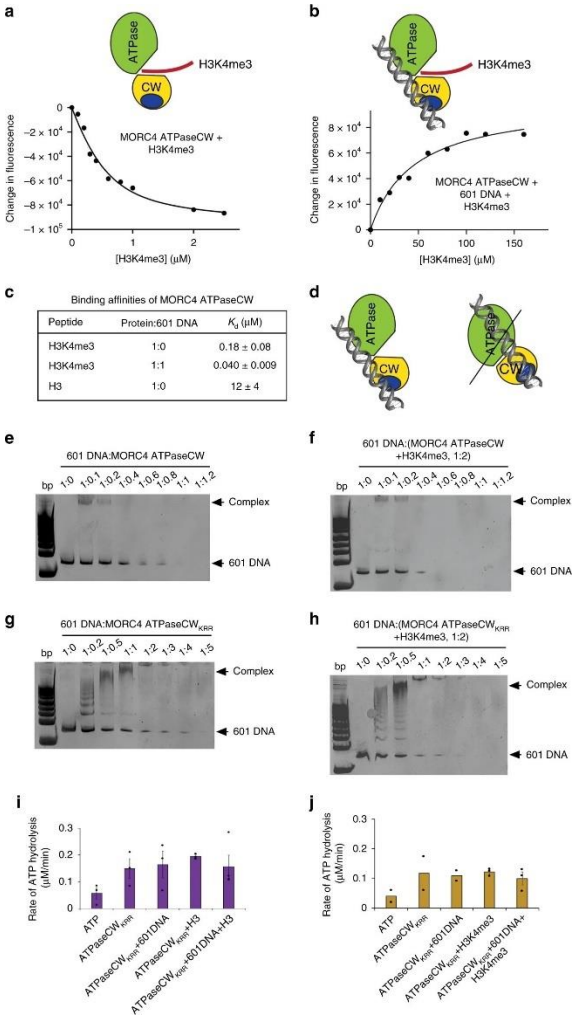


Figure 6.4: DNA binding by MORC4 CW is necessary for ATP hydrolysis. **a, b** Representative binding curves used to determine K_d values by fluorescence. Cartoon representations of MORC4 ATPaseCW binding to histone H3 tail in the absence (a) and presence (b) of DNA are shown above the curves. **c** Binding affinities of MORC4 ATPaseCW for the indicated histone H3 peptides (with or without 601 DNA) measured by tryptophan fluorescence. Data are represented as mean values +/− S.D. from three independent experiments (n = 3). Source data are provided in a Source Data file. **d** Models of the MORC4 ATPaseCW-DNA complex. **e–h** EMSA with 601 DNA in the presence of increasing amounts of WT (e, f) and mutant (g, h) MORC4 ATPaseCW in the absence (e, g) and presence (f, h) of histone H3K4me3 peptide. Experiments in (e) and (g) were repeated ten times and two times, respectively. **i, j** Rates of ATP hydrolysis by mutant MORC4 ATPaseCW$_{KRR}$ in the presence of histone H3 (i) or histone H3K4me3 peptide (j). Data are represented as mean values +/− S.D. of three independent experiments. Error was not calculated for the first three data sets containing two data points in (j) (n = 2). Source data are provided in a Source Data file.

6.3.6 Extensive DNA-binding site of the MORC4 ATPaseCW cassette

To delineate the DNA-binding site of the ATPaseCW cassette, we mutated four positively charged surface clusters in the ATPase domain, producing the K213Q/K214Q, R225Q/K227Q/K314Q, K352Q/R355Q, and K401Q/K403Q mutants of ATPaseCW, and tested binding of these mutants to 601 DNA by EMSA (Fig. 6.5). Among the mutants, the R225Q/K227Q/K314Q mutant (Fig. 6.5b), and to a lesser degree the K401Q/K403Q mutant (Fig. 6.5c), showed the most evident reduction in DNA-binding activity, whereas the reduction in DNA binding was less pronounced for other two mutants, K352Q/R355Q and K213Q/K214Q (Fig. 6.5d, e). The K460/R462/R463 cluster in the CW domain (Fig. 6.5a, blue oval) and the R225/K227/K314 and K401/K403 clusters in the ATPase domain (Fig. 6.5a, orange and light blue ovals) are separated by over 40 Å, indicating that a large surface of the ATPaseCW cassette is involved in binding to DNA. Furthermore, the R225/K227/K314 cluster is formed by R225 and K227 from one protomer and K314 from another protomer in the dimeric structure of ATPaseCW, suggesting that dimerization might be important for the association with DNA. Indeed, EMSA experiments showed that stimulation of ATPaseCW dimerization through adding AMPPNP increases binding to 601 DNA (Supplementary Figs. S6.5 and S6.6) and that at least ~30 bp DNA is needed for the ATPaseCW cassette to appreciably interact with DNA (Supplementary Fig. S6.7).

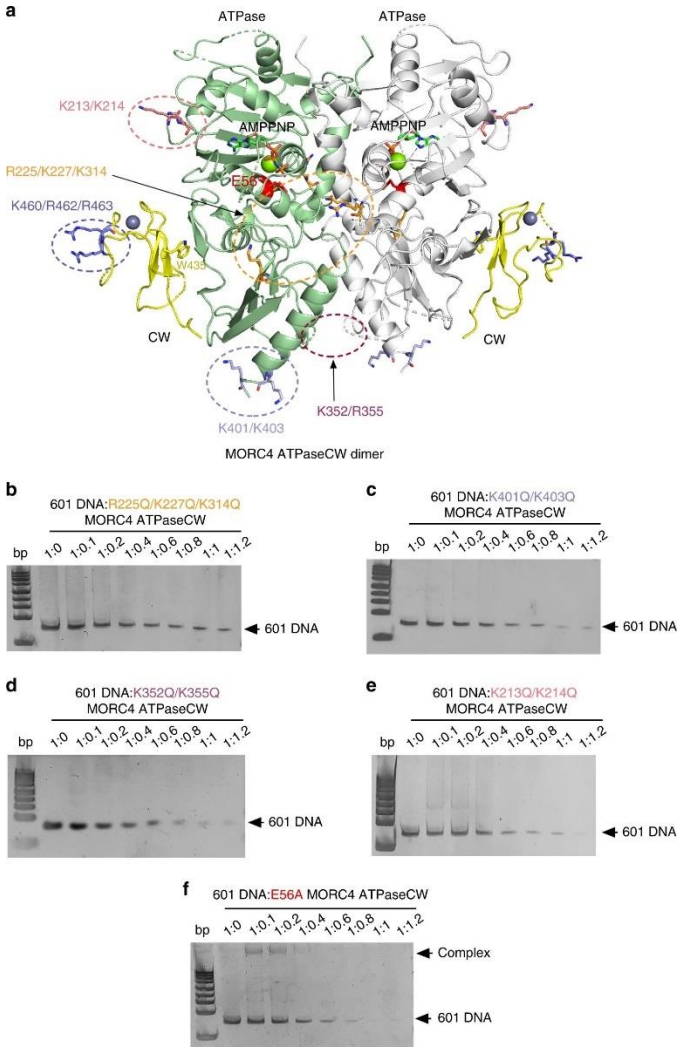


Figure 6.5: Mapping the DNA-binding site of the ATPaseCW cassette of MORC4. **a** A ribbon representation of the dimeric MORC4 ATPaseCW structure. The CW domain is colored yellow, and the ATPase domain is colored green in protomer A and gray in protomer B. The positively charged clusters in ATPaseCW are indicated by ovals and labeled. The residue E56, required for the catalytic activity of ATPaseCW, is also shown in red. **b–f** EMSA with 601 DNA in the presence of increasing amounts of indicated mutants of MORC4 ATPaseCW.

6.3.7 MORC4 impedes binding of DNA-associated proteins

To gain insight into the MORC4 activity at chromatin, we investigated the consequences of the association of the MORC4 ATPaseCW cassette with NCP, a fundamental unit of chromatin. Both the ATPase domain alone and the ATPaseCW cassette of MORC4 bound to NCP147 (the nucleosome reconstituted with the 147 bp 601 nucleosome positioning DNA sequence) in

fluorescence polarization assays ($S_{1/2}$ = 2.3 μM and 0.4 μM, respectively) and EMSA (Supplementary Figs. 6.8 and 6.9). However, interaction of the ATPase domain and the ATPaseCW cassette with NCP207 (the nucleosome reconstituted with 207 bp DNA containing the 601 sequence flanked with a 30 bp linker on either side and internally labeled with fluorescein 27 bp in from the 5' end) was increased ~5-fold and ~2-fold, respectively (Fig. 6.6a). These results suggest that extra-nucleosomal DNA enhances binding of MORC4. Importantly, binding of ATPaseCW to either NCP_{147} or NCP_{207} was ~2.5–6 fold tighter than the binding of the ATPase domain only, pointing to a contribution from the CW domain to the interaction with NCPs.

The impact of this interaction on the nucleosome dymnamics and unwrapping-wrapping equilibrium was investigated by Förster resonance energy transfer (FRET) [337]. We designed and prepared NCP_{252} using 252 bp DNA, which contained the 601 sequence with a 30 bp linker at the 5′ end, a 75 bp linker at the 3′ end, and the transcription factor Gal4 DNA-binding site replacing bases 8–26 in the 601 sequence (Fig. 6.6b). The Cy3 donor fluorophore was attached to the 5′ end of the 601 DNA adjacent to the Gal4-binding site, and the Cy5 acceptor fluorophore was attached to histone H2A(K119C). This placed Cy3 in the proximity of one of the Cy5 fluorophores, therefore a significant FRET signal is expected from a fully wrapped nucleosome, while a reduced FRET is expected when the NCP_{252} is in a more open, partially unwrapped state (Fig. 6.6c). Titration of the Gal4 DNA-binding domain into Cy3-Cy5 labeled NCP252 led to a decrease in FRET due to Gal4 binding to its target site and stabilization of the unwrapped state (Fig. 6.6d). Consistent with previous measurements [336], the value for the Gal4 concentration at which FRET efficiency is reduced by 50% ($S_{1/2}$) was found to be 20 nM.

To determine the effect of MORC4 on the nucleosome unwrapping-wrapping equilibrium, we next titrated Gal4 into NCP_{252} in the presence of ATPaseCW and ATP (Fig. 6.6e). The Gal4 ability to bind NCP_{252} was decreased ~5-fold, resulting in a $S_{1/2}$ = 100 nM (Fig. 6.6e). The presence of ATP was essential, as only a 2-fold decrease was observed in the absence of ATP ($S_{1/2}$ = 40 nM). Similar results were obtained in FRET experiments using Cy3-Cy5 labeled NCP_{147} containing the targeting site of another transcription factor, LexA, at bases 8–27 in the 601 DNA sequence (Fig.

6.5f). Addition of MORC4 ATPaseCW to the reaction resulted in a decrease in binding of LexA to NCP147 due to the stabilization of the wrapped state of NCP147. Collectively, these data reveal that binding of the MORC4 ATPaseCW cassette to the nucleosome shifts the unwrapping-wrapping equilibrium toward the wrapped state, enhancing the nucleosome stability and impeding binding of DNA-binding proteins, such as transcription factors and co-activators.

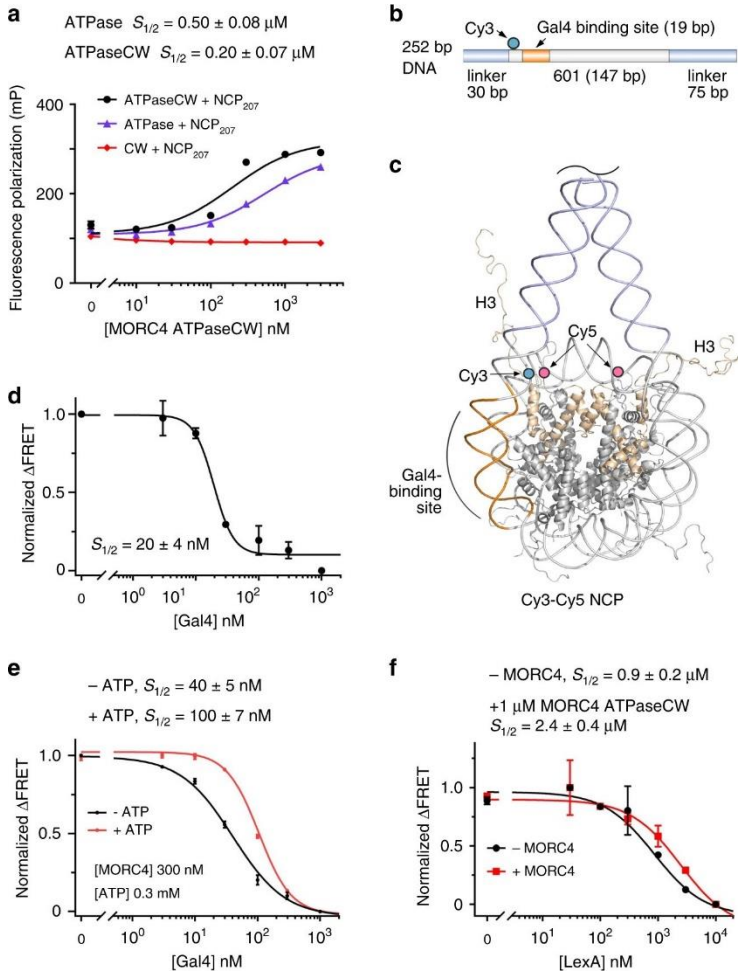


Figure 6.6: MORC4 ATPaseCW binds to and stabilizes nucleosomes. **a** Binding affinities and binding curves for the interactions of the indicated MORC4 regions with NCP_{207} as measured by fluorescence polarization. Data are represented as mean values +/– S.D. from three independent experiments (n = 3). **b** Schematic of the 252 bp DNA containing the 601 Widom sequence with a 75 bp linker at the 3′ end, a 30 bp linker at the 5′ end, and a Gal4 transcription factor binding site at bases 8–26 in the 601 Widom sequence with Cy3 positioned 34 bp from the 5′ end (cyan circle). **c** A model based on the crystal structure of the NCP with 197 bp palindromic 601 L DNA (PDB ID: 5NL0, the linker histone H1 is not shown) with histone H3 (wheat), Cy5 at H2AK119C (pink circles), Cy3 at the DNA (cyan circle), and the Gal4 target site (orange) are shown and labeled. Truncated DNA linkers are shown and colored light blue. **d, e** Normalized change in FRET

efficiency of the Cy3-Cy5 labeled NCP252 upon addition of Gal4 in the absence (d) and presence (e) of MORC4 ATPaseCW and +/−ATP. Data are represented as mean values +/− S.D. from three independent experiments (n = 3). **f** Normalized change in FRET efficiency of the Cy3-Cy5 labeled NCP147 upon the addition of LexA in the absence and presence of MORC4 ATPaseCW. Data are represented as mean values +/− S.D. from three independent experiments (n = 3).

6.3.8 ATPase activity of MORC4 regulates formation of NBs

Our findings demonstrate that the ATPaseCW cassette of MORC4 is capable of engaging chromatin. To determine the importance of this region in MORC4 functioning in vivo, we assessed the subcellular localization of MORC4 in HEK293T cells. Cells were transfected with mCherry-tagged wild-type and mutated full-length MORC4, and 48 h post transfection mCherry-tagged proteins were visualized in live cells by confocal fluorescence microscopy (Fig. 6.7a and Supplementary Fig. S6.10). We found that wild-type mCherry-MORC4 localizes to the nucleus, forming large NBs with the average size of 0.52 μm2. In 124 cells quantified, the number of MORC4 NBs ranges from 1 to 18 with an average of 4.93 per cell (Fig. 6.7b). The E56A mutant of MORC4 that retains its DNA-binding activity (Fig. 6.5f) but has catalytically impaired ATPase domain (Supplementary Fig. S6.11) showed a notable increase in the number of NBs and a decrease in the size of NBs in 127 cells quantified, indicating that the catalytic activity of MORC4 regulates the formation of NBs (Fig. 6.7a). In contrast, MORC4 W435A mutant, in which the structure of the CW domain is disrupted (Supplementary Fig. S6.12), was dispersed throughout the cell, revealing the critical role of CW in nuclear localization of MORC4 and its association with chromatin in vivo (Fig. 6.7a).

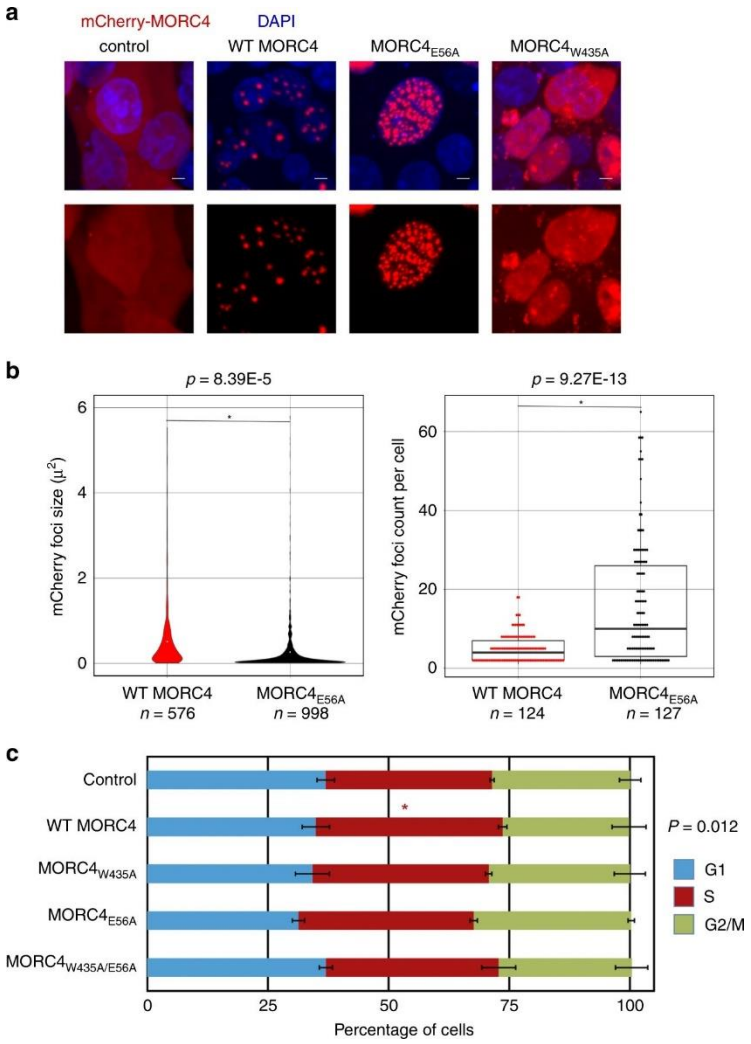


Figure 6.7: MORC4 forms NBs and affects the S phase cell cycle. **a** Representative confocal microscopy images of 293T-HEK cells overexpressing mCherry-CTRL, mCherry-MORC4 WT, mCherry-MORC4 E56A or mCherry-MORC4 W435A. Transfection were performed a minimum of three times, scale bars represent 5 μm. **b** MORC4 E56A shows a significant decrease in mCherry foci size ($p = 8.39E^{-5}$) and a significant increase in number of mCherry foci per cell ($p = 9.27E^{-13}$) compared to MORC4 WT. The analysis shown is from at least 10 independent images (WT MORC4 n = 13, MORC4 E56A n = 15), with a minimum of 120 nuclei analyzed for each condition (WT MORC4 n = 124, MORC4 E56A n = 127). **c** Cell cycle analysis following 48-h overexpression of MORC4 proteins exhibit increased percent of cells in S phase for WT MORC4 ($p = 0.012$) compared to control cells. Data represent the average of three independent experiments. Error bars represent S.E.M., * indicates significant difference from mCherry-CTRL ($p < 0.05$) by two-tailed student t-test.

6.3.9 MORC4 affects the S phase cell cycle

Chromatin organization is dynamic and undergoes extensive changes during the cell cycle, which requires chromatin condensation and decondensation. Because MORC4 enhances the nucleosome stability and homologous MORC1 was found to mediate chromatin compaction [339], we tested whether MORC4 has a role in regulating the cell cycle. We assayed the cell cycle state distribution of ~10,000 HEK293T cells expressing mCherry-tagged wild-type and mutant MORC4 by flow cytometry (Fig. 6.7c and Supplementary Fig. S6.13). We found that cells expressing wild-type MORC4 showed a modest, but significantly larger population in the S phase of the cell cycle (38.7%) compared with the control cells (34.5%). These results indicate that overexpression of MORC4 alters S phase progression in HEK293T cells. Notably, neither the ATPase mutant (E56A, 36.3%) nor the CW mutant (W435A, 36.5%) or the dual mutant (E56A/W435A, 35.8%) altered S phase progression to the same degree. These data reveal that both ATPase activity and the intact CW domain are essential (Fig. 6.7c), but no significant effect was observed for the K460A/R462A/R463A mutant, which suggested that the disruption of the DNA-binding activity of CW is not enough (Supplementary Fig. S6.13). Our results are consistent with a model whereby MORC4 stabilizes the nucleosome resulting in decreased chromatin accessibility during S phase and slowing replication of DNA. Together, our data point to a role of MORC4 in S-phase progression, which depends on both its catalytic activity and the CW domain.

6.4 Discussion

Among four members of the MORC family of proteins, MORC4 is the least well characterized. In this study we report the molecular and structural mechanism underlying the catalytic ATPase activity of MORC4. Our current model for the MORC4 activation suggests that binding of the ATPase domain and the CW domain of the ATPaseCW cassette to DNA stimulates the catalytic activity of MORC4 and promotes binding of CW to H3K4me3 (Fig. 6.4b). Notably, our data demonstrate that the DNA-binding function of CW is required for the MORC4 activation, as the

mutant protein with the impaired DNA-binding site of CW loses its ability to be stimulated to hydrolyze ATP.

Despite the overall similar architecture of the MORC proteins, their mechanisms for enzymatic activation differ markedly. In contrast to MORC4, the CW domain of MORC3 does not bind DNA and its coupling to the ATPase domain prevents DNA binding by the ATPase domain [327,329]. This results in an autoinhibited, catalytically inactive state of MORC3, which is released through the interaction of CW with histone H3K4me3. In contrast to MORC4 and MORC3, it has been shown that following DNA damage, MORC2 is phosphorylated on Ser739, and this phosphorylation is required for the ATPase activity of MORC2[26]. The distinctive mechanisms of action of MORCs correlate with the fact that these proteins have unique DNA-binding regions: (i) encompassing the ATPase and CW domains in MORC4, (ii) the ATPase domain only, with the CW domain impeding binding to DNA in MORC3 [327,329], and (iii) a coiled coil (CC) insertion between the ATPase domain and the CW domain in MORC2 [331,341]. While both DNA binding and dimerization are essential for the ATPase activity of MORC3, mutation of the MORC2 residues that disrupt dimerization interface raises the rate of ATP hydrolysis, and this observation led to a model of MORC2 cycling between monomeric and dimeric states [331].

Furthermore, MORCs are characterized by different chromatin engaging modes, as their CW domains have dissimilar functions. MORC2 contains a CW domain that is lacking one of the aromatic cage tryptophan residues and thus does not recognize histones. The CW domain of MORC3 is incapable of binding to DNA, and its histone-binding function mediates both the chromatin association and ATPase activity of MORC3. Unlike MORC3 CW, the CW domain of MORC4 is essential in chromatin targeting and binds to histones but it is its DNA-binding function that is necessary for the ATPase activity of MORC4. The CW domain does not inhibit the DNA-binding activity of MORC4 and MORC2, whereas the CW domain of MORC3 does.

Given the stark differences in the mechanisms of action of MORCs, it is expected that each MORC family member may be involved in distinct biological events. Here, we show that in cells, MORC4 plays a role in S-phase progression, possibly by decreasing chromatin accessibility during S phase and slowing replication of DNA, as in vitro results demonstrate that MORC4 promotes

the DNA wrapping in the nucleosome in FRET experiments. While MORC2 is found in the nuclear and cytosolic pools, MORC4 and MORC3 localize exclusively to the nucleus where they form NBs. We have previously shown that MORC3 NBs have LLPS gel-like properties [334], and it will be essential in future studies to determine the nature of MORC4 NBs. Furthermore, to better understand their roles in chromatin condensation or decondensation, it will be imperative to elucidate the structural modes of MORCs' engagement with the nucleosome.

6.5 Methods

Protein expression and purification

The ATPase domain (aa 29–411) and the ATPaseCW cassette (aa 29–486) of human MORC4 (Q8TE76) were cloned into a pDEST17 vector and pDEST15 vector, respectively. The CW domain (aa 417–475) of MORC4 was cloned into a pGEX 6p-1 vector. Proteins were expressed in Rosetta2 (DE3) pLysS or BL21 (DE3) RIL in LB or minimal media supplemented with 15NH4Cl. Protein expression was induced with 0.2 mM IPTG for 16 h at 16 °C. The His-tagged ATPase domain was purified on HisPur Ni-NTA agarose beads (Thermo-Fisher) in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM MgCl2, and 2 mM DTT. The GST-tagged ATPaseCW cassette was purified on Pierce glutathione agarose beads (Thermo-Fisher) in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM MgCl2, and 2 mM DTT. GST-CW was purified on Pierce glutathione agarose beads (Thermo-Fisher) in 20 mM Tris-HCl (pH 7.5) buffer supplemented with 150 mM NaCl and 2 mM DTT. The GST tag was cleaved overnight at 4 °C with either TEV or PreScission proteases. Unlabeled proteins were further purified by size exclusion chromatography and concentrated in Millipore concentrators. All mutants were generated by site-directed mutagenesis using the Stratagene QuikChange mutagenesis protocol, grown and purified as wild-type proteins. MORC3 CW was expressed and purified as reported [329].

X-ray crystallography

The human MORC4 ATPaseCW cassette (aa 29–486) solution was concentrated to 6 mg/mL and incubated with the H3K4me3 peptide (aa 1–12 of H3, in a 1:2 protein-to-peptide molar ratio) along with 2 mM AMPPNP at 25 °C for 20 min in buffer containing 20 mM Tris-HCl (pH 7.5), 5 mM MgCl2 and 2 mM DTT. Crystals were obtained at 18 °C using the sitting-drop vapor diffusion method by mixing equal volumes of protein solution with well solution composed of 0.1 M Hepes (pH 7.5) and 20% Jeffamine ED-2003. X-ray diffraction data were collected from a single crystal at the ALS 4.2.2 beamline, Berkeley administrated by the Molecular Biology Consortium. HKL2000 was used for indexing, scaling, and data reduction. The phase solution was obtained by molecular replacement using the MORC3 ATPaseCW structure (PDB ID 6O1E) as a search model. Manual model building was performed using Coot [342], and the structure was refined using Phenix [343]. The final structure was verified by MOLProbity [344] and the PDB validation server. The X-ray diffraction and structure refinement statistics are summarized in Supplementary Table 1.

Fluorescence spectroscopy

Spectra were recorded at 25 °C on a Fluoromax-3 spectrofluorometer (HORIBA). The samples containing 0.5–1 μM wild-type or mutated ATPaseCW or CW in 20 mM Tris pH 6.8, 100 mM NaCl, 2 mM DTT buffer and progressively increasing concentrations of the histone peptides were excited at 295 nm. Emission spectra were recorded between 320 and 360 nm with a 0.5 nm step size and a 0.5 s integration time and averaged over three scans. The K_d values were determined using a nonlinear least-squares analysis and the equation:

$$\Delta I = \Delta I_{max} \frac{(([L] + [P] + K_d) - \sqrt{([L] + P] + K_d)^2 - 4[P][L]}))}{2[P]}$$

where [L] is the concentration of the histone peptide, [P] is the concentration of the protein, ΔI is the observed change of signal intensity, and ΔI_{max} is the difference in signal intensity of the free and bound states of the protein, or $\Delta I = \Delta I_{max}$[L]/(K_D +[L]). The K_D values were averaged over two or three separate experiments, with error calculated as the SD between the runs.

ATPase activity assays

The ATPase assays were performed using the EnzChek Phosphate Assay Kit (Invitrogen, item # E6646). The reactions were carried out on 1.0 μM of MORC4 His-ATPase or ATPaseCW (WT and mutants), in the presence and absence of 1 μM 601 DNA or 50 μM of unmodified H3 or H3K4me3 peptide (aa 1–12 of H3) in a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM MgCl2, 0.1 mM sodium azide, 200 μM MESG, and 1 U of PNP. In assay with the nucleosome, 0.5 μM of MORC4 ATPaseCW and 0.5 μM NCP167 (Activemotif) were used. The reaction was started by adding 2 mM ATP to the mixture at room temperature, and the release of inorganic phosphate was monitored by measuring the absorbance at 360 nm on a Nanodrop 2000c spectrophotometer (Thermo Scientific). In the presence of inorganic phosphate, produced by the hydrolysis of ATP to ADP, MESG is enzymatically converted to ribose 1-phosphate and MESG by PNP, resulting in a shift in the wavelength absorbance from 330 nm for MESG to 360 nm for the product. The rate of ATP self-hydrolysis was measured in parallel. Error was calculated as the S.D. of at least three separate experiments (two in Fig. 6.4j).

NMR experiments

NMR experiments were carried out at 298 K on a Varian INOVA 600 and 900 MHz spectrometers. His-tagged MORC4 ATPase (1–411) was expressed and purified as described above and used without cleavage of the tag. NMR samples contained 0.1 mM uniformly 15N-labeled WT or mutated CW in 20 mM Tris-HCl (pH 7) buffer supplemented with 100 mM NaCl, 5 mM DTT, and 10% D_2O. Binding was characterized by monitoring CSPs in the proteins induced by H3 peptides

or the ATPase domain. For the CWKRR mutant, NMR experiments were performed in 20 mM Tris-HCl (pH 6.8) buffer supplemented with 150 mM NaCl, 2 mM DTT, and 10% D2O.

Peptide microarrays

Histone peptide microarrays were produced, performed, and analyzed as described [31] with some modifications. Microarrays were analyzed by the Typhoon Trio+ (GE) and quantified using the ImageQuant TL software. The analysis was performed by first averaging the triplicate values for each peptide in a subarray; these were then linearly scaled based on the minimum and maximum values to be between 0 and 1. These scaled values were then averaged together and sorted based on the modifications stated in the plot legends. Averages and standard deviations are from ≥3 arrays; these values for every peptide tested can be found in Supplementary Data 1.

Histone peptide pulldown assays

The in-solution histone peptide pulldowns experiments were performed as described [31]. In brief, 50 pmol of MORC4 CW was incubated with 500 pmol of each biotinylated histone peptide in 50 mM Tris, 0.1% NP-40, 0.5% BSA, 500 mM NaCl, pH 8.0 buffer for one hour at 4 °C with rotation. Five microliters of magnetic streptavidin-coated beads were added to each reaction and the mixtures were incubated for another hour with rotation. The beads were then washed three times and analyzed using SDS–PAGE and western blotting. Anti-GST (13-0022, 1:1000) antibody was from EpiCypher and anti-Rabbit-HRP (NA934V, 1:20,000) antibody was from GE. The input control lane represents 1% of the input material used in a pulldown. The images shown are representative of three experiments.

EMSA with DNA

EMSA experiments were performed essentially as described [346]. In brief, increasing amounts of WT or mutant CW or ATPaseCW cassette of MORC4 were incubated with 147 bp 601 DNA (5 pmol) or 50 ng of O'RangeRuler 5 bp ladder (Thermo Scientific) in a DNA-binding buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2.5 mM MgCl2, and 2 mM DTT for 0.5 h on

ice. The reaction mixtures were loaded on 5% native polyacrylamide gels, and electrophoresis was performed in 0.2× Tris-borate-EDTA (TBE) buffer at 100 V for 1.2 h on ice. Gels were stained with SYBR Gold (Invitrogen).

Cross-linking assay

His-tagged MORC4 (aa 1–486) and MORC4 (29–283, I30A) were expressed and purified as described above. 10 μM proteins were cross-linked in 10 mM Hepes pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0.5 mM TCEP, and 1 mM AMPPNP or ADP. BS3 cross-linker (Thermo) was resuspended to 2 mM and used at a concentration of 0.1 mM. Cross-linking was allowed to proceed for 10 min. Cross-linked samples were quenched in 0.1 M Tris pH 7.5, resolved on SDS/PAGE, and stained with Coomassie blue.

Fluorescence polarization

Fluorescence polarization measurements were acquired with a Tecan infinite M1000Pro plate reader by exciting at 470 nm and measuring polarized emission at 519 nm with 5 nm excitation and emission bandwidths. The fluorescence polarization was calculated from the emission polarized parallel and perpendicular to the polarized excitation light. Fluorescence polarization measurements were carried out using increasing concentrations of MORC4 ATPaseCW, ATPase alone, or CW alone with 5 nM 207 bp Fluorescein-labeled NCPs in 15 mM Tris pH 7.5, 75 mM NaCl, 0.00625% Tween 20, and 5 mM dithiothreitol buffer in 30 μL reaction volumes. The samples were loaded into a Corning round bottom black polystyrene plate and incubated at room temperature for 10 min before fluorescence polarization measurements of the samples were taken. The data were then fit to a non-cooperative binding isotherm to determine $S_{1/2}$ values. Errors represent a S.D. between the $S_{1/2}$ values based on three runs.

NCP preparation for fluorescence polarization and FRET

Histone octamers unlabeled or labeled with Cy5 at H2AK119 were prepared as previously described[33]. NCPs were reconstituted by combining octamer with 1.25× excess DNA and performing slow salt dialysis. Free DNA was then separated from fully reconstituted NCPs by sucrose gradient purification. NCPs for fluorescence polarization assays were reconstituted from unlabeled octamer and 207 bp DNA containing the 601 Widom sequence flanked with a 30 bp linker on either side and internally labeled with Fluorescein 27 bp in from the 5′ end. Cy3-Cy5 labeled NCPs for FRET assays were reconstituted with Cy5 labeled octamer plus either 147 bp 601 Widom DNA containing a LexA transcription factor binding site at bases 8–27 [34], or 252 bp DNA containing the 601 Widom sequence with a 75 bp linker at the 3′ end, a 30 bp linker at the 5’ end, and a Gal4 transcription factor binding site at bases 8–26 in the 601 Widom sequence. The 147 bp LexA-binding DNA was labeled with Cy3 on the 5′ end, and the 252 bp Gal4-binding DNA was internally labeled with Cy3 34 bp in from the 5′ end.

LexA and Gal4 protein preparation

LexA was expressed in E. coli, separated from genomic DNA and the proteome by polyethyleneimine precipitation, and salted out with ammonium sulfate. LexA was resuspended in buffer A (20 mM potassium phosphate pH 7, 0.1 mM EDTA, 10% glycerol, 1 mM DTT) + 200 mM NaCl and purified by a linear gradient to A + 800 mM NaCl over either a cellulose phosphate or HiTrap Heparin HP column. Final LexA purification was performed on a hydroxyapatite column and dialyzed into 10 mM PIPES pH 7.0, 0.1 mM EDTA, 10% glycerol, and 200 mM NaCl for storage at −80 °C. The His-tagged DNA-binding domain (aa 1–147) of Gal4 was expressed in E. coli and purified from lysate by a linear gradient of buffer B (25 mM Tris pH 7.5, 200 mM NaCl, 0.2% Tween 20, 20 μM zinc acetate, 1 mM dithiothreitol, and 1 mM PMSF) + 10 mM imidazole to B + 200 mM imidazole over a Nickel-NTA column. Gal4 (1–147) was then further purified by a linear gradient of buffer C (25 mM Tris pH 7.5, 20 μM zinc acetate, 1 mM

dithiothreitol, and 1 mM PMSF) + 200 mM NaCl to C + 800 mM NaCl over a Tosoh cation exchange column. Purified Gal4 (1–147) was then dialyzed into 10 mM HEPES pH 7.5, 200 mM NaCl, 10% glycerol, 20 μM zinc acetate, 1 mM DTT, and 1 mM PMSF for storage at −80 °C.

FRET

Cy3-Cy5 NCP FRET efficiency measurements were carried out on a Horiba Scientific Fluoromax 4. Samples were excited at 510 and 610 nm and the photoluminescence spectra were collected from 530 to 750 nm and 630 to 750 nm for donor and acceptor excitations, respectively. Each wavelength was integrated for one second, and the excitation and emission slit width were set to 5 nm with 2 nm emission wavelength steps. FRET measurements were computed through the (Ratio)A method. LexA titrations were carried out in 20 mM Tris pH 7.5, 75 mM NaCl, 10% glycerol, 0.00625% Tween 20 buffer with 5 nM 147 bp Cy3-Cy5 NCPs mixed in 20 μL reaction volumes and allowed to incubate for 5 mins at 4 °C before FRET of the sample was measured. To measure changes in accessibility of the NCP_{147} due to MORC4 ATPaseCW binding, LexA titrations were repeated in the presence of NCP binding saturating concentrations of MORC4 ATPaseCW. Gal4 (1–147) titrations were carried out in 10 mM Tris pH 8, 130 mM NaCl, 10% glycerol, 0.0075% Tween 20 buffer with 1 nM 252 bp Cy3-Cy5 nucleosomes mixed in 60 μL reaction volumes and allowed to incubate for 10 mins at room temperature before FRET of the sample was measured. To measure changes in accessibility of the NCP_{252} due to MORC4 ATPaseCW binding, Gal4 (1–147) titrations were repeated in the presence of NCP binding saturating concentrations of MORC4 ATPaseCW +/− 0.3 mM ATP. FRET efficiency values were normalized against the FRET efficiency value of NCP in the absence of titrant. The titrations were then fit to a non-cooperative binding isotherm to determine $S_{1/2}$ values. Errors represent a S.D. between the $S_{1/2}$ values based on three runs.

EMSA with NCPs

EMSAs were performed by mixing increasing amounts of MORC4 ATPase with 5 nM 147 bp fluorescein-labeled NCP in 25 mM Tris-HCl pH 7.5, 75 mM NaCl, 25 mM Tris-HCl pH 7.5, 0.005% Tween 20, and 10% glycerol buffer in a 12 µL reaction volume. Each sample was incubated at 4 °C for 5 min and then loaded onto a 5% native polyacrylamide gel. Electrophoresis was performed in 0.3× Tris-borate-EDTA (TBE) at 300 V for 90 min. Fluorescein fluorescence images were acquired with a Typhoon Phosphor Imager.

Cell culture

293T-HEK cells were maintained in Dulbecco's Modified Essential Medium (DMEM) with 10% FB Essence (VWR), 1% penicillin/streptomycin and L-glutamine.

MORC4 localization in HEK cells

WT mCherry-MORC4, mCherry-MORC4 E56A, mCherry-MORC4 W435A, mCherry-MORC4 E56A/W435A, mCherry-MORC4 K460A/R462A/R463A, or mCherry control plasmid were transfected into 2 million 293T-HEK cells seeded on 25 mm circular coverslips (Carolina Biological Supply Item#: 633057) in 10-cm-diameter tissue culture dish by Lipofectamine 3000 (Life Technology, L3000-075) following the manufacturer's instructions. Two days after transfection, cells were stained with Hoechst 333241 for 30 min and imaged live on a ZeissAxiovert 200 M inverted microscope fitted with a ×40 oil objective. Images were acquired with Slidebook 6 software and exported for analysis in Image J.

mCherry foci size and foci per cell calculations

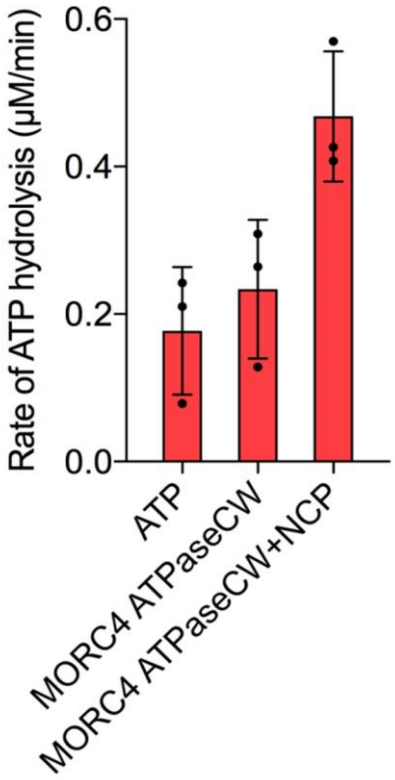
mCherry foci sizes were calculated using the ImageJ image processing program (version 1.51) using the Analyze Particle tool. mCherry particle sizes were collected for MORC4 WT and MORC4 W435A transfected cells. Data were plotted with R (version 3.4.1) using ggplot2 (version 3.2.1). Cell size and mCherry foci count per cell were calculated using CellProfile cell image analysis

software (version 3.1.5) to identify nuclei and mCherry foci, relate mCherry foci to nuclei containing them and measure the size of nuclei and number mCherry foci contained. Data were plotted with R (version 3.4.1) using ggplot2 (version 3.2.1).

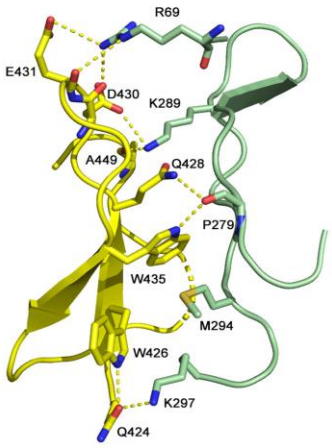
Flow cytometry

For DNA content analysis, 106 cells previously transfected with mCherry MORC4 constructs and labeled with 25 μM Ethynyl deoxyUridine (EdU) for 1 h in 100 μL were washed with PBS (NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, KH2PO4 1.8 mM) and fixed in ice cold 70% ethanol overnight. Fixed cells were washed and resuspended in PBS containing 0.05% NP40 and incubated at room temperature for 30 min. Cells were again washed with PBS. Cells were resuspended in Alexa Fluor 647 click staining solution (100 mM Tris pH 7.6, 4 mM CuSO4, 3 μM AF647 azide, 100 mM Sodium Ascorbate) for 1 h. Following staining, cells were washed with PBS and then stained for total DNA content using 10 μg/mL propidium iodide (PI) and 0.2 mg/ml RNase A for 60 min before analysis on a FACScan instrument. Data were analyzed using FlowJo (Version 10) with cells gated to remove doublets. EdU vs PI plots were gated in quadrants to identify G1, early S, late S and G2/M cells. S phase fractions were summed for total S phase population. Four biological replicates were conducted and the average percentages for each phase were plotted.

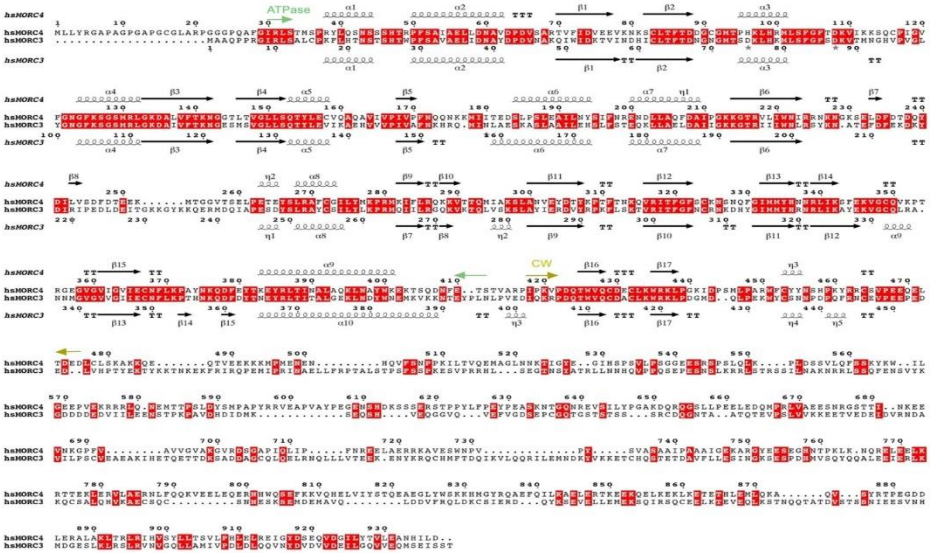
6.6 Supplementary Information



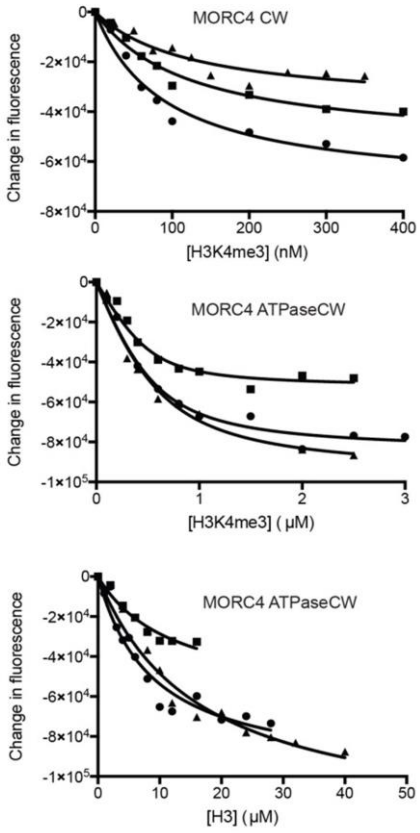
Supplementary Figure S6.1: Rates of ATP hydrolysis by the ATPase-CW cassette of MORC4 in the presence and absence of NCP167 (the nucleosome containing 167 bp DNA). Data are represented as mean values +/- S.D. from three independent experiments (n=3). Source data are provided in a Source Data file.



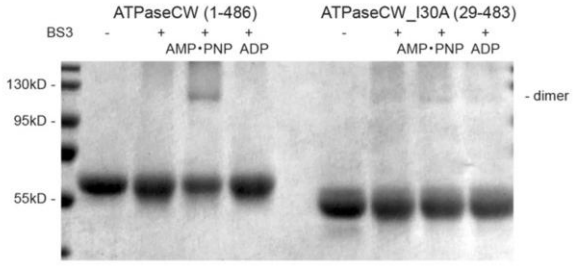
Supplementary Figure S6.2: A zoom-in view of the ATPase:CW interface from the structure of the ATPaseCW cassette. Dashed lines indicate hydrogen bonds between the ATPase domain residues (green) and the CW domain residues (yellow).



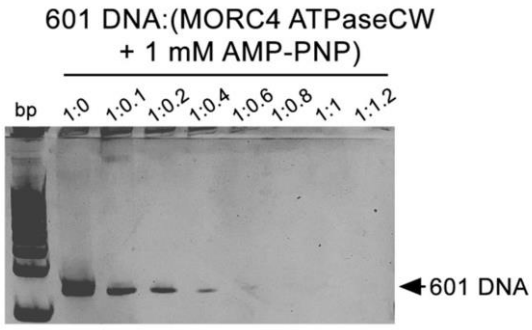
Supplementary Figure S6.3: Alignment of the amino acid sequences from MORC4 and MORC3. Identical residues are highlighted red. The ATPase and CW domains are indicated by green and yellow arrows and labeled.



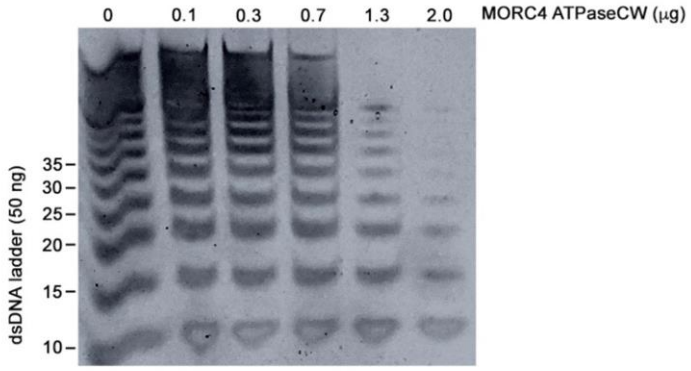
Supplementary Figure S6.4: Binding curves used to determine the K_d values by tryptophan fluorescence.



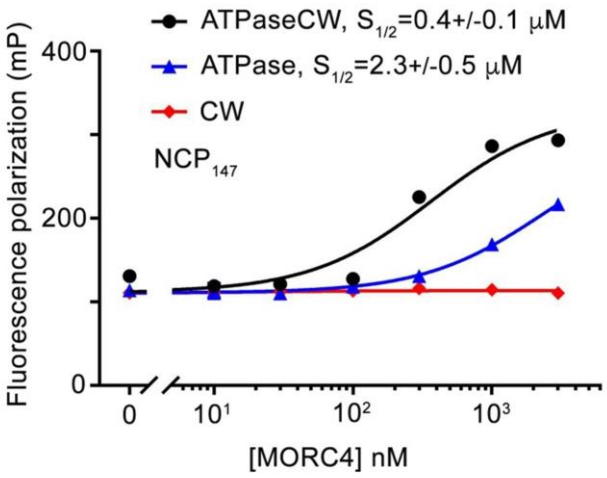
Supplementary Figure S6.5: Cross-linking assays using WT MORC4 ATPaseCW and I30A mutant of MORC4 ATPaseCW (designed based on sequence alignment with MORC3 to abolish dimerization) in the absence or presence of AMPPNP or ADP. AMPPNP but not ADP induced dimerization. Experiment with ATPaseCW (1-486) was repeated two times.



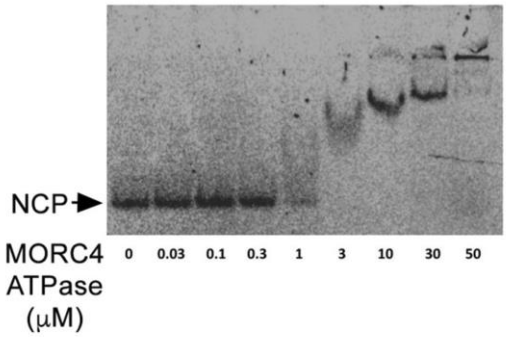
Supplementary Figure S6.6: EMSA with 601 DNA in the presence of 1 mM AMPPNP and increasing amounts of WT MORC4 ATPaseCW.



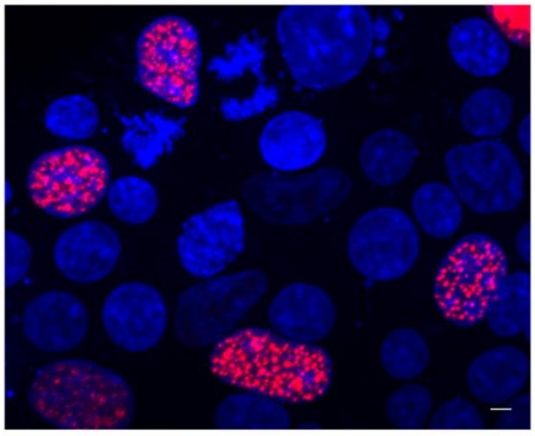
Supplementary Figure S6.7: EMSA with 5bp dsDNA ladder (50 ng) and increasing amounts of WT MORC4 ATPaseCW,
as indicated above the gel. Experiment was repeated three times.



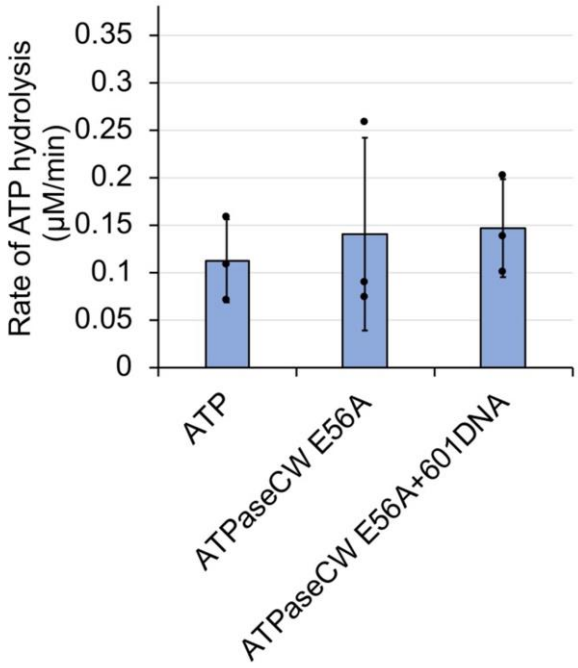
Supplementary Figure S6.8: Binding affinities and binding curves for the interactions of the indicated MORC4 regions with NCP147 as measured by fluorescence polarization. Data are represented as mean values +/- S.D. from three independent experiments (n=3).



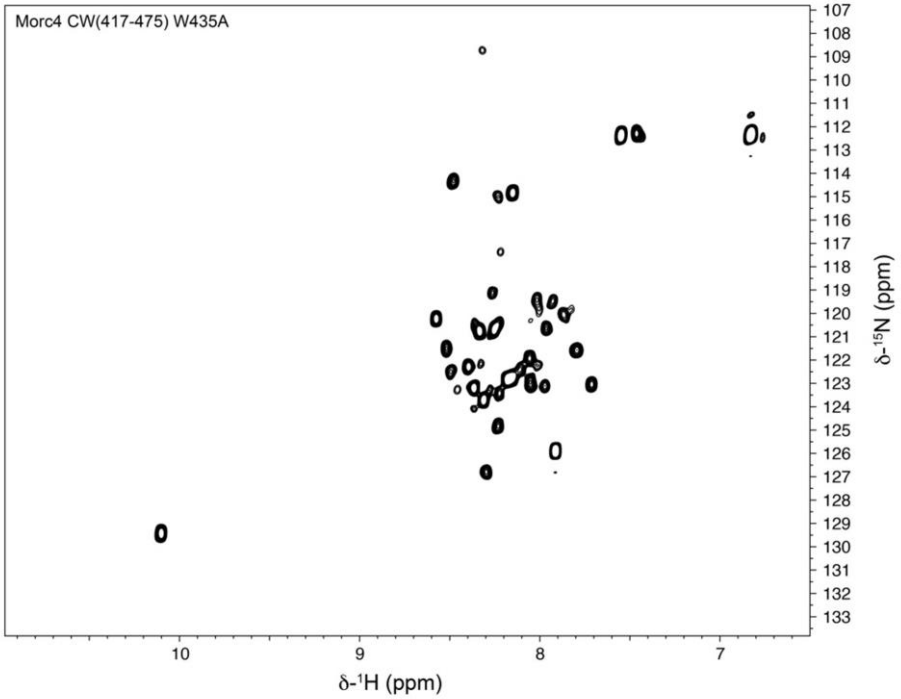
Supplementary Figure S6.9: EMSA with NCP_{147} in the presence of increasing amounts of the ATPase domain of MORC4.
Experiment was repeated three times.



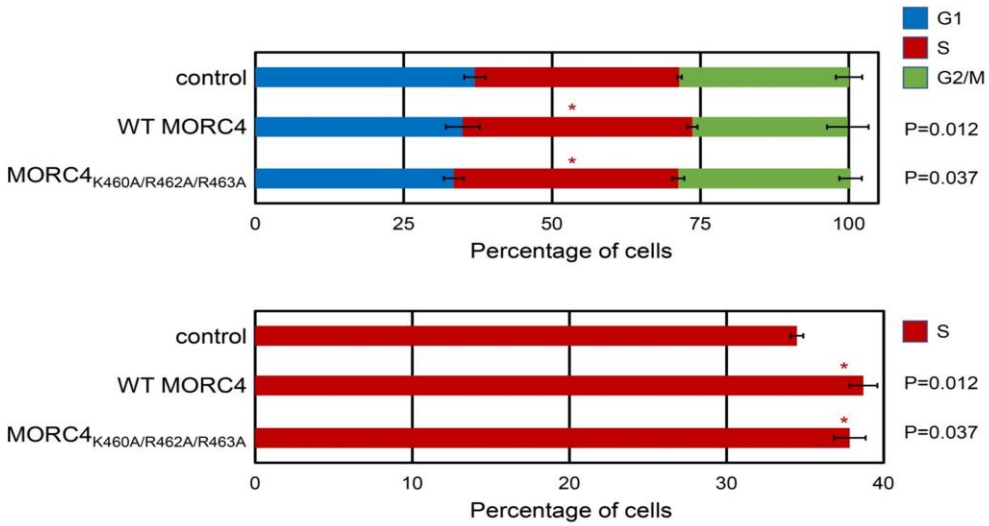
Supplementary Figure S6.10: Representative confocal microscopy images of 293T-HEK cells overexpressing mCherry-MORC4 E56A. Transfection were performed a minimum of 3 times, scale bar represents 5 μm.



Supplementary Figure S6.11: The E56A mutant of MORC4 is catalytically inactive but binds DNA as WT MORC4. Rates of ATP hydrolysis by the E56A mutant of the ATPaseCW cassette of MORC4 in the presence and absence of 601 DNA. Data are represented as mean values +/- S.D. from three independent experiments (n=3). Source data are provided in a Source Data file.



Supplementary Figure S6.12: $^{1}H,^{15}N$ HSQC spectrum of the ^{15}N-labeled MORC4 CW_{W435A} mutant indicates an unfolded protein.



Supplementary Figure S6.13: Cell cycle analysis following 48 hour overexpression of MORC4 proteins exhibit increased percent of cells in S phase for WT MORC4 (p=0.012) and MORC4 K460A/R462A/R463A mutant (p=0.037) compared to control cells. Data represent the average of three independent experiments. Error bars represent S.E.M, * indicates significant difference from mCherry-CTRL (p<0.05) by two-tailed student t-test.

Supplementary Table 1. Data collection and refinement statistics

	MORC4 ATPaseCW/AMPPN
Data collection	
Space group	P 1 21 1
Cell dimensions	
a, *b*, *c* (Å)	52.4, 109.9, 70.4
α, β, γ (°)	90.0, 96.0, 90.0
Resolution (Å)	2.9(3.1-2.9) *
R_{pim}	7.9(28.5)
$I / \sigma I$	10.1(2.0)
Completeness (%)	94.7(75.5)
Redundancy	3.5(2.4)
Refinement	
Resolution (Å)	47.1-2.9
No. reflections	15106
R_{work} / R_{free}	0.2397/0.2644
No. atoms	6338
ATPaseCW	6242
AMPPNP/Mg/Zn	66
Water	30
B-factors	39.66
ATPaseCW	39.76
AMPPNP/Mg/Zn	32.17
Water	34.64
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.588
Ramachandran plot	95.92
Most favored (%)	3.94
Allowed (%)	0.14
Outliers (%)	

*Values in parentheses are for highest-resolution shell.

Chapter 7. Discussion and Future Directions

There are many different histone post-translational reader proteins that bind to a wide array of different modifications made to each histone in the nucleosome octamer. In this work we tried to further characterize four of them containing domains that target methylated lysine residues. Additional information about their biological functions and mechanisms were learned, hopefully contributing to a better understanding of their cellular roles and the diseases each one has been linked to. In this chapter I will briefly summarize our findings for each protein and discuss possible next steps for each.

LEDGF is considered to be a key host factor for HIV-1 viral DNA integration into chromatin DNA. It is known that the viral DNA is targeted towards actively transcribing gene regions and seems to prefer nucleosomes over naked DNA for integration, specifically the major grooves of the nucleosomal DNA. Currently the mechanisms by which HIV-1 viral DNA is targeted to active genes are not well understood. Since LEDGF is known to interact with the HIV-1 integrase's catalytic core domain and guide the protein to sites of integration, we tried to add further information to how LEDGF interacts with chromatin and HIV-1 integrase in the context of nucleosomes with and without the H3K36me3 post translational modification. An epigenetic mark highly associated with regions of active genes, and a known binding target of LEDGF's PWWP domain. In HIV-1 viral DNA integration assays, LEDGF was shown to assist integration of the viral DNA into naked DNA and nucleosome substrates when combined with the HIV-1 integrase, but actually deterred integration when combined with the full pre-integration complex taken from infected T-cells which contained the integrase plus other factors associated with HIV DNA integration. In the presence of the H3K36me3 mark however, LEDGF was able to assist integration mediated by the pre-integration complex.

Previously we were unable to directly determine LEDGF's exact binding regime to the nucleosome due to the protein's inability to affect the FRET of our Cy3-Cy5 nucleosomes, but we were able to use anisotropy instead to acquire a clear binding curve of LEDGF to nucleosomes with and without a H3K36C tri-methyl lysine analog. With this, we were able to show that LEDGF decreases accessibility of nucleosomes to transcription factors like Gal4 by a factor of 2 at LEDGF binding saturation conditions. In single-molecule experiments, this decrease in accessibility appeared to be due to the protein reducing transcription factor binding on rate more so than off rate.

The new data in this work helps to further characterize how LEDGF interacts with chromatin lacking or containing the H3K36me3 mark in the process of HIV-1 DNA integration, but there remain inconsistencies in the data that keep LEDGF's mechanisms poorly understood. H3K36me3 is a mark found in actively transcribing genes, and LEDGF targets it, but in the FRET assays LEDGF appears to decrease accessibility of the nucleosome instead of increasing accessibility. The latter would've been more expected considering LEDGF plays a key role in assisting HIV-1 integration as shown in this work. Further studies need to be done to confirm LEDGF's true role when interacting with H3K36me3 modified chromatin. While the mark appeared to play a significant role in the HIV-1 DNA integration interactions with LEDGF, the mark appeared to have no effect in the LEDGF FRET assays showing the protein's impact on accessibility. The version of the H3K36me3 mark we use in this study is an analog that, while chemically is very similar to the tri-methylated lysine, has a sulfide in place of a methylene. This difference could possibly change how LEDGF and other H3K36me3-binding proteins interact with it. It is also possible that the conditions we used in the FRET assays were not enough to show the true nature of LEDGF's interaction with H3K36me3-modified chromatin. The FRET assays used 75 mM NaCl with no $MgCl_2$. It seems plausible that in higher, more physiological salt conditions hydrophobic interactions will play more of a role.

BRPF1 is a subunit of the native monocytic leukemic zinc-finger (MOZ) and MOZ-related factor (MORF) epigenetic writer complexes, which are capable of histone acetyltransferase activity. The full-length protein has multiple PTM-binding domains such as a histone lysine acetylation binding bromodomain, and a lysine methylation binding PWWP domain. In this work

we focused on characterizing the plant homeodomain-zinc knuckle-plant homeodomain (PZP) region of the BRPF1 protein. The plant homeodomains (PHD) have been found to bind to methylated lysines, particularly H3K4me3, as well as acetylated lysines occasionally. Here we study the binding mechanisms of the BRPF1 PZP domain to nucleosome chromatin in order to gain a better understanding of how this protein associates with MOZ/MORF chromatin substrates.

From structural analysis of the BRPF1 PZP domain interacting with an H3 tail peptide and nucleosomes, it appears that each PHD finger binds to a nucleosome differently. Negatively charged residues in the first PHD finger account for the domain's binding affinity to the H3 tail while positively charged residues in the second PHD finger and zinc knuckle bind to nucleosomal DNA. Binding assays utilizing mutations made to each PHD finger showed that the binding affinity for nucleosomal DNA was much higher compared to the H3 tail for the BRPF1 PZP domain.

While these results are interesting, more needs to be done. Additional information on which PTMs the BRPF1 PHD fingers target would be useful, whether that be H3K4me3 or others. While the PZP domain prefers nucleosomal DNA to the unmodified H3 tail, it wouldn't be surprising to see certain modifications raise the binding affinity of the domain to the H3 tail, even dominating the overall interaction compared to the binding of the nucleosomal DNA. This information would help us understand how the BRPF1 PZP domain targets the MOZ/MORF complexes to regions of differently modified chromatin and shed some light on the functional significance of the complexes' acetyltransferase activity.

CHD3 is a subunit of the Nucleosome Remodeling and Deactylation complex (NuRD). CHD3 plays a role in targeting the complex to specific areas of chromatin via two chromodomains and two PHD fingers. Since the PHD's were not well characterized, we chose to focus on them in this study. We isolated each PHD finger separately as well as the entire region of the protein linking both PHD fingers together and characterized their interaction with modified histone peptides and nucleosomes.

Through NMR and binding assays, we concluded that PHD1 and PHD2 of CHD3 bind preferentially to H3K9me3 and H3K9ac, with about the same affinity, over unmodified H3 tails

in the context of nucleosomes and H3 peptides, while FRET assays showed that the PHD1-2 linked region was able to increase the accessibility of H3K9me3-nucleosomes to the LexA transcription factor. This required both PHDs as neither PHD on its own was enough to have an effect on LexA binding to nucleosomes at PHD1/2 binding saturation. Interestingly, each PHD domain appears to have the capacity to bind an H3 tail, so it seems possible that both domains work together to impact the accessibility of the nucleosome. H3K4me3 actually inhibited binding of the PHD fingers, leading to lower binding affinity of the PHDs to the H3 tail even over unmodified H3 tails.

The PHD1-2 linked region targeting H3K9ac seems important for the NuRD complex localizing at acetylated promoter-proximal regions of targeted genes for deacetylase activity, but it is curious that the linked PHDs targets H3K9me3, a mark widely associated with silent genes, while opening the nucleosome for transcription alone. More studies need to be done to learn more about how CHD3 PHD1-2's impact on nucleosome accessibility plays a role in the NuRD complex. Since NuRD is capable of nucleosome remodeling, it seems plausible that the PHD fingers are working together to destabilize the nucleosome ahead of remodeling activity.

Human **MORC4** is the least characterized of the Morc family of chromatin binding proteins, and has been associated with acute and chronic pancreatitis and inflammatory bowel disorders. The protein has also been linked to breast cancer and diffuse large B cell lymphoma. Despite this, its biological functions remain largely unknown.

The Morc family proteins are characterized by an N terminal gyrase, hsp90, kinase, Mut L (GHKL) ATPase domain followed by a CW-type zinc finger domain. ATPase activity assays showed that the Morc4 GHKL domain had little ATPase activity on its own but binding of a Morc4 ATPase-CW cassette to DNA showed stimulation of the ATPase's activity. Binding and enzymatic assays showed that the CW domain is capable of binding to DNA in addition to the ATPase domain, and its interaction with DNA is important for ATPase activity. Furthermore, the Morc4 CW domain was shown to be an epigenetic reader that targets H3K4me3.

Most interestingly, FRET assays showed that binding of the protein to a nucleosome appeared to decrease accessibility of the nucleosome to TFs like Gal4, and that the ATPase activity of the GHKL domain plays a role since in the presence of ATP, Morc4's impact on nucleosome

accessibility was greater by a factor of ~2.5. This appears to indicate that Morc4 plays a role in gene silencing similarly to several plant Morcs and human Morc2. The mechanism behind how this happens remains unknown. It is possible that the Morc4 GHKL domain is capable of DNA gyrase activity, but how the typical DNA nicking action of a gyrase would lead to a decrease in nucleosome accessibility instead of an increase is unclear.

Further studies on the function of the Morc4 ATPase activity would assist in determining how Morc4 is able to impact chromatin accessibility by hydrolyzing ATP. Assays to confirm it can perform the functions of a DNA gyrase, such as the simple assay of showing it can nick supercoiled DNA to relax it, would be useful. It is also possible that human Morc4 could act as a chromatin remodeler similarly to human Morc2, which would help explain its role as a gene silencer.

Bibliography

1. Amber R. Cutter and Jeffrey J. Hayes. (2015). A brief review of nucleosome structure. *FEBS Letters*, 589(20):2914–2922.
2. K Luger, a W M¨ader, R K Richmond, D F Sargent, and T J Richmond. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260.
3. Brehove, M., Wang, T., North, J., Luo, Y., Dreher, S. J., Shimko, J. C., … Poirier, M. G. (2015). Histone core phosphorylation regulates DNA accessibility. *Journal of Biological Chemistry*, 290(37), 22612–22621.
4. Marfella, C. G., & Imbalzano, A. N. (2007). The Chd family of chromatin remodelers. *Mutation research*, 618(1-2), 30–40.
5. Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765), 41–45.
6. Huang, H., Lin, S., Garcia, B. A., & Zhao, Y. (2015). Quantitative proteomic analysis of histone modifications. *Chemical reviews*, 115(6), 2376–2418.
7. Bowman, G. D., & Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chemical reviews*, 115(6), 2274–2295.
8. Gardner, K. E., Allis, C. D., & Strahl, B. D. (2011). Operating on chromatin, a colorful language where context matters. *Journal of molecular biology*, 409(1), 36–46.
9. Zhao, Y., & Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor perspectives in biology*, 7(9), a025064.
10. Schwämmle, V., Aspalter, C. M., Sidoli, S., & Jensen, O. N. (2014). Large scale analysis of co-existing post-translational modifications in histone tails reveals global fine structure of cross-talk. *Molecular & cellular proteomics* : MCP, 13(7), 1855–1865.
11. Dien et al., 2013. (2008). Perceiving the epigenetic landscape through histone readers. *Bone*, 23(1), 1–7.
12. Paul R Selvin. (1995). [13] Fluorescence resonance energy transfer. In Biochemical Spectroscopy, volume 246 of *Methods in Enzymology*, pages 300–334. Academic Press.
13. Luo, Y., North, J. A., & Poirier, M. G. (2014). Single molecule fluorescence methodologies for investigating transcription factor binding kinetics to nucleosomes and *DNA*. *Methods*, 70(2–3), 108–118.
14. Donovan, B. T. (2019). Nucleosome Regulation of Transcription Factor Binding Kinetics : Implications for Gene Expression. [Doctoral Dissertation, The Ohio State University].
15. R. M. Clegg. (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods in Enzymology*, 211:353–388.
16. J. Lackowicz. “Fluorescence Anisotropy”. In: *Princ. Fluoresc. Spectrosc.* (2006), pp. 353–382. issn: 00037028. doi: 10.1007/3-540-29623-9_7040
17. M. Ameloot et al. “Fluorescence anisotropy measurements in solution: Methods and reference mate- rials (IUPAC Technical Report).” In: *Pure Appl. . . .* 85.3 (2013), pp. 589–608. issn: 1365-3075. doi: 10.1351/PAC-REP-11-11-12
18. B. Valeur. “Fluorescence Polarization. Emission Anisotropy”. In: *Mol. Fluoresc.* 8 (2001), pp. 125– 154. doi: 10.1002/3527600248.ch5
19. Valeur, B., & Berberan-Santos, M. N. (2012). Handbook of Fluorescence Spectroscopy and Imaging Fluorescence Applications in Biotechnology and Life Sciences Surface Enhanced Raman Spectroscopy Applied and Industrial Photochemistry.
20. Ruthenburg, A. J., Allis, C. D., & Wysocka, J. (2007). Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell*, 25(1), 15–30.
21. Honda, B. M., Dixon, G. H., & Candido, E. P. M. (1975). Sites of in vivo histone methylation in developing trout testis.

22. Dahl, J. A., Jung, I., Aanes, H., Greggains, G. D., Manaf, A., Lerdrup, M., … Klungland, A. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*, 537(7621), 548–552.
23. Park, S., Kim, G. W., Kwon, S. H., & Lee, J. S. (2020). Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *FEBS Journal*, 287(14), 2891–2902.
24. Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B., & Whittle, A. J. (2012). H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Journal of Signal Processing Systems*, 72(1), 1–16.
25. Shimon Weiss. Fluorescence Spectroscopy of Single Biomolecules. *Science*, 283(March):1676–1684, 1999.
26. Rahul Roy, Sungchul Hohng, and Taekjip Ha. (2008). A practical guide to single- molecule FRET. *Nat. Methods*, 5(6):507–516.
27. M.D. Gibson, M. Brehove, Y. Luo, J. North, and M.G. Poirier. (2016) Methods for Investigating DNA Accessibility with Single Nucleosomes, volume 581. *Elsevier Inc.*, 1 edition.
28. Ortiz-Hernandez, G. L., Sanchez-Hernandez, E. S., & Casiano, C. A. (2020). Twenty years of research on the DFS70/LEDGF autoantibody-autoantigen system: Many lessons learned but still many questions. *Autoimmunity Highlights*, 11(1).
29. LeRoy, G., Oksuz, O., Descostes, N., Aoi, Y., Ganai, R. A., Kara, H. O., … Reinberg, D. (2019). LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. *Science Advances*, 5(10), 1–12.
30. Blokken, J., De Rijck, J., Christ, F., & Debyser, Z. (2017). Protein–protein and protein–chromatin interactions of LEDGF/p75 as novel drug targets. *Drug Discovery Today: Technologies*, 24, 25–31.
31. Eidahl, J. O., Crowe, B. L., North, J. A., McKee, C. J., Shkriabai, N., Feng, L., … Kvaratskhelia, M. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Research*, 41(6), 3924–3936.
32. Padeken, J., Methot, S. P., & Gasser, S. M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nature Reviews Molecular Cell Biology*, 23(9), 623–640.
33. Grewal, S. I. S. (2023). The molecular basis of heterochromatin assembly and epigenetic inheritance. *Molecular Cell*, 83(11), 1767–1785.
34. Hall, M. A., Shundrovsky, A., Bai, L., Fulbright, R. M., Lis, J. T., & Wang, M. D. (2009). High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nature Structural and Molecular Biology*, 16(2), 124–129.
35. Li, G., Levitus, M., Bustamante, C., & Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. *Nature Structural and Molecular Biology*, 12(1), 46–53.
36. Hill, A. V. (1910). The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *The Journal of Physiology*, 40, i--vii.
37. Luo, Y., North, J. A., & Poirier, M. G. (2014). Single molecule fluorescence methodologies for investigating transcription factor binding kinetics to nucleosomes and DNA. *Methods (San Diego, Calif.)*, *70(2-3)*, 108–118.
38. Polach, K. J., & Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *Journal of molecular biology*, *254(2)*, 130–149.
39. Luo, Y., North, J. A., Rose, S. D., & Poirier, M. G. (2014). Nucleosomes accelerate transcription factor dissociation. *Nucleic acids research*, 42(5), 3017–3027.
40. Carpenter, E. J. W. and P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Bone*, 23(1), 1–7.
41. Allfrey, V. G., & Mirsky, A. E. (1964). Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science (New York, N.Y.)*, 144(3618), 559.
42. ALLFREY, V. G., FAULKNER, R., & MIRSKY, A. E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION

OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, 51(5), 786–794.
43. Sharda, A., & Humphrey, T. C. (2022). The role of histone H3K36me3 writers, readers and erasers in maintaining genome stability. *DNA Repair*, 119, 103407.
44. Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X., & Chen, E. S. (2022). Structural and functional specificity of H3K36 methylation. *Epigenetics and Chromatin*, 15(1), 1–20.
45. Chandradoss, S. D., Haagsma, A. C., Lee, Y. K., Hwang, J. H., Nam, J. M., & Joo, C. (2014). Surface passivation for single-molecule protein studies. *Journal of Visualized Experiments*, (86), 1–8.
46. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L., Jr, & Wiggins, C. H. (2009). Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophysical journal*, 97(12), 3196–3205.
47. Rosa, S., & Shaw, P. (2013). Insights into chromatin structure and dynamics in plants. *Biology*, 2(4), 1378–1410.
48. Kornberg R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science (New York, N.Y.),* 184(4139), 868–871.
49. Marsano, R. M., & Dimitri, P. (2022). Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. *Cells*, 11(5), 761.
50. Hattori, M. (2005). Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, 50(2), 162–168.
51. Miller, J. L., & Grant, P. A. (2013). The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Sub-cellular biochemistry*, 61, 289–317.
52. Kamakaka, R. T., & Biggins, S. (2005). Histone variants: deviants?. *Genes & development*, 19(3), 295–310.
53. Melters, D. P., Nye, J., Zhao, H., & Dalal, Y. (2015). Chromatin Dynamics in Vivo: A Game of Musical Chairs. *Genes*, 6(3), 751–776.
54. Suto, R. K., Clarkson, M. J., Tremethick, D. J., & Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nature structural biology*, 7(12), 1121–1124.
55. Chen, P., Wang, Y., & Li, G. (2014). Dynamics of histone variant H3.3 and its coregulation with H2A.Z at enhancers and promoters. *Nucleus (Austin, Tex.)*, 5(1), 21–27.
56. Watanabe, S., Radman-Livaja, M., Rando, O. J., & Peterson, C. L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science (New York, N.Y.),* 340(6129), 195–199.
57. Chawla A, Wang C, Patton C, Murray M, Punekar Y, de Ruiter A, Steinhart C. (2018). A review of long-term toxicity of antiretroviral treatment regimens and implications for an aging population. *Infect Dis Ther 7*: 183–195.
58. Wilen CB, Tilton JC, Doms RW. (2012). HIV: cell binding and entry. *Cold Spring Harb Perspect Med 2*:a006866.
59. Wyatt R, Sodroski J. (1998). The HIV-1 envelope glycoproteins: fusogens, anti-gens, and immunogens. *Science* 280:1884–1888..
60. Campbell EM, Hope TJ. (2015). HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat Rev Microbiol* 13:471–483.
61. Fassati A. (2012). Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res* 170:15–24.
62. Gres AT, Kirby KA, KewalRamani VN, Tanner JJ, Pornillos O, Sarafianos SG. (2015). Structural virology: X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* 349:99–103.
63. Mattei S, Glass B, Hagen WJH, Kräusslich H-G, Briggs JAG. (2016). The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* 354:1434–1437.
64. Perilla JR, Gronenborn AM. (2016). Molecular architecture of the retroviral capsid. *Trends Biochem* Sci 41:410–420.
65. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 497: 643–646.

66. Hu WS, Hughes SH. (2012). HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2:a006882.
67. Craigie R, Bushman FD. (2012). HIV DNA integration. *Cold Spring Harb Perspect Med* 2:a006890.
68. Bowerman B, Brown PO, Bishop JM, Varmus HE. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev 3*: 469–478.
69. Farnet CM, Haseltine WA. (1990). Integration of human immunodeficiency virus type 1 DNA in vitro. *Proc Natl Acad Sci U S A* 87:4164–4168.
70. Farnet CM, Haseltine WA. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J Virol* 65:1910–1915.
71. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M. (1993). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1with viral nucleic acids following acute infection. *Proc Natl Acad Sci U S A* 90:6125–6129.
72. Ellison V, Brown PO. (1994). A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc Natl Acad Sci U S A* 91:7316–7320.
73. Harris D, Engelman A. (2000). Both the structure and DNA binding function of the barrier-to-autointegration factor contribute to reconstitution of HIV type 1 integration in vitro. *J Biol Chem* 275:39671–39677.
74. Deeks SG, Archin N, Cannon P, Collins S, Jones RB, de Jong MAWP, Lambotte O, Lamplough R, Ndung'u T,Sugarman J,Tiemessen CT, Vandekerckhove L, Lewin SR (2021). Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy. *Nature Medicine* 27:2085–2098.
75. Chun T-W, Stuyver L, Mizell SB, Ehler LA, Mican JAM, Baseler M, Lloyd AL, Nowak MA, Fauci AS. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A* 94:13193–13197
76. Chun TW, Fauci AS. (2012). HIV reservoirs: pathogenesis and obstacles to viral eradication and cure. *AIDS* 26:1261–1268.
77. Lesbats P, Engelman AN, Cherepanov P. (2016). Retroviral DNA integration. *Chem Rev* 116:12730–12757.
78. Maertens GN, Engelman AN, Cherepanov P. (2022). Structure and function of retroviral integrase. *Nat Rev Microbiol* 20:20–34.
79. Métifiot M, Marchand C, Pommier Y. (2013). HIV integrase inhibitors: 20- year landmark and challenges. *Adv Pharmacol* 67:75–105.
80. Mbhele N, Chimukangara B, Gordon M. (2021). HIV-1 integrase strand transfer inhibitors: a review of current drugs, recent advances and drug resistance. *Int J Antimicrob Agents* 57:106343.
81. Craigie R. (2018). Nucleoprotein intermediates in HIV-1 DNA integration: structure and function of HIV-1 intasomes. *Subcell Biochem* 88:189–210.
82. Engelman A, Mizuuchi K, Craigie R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67:1211–1221
83. Bushman FD, Craigie R. (1991). Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci U S A* 88:1339–1343.
84. Fujiwara T, Mizuuchi K. (1988). Retroviral DNA integration: structure of an integration intermediate. *Cell* 54:497–504.
85. Brown PO, Bowerman B, Varmus HE, Bishop JM. (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci U S A* 86:2525–2529.
86. Serrao E, Krishnan L, Shun M-C, Li X, Cherepanov P, Engelman A, Maertens GN. (2014). Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res* 42:5164–5176.
87. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. (2002). HIV-1 integration in the human genome favors active genes and local hot spots. *Cell* 110:521–529.

88. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2:E234.
89. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res* 17:1186–1194.
90. Brady T, Agosto LM, Malani N, Berry CC, O'Doherty U, Bushman F. (2009). HIV integration site distributions in resting and activated CD41 T cells infected in culture. *AIDS* 23:1461–1471.
91. Wang GP, Levine BL, Binder GK, Berry CC, Malani N, McGarrity G, Tebas P, June CH, Bushman FD. (2009). Analysis of lentiviral vector integration in HIV1 study subjects receiving autologous infusions of gene modified CD41 T cells. *Mol Ther* 17:844–850.
92. Cattoglio C, Pellin D, Rizzi E, Maruggi G, Corti G, Miselli F, Sartori D, Guffanti A, Di Serio C, Ambrosi A, De Bellis G, Mavilio F. (2010). High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* 116: 5507–5517.
93. Roth SL, Malani N, Bushman FD. (2011). Gammaretroviral integration into nucleosomal target DNA in vivo. *J Virol* 85:7393–7401.
94. Taganov KD, Cuesta I, Daniel R, Cirillo LA, Katz RA, Zaret KS, Skalka AM. (2004). Integrase-specific enhancement and suppression of retroviral DNA integration by compacted chromatin structure in vitro. *J Virol* 78:5848–5855.
95. Lesbats P,Botbol Y,Chevereau G, Vaillant C, Calmels C,Arneodo A, Andreola M-L, Lavigne M, Parissi V. (2011). Functional coupling between HIV- 1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog* 7:e1001280.
96. Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, Pongor S, Luzzati R, Recchia A, Mavilio F, Giacca M, Lusic M. (2015). Nuclear architecture dictates HIV-1 integration site selection. *Nature* 521:227–231.
97. Matysiak J, Lesbats P, Mauro E, Lapaillerie D, Dupuy J-W, Lopez AP, Benleulmi MS, Calmels C, Andreola M-L, Ruff M, Llano M, Delelis O, Lavigne M, Parissi V. (2017). Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology* 14:39.
98. Lelek M, Casartelli N, Pellin D, Rizzi E, Souque P, Severgnini M, Di Serio C, Fricke T, Diaz-Griffero F, Zimmer C, Charneau P, Di Nunzio F. (2015). Chromatin organization at the nuclear pore favors HIV replication. *Nat Commun* 6:6483.
99. Bedwell GJ, Jang S, Li W, Singh PK, Engelman AN. (2021). rigrag: high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. *Nucleic Acids Res* 49:7330–7346.
100. Albanese A, Arosio D, Terreni M, Cereseto A. (2008). HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery. *PLoS One* 3:e2413.
101. Sowd GA, Serrao E, Wang H, Wang W, Fadel HJ, Poeschla EM, Engelman AN. (2016). A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad Sci U S A* 113:E1054–63.
102. Lucic B, Chen H-C, Kuzman M, Zorita E, Wegner J, Minneker V, Wang W, Fronza R, Laufs S, Schmidt M, Stadhouders R, Roukos V, Vlahovicek K, Filion GJ, Lusic M. (2019). Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat Commun* 10:4059.
103. Elleder D, Pavlícek A, Paces J, Hejnar J. (2002). Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. *FEBS Lett* 517:285–286.
104. Ho Y-C, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD, Siliciano RF. (2013). Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155:540–551.
105. Cohn LB,Silva IT,Oliveira TY, RosalesRA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD,Siliciano

RF,Jankovic M,Nussenzweig MC. (2015). HIV-1 integration landscape during latent and active infection. *Cell* 160:420–432.
106. McGinty RK, Tan S. (2015). Nucleosome structure and function. *Chem Rev* 115:2255–2273.
107. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260.
108. Simpson RT. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17: 5524–5531.
109. Allan J, Hartman PG, Crane-Robinson C, Aviles FX. (1980). The structure of histone H1 and its location in chromatin. *Nature* 288:675–679.
110. Woodcock CL, Skoultchi AI, Fan Y. (2006). Role of linker histone in chroma- tin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res* 14:17–25.
111. Bowman GD, Poirier MG. (2015). Posttranslational modifications of histones that influence nucleosome dynamics. *Chem Rev* 115:2274–2295.
112. Allis CD, Jenuwein T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17:487–500.
113. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11:1287–1289.
114. Emiliani S, Mousnier A, Busschots K, Maroun M,Van Maele B,Tempé D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, Christ F, Rain J-C, Dargemont C, Debyser Z, Benarous R. (2005). Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J Biol Chem* 280:25517–25523.
115. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM. (2006). An essential role for LEDGF/p75 in HIV integration. *Science* 314:461–464.
116. Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, Poeschla EM. (2006). Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J Mol Biol* 360: 760–773.
117. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. (2006). Biochemical and genetic analyses of integrase-interacting proteins lens epithelium- derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology* 346:415–426.
118. Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, Van den Haute C, Witvrouw M, Debyser Z. (2006). Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. *J Virol* 80:1886–1896.
119. Cherepanov P. (2007). LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res* 35:113–124.
120. Hombrouck A, De Rijck J, Hendrix J, Vandekerckhove L, Voet A, De Maeyer M, Witvrouw M, Engelborghs Y, Christ F, Gijsbers R, Debyser Z. (2007). Virus evolution reveals an exclusive role for LEDGF/p75 in chromosomal tethering of HIV. *PLoS Pathog* 3:e47.
121. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD. (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* 2:e1340.
122. Shun M-C, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, Engelman A. (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specificHIV-1 integration. *Genes Dev* 21:1767–1778.
123. Engelman A, Cherepanov P. (2008). The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog* 4:e1000046.
124. BalasubramaniamM, Zhou J,Addai A,Martinez P,Pandhare J,Aiken C, Dash C. (2019). PF74 inhibits HIV-1 integration by altering the composi- tion of the preintegration complex. *J Virol* 93.
125. Addai AB, Pandhare J, Paromov V, Mantri CK, Pratap S, Dash C. (2015). Cocaine modulates HIV-1 integration in primary CD41 T cells: implications in HIV-1 pathogenesis in drug-abusing patients. *J Leukoc Biol* 97: 779–790.

126. Balasubramaniam M, Davids B, Addai AB, Pandhare J, Dash C. (2017). Measurement of in vitro integration activity of HIV-1 preintegration complexes. *J Vis Exp* 2017:54581.
127. Davids B-O, Balasubramaniam M, Sapp N, Prakash P, Ingram S, Li M, Craigie R, Hollis T, Pandhare J, Dash C. (2021). Human three prime repair exonuclease 1 promotes HIV-1 integration by preferentially degrading unprocessed viral DNA. *J Virol* 95:e0055521.
128. Balasubramaniam M, Davids B-O, Bryer A, Xu C, Thapa S, Shi J, Aiken C, Pandhare J, Perilla JR, Dash C. (2022). HIV-1 mutants that escape the cyto- toxic T-lymphocytes are defective in viral DNA integration. *PNAS Nexus* 1:pgac064.
129. Schnitzler GR. (2001). Isolation of histones and nucleosome cores from mammalian cells. *Curr Protoc Mol Biol Chapter* 21:Unit 21.5.
130. Van Holde KE. (1989). Chromatin. In Springer series in molecular biology, XII, p 497. Springer, New York, NY.
131. Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, Lindemann D, Engelman AN, Costa A, Cherepanov P. (2015). Structural basis for retroviral integration into nucleosomes. *Nature* 523:366–369.
132. Lowary PT, Widom J. (1998). New DNA sequence rules for high affinity bind- ing to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 276:19–42.
133. Luger K, Rechsteiner TJ, Richmond TJ. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* 304:3–19.
134. Li M, Craigie R. (2009). Nucleoprotein complex intermediates in HIV-1 integration. *Methods* 47:237–242.
135. Li M, Jurado KA, Lin S, Engelman A, Craigie R. (2014). Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One* 9:e105078.
136. Christensen DE, Ganser-Pornillos BK, Johnson JS, Pornillos O, Sundquist WI. (2020). Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science* 370:eabc8420.
137. Thåström A, Lowary PT, Widlund HR, Cao H, Kubista M, Widom J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol* 288:213–229.
138. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. (2009). High throughput characterization of combinatorial his- tone codes. *Mol Cell Proteomics* 8:2266–2284.
139. Rando OJ. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev* 22:148–155.
140. Battivelli E, Dahabieh MS, Abdel-Mohsen M, Svensson JP, Tojal Da Silva I, Cohn LB, Gramatica A, Deeks S, Greene WC, Pillai SK, Verdin E. (2018). Dis- tinct chromatin functional states correlate with HIV latency reactivation in infected primary CD41 T cells. *Elife* 7:e34655.
141. Wu VH, Nobles CL, Kuri-Cervantes L, McCormick K, Everett JK, Nguyen S, del Rio Estrada PM, González-Navarro M, Torres-Ruiz F, Ávila-Ríos S, Reyes-Terán G, Bushman FD, Betts MR. (2020). Assessment of HIV-1 integration in tissues and subsets across infection stages. *JCI Insight* 5: e139783.
142. Musselman CA, Gibson MD, Hartwick EW, North JA, Gatchalian J, Poirier MG, Kutateladze TG. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun* 4:2969.
143. Gibson MD, Gatchalian J, Slater A, Kutateladze TG, Poirier MG. (2017). PHF1 Tudor and N-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. *Nucleic Acids Res* 45:3767–3776.
144. Simon MD, Chu F, Racki LR, de la Cruz CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* 128:1003–1012.
145. Compton JL, Bellard M, Chambon P. (1976). Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes. *Proc Natl Acad Sci U S A* 73:4382–4386.

146. Burge NL, Thuma JL, Hong ZZ, Jamison KB, Ottesen JJ, Poirier MG. (2022). H1.0 C terminal domain is integral for altering transcription factor binding within nucleosomes. *Biochemistry* 61:625–638.
147. Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. (2014). Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* 42:10209–10225.
148. LusicM, Siliciano RF. (2017). Nuclear landscape of HIV-1 infection and integration. *Nat Rev Microbiol* 15:69–82.
149. Engelman AN, Singh PK. (2018). Cellular and molecular mechanisms of HIV-1 integration targeting. *Cell Mol Life Sci* 75:2491–2507.
150. Poletti V, Mavilio F. (2018). Interactions between retroviruses and the host cell genome. *Mol Ther Methods Clin Dev* 8:31–41.
151. Meehan AM, Poeschla EM. (2010). Chromatin tethering and retroviral integration: recent discoveries and parallels with DNA viruses. *Biochim Biophys Acta* 1799:182–191.
152. LeRoy G, Oksuz O, Descostes N, Aoi Y, Ganai RA, Kara HO, Yu J-R, Lee C- H, Stafford J, Shilatifard A, Reinberg D. (2019). LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. *Sci Adv* 5:eaay3068.
153. Cherepanov P, Devroe E, Silver PA, Engelman A. (2004). Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem* 279:48883–48892.
154. Cherepanov P, Sun Z-YJ, Rahman S, Maertens G, Wagner G, Engelman A. (2005). Solution structure of the HIV-1 integrase-binding domain in LEDGF/ p75. *Nat Struct Mol Biol* 12:526–532.
155. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, Poeschla E. (2005). Identification of the LEDGF/p75 HIV-1 integrase-interaction do- main and NLS reveals NLS-independent chromatin tethering. *J Cell Sci* 118:1733–1743.
156. Pandey KK, Sinha S, Grandgenett DP. (2007). Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase- mediated concerted integration. *J Virol* 81:3969–3979.
157. Shimko JC, North JA, Bruns AN, Poirier MG, Ottesen JJ. (2011). Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. *J Mol Biol* 408:187–204.
158. Le JV, Luo Y, Darcy MA, Lucas CR, Goodwin MF, Poirier MG, Castro CE. (2016). Probing nucleosome stability with a DNA Origami Nanocaliper. *ACS Nano* 10:7073–7084.
159. Wang H, Farnung L, Dienemann C, Cramer P. (2020). Structure of H3K36- methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. *Nat Struct Mol Biol* 27:8–13.
160. Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, Plumb M, Graham RL, Gorelick RJ, Hess S, Poirier MG, Foster MP, Kvaratskhelia M. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mono- nucleosomes. *Nucleic Acids Res* 41:3924–3936.
161. Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M, Grewal SIS. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* 451:734–737.
162. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49.
163. Woo H, Dam Ha S, Lee SB, Buratowski S, Kim T. (2017). Modulation of gene expression dynamics by cotranscriptional histone methylations. *Exp Mol Med* 49:e326.
164. Wagner EJ, Carpenter PB. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 13:115–126.
165. Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li G-M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSa. *Cell* 153:590–600.
166. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543:72–77.

167. Sundarraj J, Taylor GCA, von Kriegsheim A, Pradeepa MM. (2017). H3K36me3 and PSIP1/LEDGF associate with several DNA repair proteins, suggesting their role in efficient DNA repair at actively transcribing loci. *Wellcome Open Res* 2:83.
168. Vansant G, Chen H-C, Zorita E, Trejbalová K, Miklík D, Filion G, Debyser Z. (2020). The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res* 48:7801–7817.
169. Vansant G, Vranckx LS, Zurnic I, Van Looveren D, Van de Velde P, Nobles C, Gijsbers R, Christ F, Debyser Z. (2019). Impact of LEDGIN treatment during virus production on residual HIV-1 transcription. *Retrovirology* 16:8.
170. Robinson PJ, Rhodes D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* 16:336–343.
171. Bednar J, Garcia-Saez I, Boopathi R, Cutter AR, Papai G, Reymer A, Syed SH, Lone IN, Tonchev O, Crucifix C, Menoni H, Papin C, Skoufias DA, Kurumizaka H, Lavery R, Hamiche A, Hayes JJ, Schultz P, Angelov D, Petosa C, Dimitrov S. (2017). Structure and dynamics of a 197 bp nucleo- some in complex with linker histone H1. *Mol Cell* 66:384–397.
172. Sancho M, Diani E, Beato M, Jordan A. (2008). Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet* 4:e1000227.
173. Chaudhuri E, Dash S, Balasubramaniam M, Padron A, Holland J, Sowd GA, Villalta F, Engelman AN, Pandhare J, Dash C. (2020). The HIV-1 capsid- binding host factor CPSF6 is posttranscriptionally regulated by the cellu- lar microRNA miR-125b. *J Biol Chem* 295:5081–5094.
174. Benleulmi MS, Matysiak J, Robert X, Miskey C, Mauro E, Lapaillerie D, LesbatsP,Chaignepain S, Henriquez DR, CalmelsC,Oladosu O, Thierry E, Leon O, Lavigne M,Andreola M-L,Delelis O, IvicsZ,Ruff M,Gouet P, Parissi V. (2017). Modulation of the functional association between the HIV-1 inta- some and the nucleosome by histone amino-terminal tails. *Retrovirology* 14:54.
175. Ehrenhofer-Murray AE. (2004). Chromatin dynamics at DNA replication, transcription and repair. *Eur J Biochem* 271:2335–2349.
176. Ransom M, Dennehey BK, Tyler JK. (2010). Chaperoning histones during DNA replication and repair. *Cell* 140:183–195.
177. Lai WKM, Pugh BF. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* 18: 548–562.
178. Mackler RM, Lopez MA, Osterhage MJ, Yoder KE. (2018). Prototype foamy virus integrase is promiscuous for target choice. *Biochem Biophys Res Commun* 503:1241–1246.
179. Mackler RM, Jones ND, Gardner AM, Lopez MA, Howard CJ, Fishel R, Yoder KE. (2019). Nucleosome DNA unwrapping does not affect proto- type foamy virus integration efficiency or site selection. *PLoS One* 14: e0212764.
180. Kotlar RM, Jones ND, Senavirathne G, Gardner AM, Messer RK, Tan YY, Rabe AJ, Fishel R, Yoder KE. (2021). Retroviral prototype foamy virus intasome binding to a nucleosome target does not determine integration efficiency. *J Biol Chem* 296:100550.
181. Rabe AJ, Tan YY, Larue RC, Yoder KE. (2021). Prototype foamy virus inte- grase displays unique biochemical activities among retroviral integrases. *Biomolecules* 11:1910.
182. Wilson MD, Renault L, Maskell DP, Ghoneim M, Pye VE, Nans A, Rueda DS, Cherepanov P, Costa A. (2019). Retroviral integration into nucleo- somes through DNA looping and sliding along the histone octamer. *Nat Commun* 10:4189.
183. Maertens GN, Hare S, Cherepanov P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468: 326–329.
184. Pasi M, Mornico D, Volant S, Juchet A, Batisse J, Bouchier C, Parissi V, Ruff M, Lavery R, Lavigne M. (2016). DNA minicircles clarify the specific role of DNA structure on retroviral integration. *Nucleic Acids Res* 44:7830–7847.
185. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474:516–520.
186. Chereji RV, Bryson TD, Henikoff S. (2019). Quantitative MNase-seq accu- rately maps nucleosome occupancy levels. *Genome Biol* 20:198.

187. Frenster JH, Allfrey VG, Mirsky AE. (1963). Repressed and active chromatin isolated from interphase lymphocytes. *Proc Natl Acad Sci U S A* 50: 1026–1032.
188. Huisinga KL, Brower-Toland B, Elgin SC. (2006). The contradictory defini- tions of heterochromatin: transcription and silencing. *Chromosoma* 115: 110–122.
189. Bannister AJ, Kouzarides T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21:381–395.
190. Matzke MA, Mosher RA. (2014). RNA-directed DNA methylation: an epige- netic pathway of increasing complexity. *Nat Rev Genet* 15:394–408.
191. De Ravin SS, Su L, Theobald N, Choi U, Macpherson JL, Poidinger M, Symonds G, Pond SM, Ferris AL, Hughes SH, Malech HL, Wu X. (2014). Enhancers are major targets for murine leukemia virus vector integration. *J Virol* 88:4504–4513.
192. Meehan AM, Saenz DT, Morrison JH, Garcia-Rivera JA, Peretz M, Llano M, Poeschla EM. (2009). LEDGF/p75 proteins with alternative chromatin tethers are functional HIV-1 cofactors. *PLoS Pathog* 5:e1000522.
193. Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, Wang GG, Shun M-C, Allis CD, Engelman A, Hughes SH. (2010). Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci U S A* 107:3135–3140.
194. Braaten D, Ansari H, Luban J. (1997). The hydrophobic pocket of cyclophilin is the binding site for the human immunodeficiency virus type 1 Gag polypro- tein. *J Virol* 71:2107–2113.
195. Bedwell GJ, Engelman AN. (2021). Factors that mold the nuclear land- scape of HIV-1 integration. *Nucleic Acids Res* 49:621–635.
196. Van Nuland R, van Schaik FM, Simonis M, van Heesch S, Cuppen E, Boelens R, Timmers HM, van Ingen H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1- PWWP domain. *Epigenetics Chromatin* 6:12.
197. Shun M-C, Botbol Y, Li X, Di Nunzio F, Daigle JE, Yan N, Lieberman J, Lavigne M, Engelman A. (2008). Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human im- munodeficiency virus type 1 infectivity. *J Virol* 82:11555–11567.
198. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J, Engelborghs Y, Christ F, Debyser Z. (2006). Overexpression of the lens epithelium-derived growth factor/p75 integrase binding do- main inhibits human immunodeficiency virus replication. *J Virol* 80: 11498–11509.
199. Meehan AM, Saenz DT, Morrison J, Hu C, Peretz M, Poeschla EM. (2011). LEDGF dominant interference proteins demonstrate prenuclear exposure of HIV-1 integrase and synergize with LEDGF depletion to destroy viral infectivity. *J Virol* 85:3570–3583.
200. Morchikh M, Naughtin M, Di Nunzio F, Xavier J, Charneau P, Jacob Y, Lavigne M. (2013). TOX4 and NOVA1 proteins are partners of the LEDGF PWWP domain and affect HIV-1 replication. *PLoS One* 8:e81217.
201. Satchwell SC, Drew HR, Travers AA. (1986). Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* 191:659–675.
202. Chua EYD, Vasudevan D, Davey GE, Wu B, Davey CA. (2012). The mechan- ics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res* 40:6338–6352.
203. Katzman M, Sudol M. (1995). Mapping domains of retroviral integrase respon- sible for viral DNA specificity and target site selection by analysis of chimera between human immunodeficiency virus type 1 and visna virus integrases. *J Virol* 69:5687–5696.
204. Shibagaki Y, Chow SA. (1997). Central core domain of retroviral integrase is responsible for target site selection. *J Biol Chem* 272:8361–8369.
205. Janssen A, Colmenares SU, Karpen GH. (2018). Heterochromatin: guardian of the genome. *Annu Rev Cell Dev Biol* 34:265–288.
206. Terme J-M, Sesé B, Millán-Ariño L, Mayor R, Belmonte JCI, Barrero MJ, Jordan A. (2011). Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J Biol Chem* 286:35347–35357.

207. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 9:279–289.
208. Clark DJ, Thomas JO. (1986). Salt-dependent co-operative interaction of histone H1 with linear DNA. *J Mol Biol* 187:569–580.
209. Fang H, Clark DJ, Hayes JJ. (2012). DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain. *Nucleic Acids Res* 40: 1475–1484.
210. Xiao B, Freedman BS, Miller KE, Heald R, Marko JF. (2012). Histone H1 com- pacts DNA under force and during chromatin assembly. *Mol Biol Cell* 23: 4864–4871.
211. Platt EJ, Bilska M, Kozak SL, Kabat D, Montefiori DC. (2009). Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immu- nodeficiency virus type 1. *J Virol* 83:8289–8292.
212. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. (1998). Effects of CCR5 and CD4 cell surface concentrations on infections by macrophage- tropic isolates of human immunodeficiency virus type 1. *J Virol* 72: 2855–2864.
213. Clegg RM. (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol* 211:353–388.
214. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011). Integrative genomics viewer. *Nat Biotechnol* 29:24–26.
215. Allen, H.F., Wade, P.A., and Kutateladze, T.G. (2013). The NuRD architecture. *Cell. Mol. Life Sci.* 70, 3513–3524.
216. Allen HF, Wade PA, Kutateladze TG. (2013). The NuRD architecture. *Cellular and molecular life sciences:* CMLS. 70:3513–3524.
217. Andrews FH, Strahl BD, Kutateladze TG. (2016). Insights into newly discovered marks and readers of epigenetic information. *Nature chemical biology*. 12:662–668.
218. Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, Francis D, Bredel M, Vogel H, Mills AA. (2007). CHD5 is a tumor suppressor at human 1p36. *Cell*. 128:459–475.
219. Bergs JW, Neuendorff N, van der Heijden G, Wassenaar E, Rexin P, Elsasser HP, Moll R, Baarends WM, Brehm A. (2014). Differential expression and sex chromosome association of CHD3/4 and CHD5 during spermatogenesis. *PloS one*. 9:e98203.
220. Cai Y, Geutjes EJ, de Lint K, Roepman P, Bruurs L, Yu LR, Wang W, van Blijswijk J, Mohammad H, de Rink I, et al. (2014). The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. *Oncogene*. 33:2157–2168.
221. Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ, Al-Sady B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Molecular cell*. 41:67–81.
222. Canzio D, Liao M, Naber N, Pate E, Larson A, Wu S, Marina DB, Garcia JF, Madhani HD, Cooke R, et al. (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature*. 496:377–381.
223. Clapier CR, Cairns BR. (2009). The biology of chromatin remodeling complexes. *Annual review of biochemistry*. 78:273–304.
224. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 447:799–816.
225. Cosgrove MS, Boeke JD, Wolberger C. (2004). Regulated nucleosome mobility and the histone code. *Nature structural & molecular biology*. 11:1037–1043.
226. Denslow SA, Wade PA. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene*. 26:5433–5438.
227. Dyer PN, Edayathumangalam RS, White CL, Bao Y, Chakravarthy S, Muthurajan UM, Luger K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in enzymology*. 375:23–44.

228. Garske AL, Oliver SS, Wagner EK, Musselman CA, LeRoy G, Garcia BA, Kutateladze TG, Denu JM. (2010). Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nature chemical biology*. 6:283–290.
229. Gibson MD, Gatchalian J, Slater A, Kutateladze TG, Poirier MG. (2017). PHF1 Tudor and N-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. *Nucleic acids research*.
230. Gomez-Del Arco P, Perdiguero E, Yunes-Leites PS, Acin-Perez R, Zeini M, Garcia-Gomez A, Sreenivasan K, Jimenez-Alcazar M, Segales J, Lopez-Maderuelo D, et al. (2016). The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. *Cell metabolism*. 23:881–892.
231. Goodarzi AA, Kurka T, Jeggo PA. (2011). KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nature structural & molecular biology*. 18:831–839.
232. Ikebe J, Sakuraba S, Kono H. (2016). H3 Histone Tail Conformation within the Nucleosome and the Impact of K14 Acetylation Studied Using Enhanced Sampling Simulation. *PLoS computational biology*. 12:e1004788.
233. Ingram KG, Curtis CD, Silasi-Mansat R, Lupu F, Griffin CT. (2013). The NuRD chromatin-remodeling enzyme CHD4 promotes embryonic vascular integrity by transcriptionally regulating extracellular matrix proteolysis. *PLoS genetics*. 9:e1004031.
234. Klement K, Luijsterburg MS, Pinder JB, Cena CS, Del Nero V, Wintersinger CM, Dellaire G, van Attikum H, Goodarzi AA. (2014). Opposing ISWI- and CHD-class chromatin remodeling activities orchestrate heterochromatic DNA repair. *The Journal of cell biology*. 207:717–733.
235. Kolla V, Naraparaju K, Zhuang T, Higashi M, Kolla S, Blobel GA, Brodeur GM. (2015). The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. *The Biochemical journal*. 468:345–352.
236. Kolla V, Zhuang T, Higashi M, Naraparaju K, Brodeur GM. (2014). Role of CHD5 in human cancers: 10 years later. *Cancer research*. 74:652–658.
237. Lai AY, Wade PA. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature reviews Cancer*. 11:588–596.
238. Li G, Widom J. (2004). Nucleosomes facilitate their own invasion. *Nature structural & molecular biology*. 11:763–769.
239. Li W, Mills AA. (2014). Architects of the genome: CHD dysfunction in cancer, developmental disorders and neurological syndromes. *Epigenomics*. 6:381–395.
240. Mansfield RE, Musselman CA, Kwan AH, Oliver SS, Garske AL, Davrazou F, Denu JM, Kutateladze TG, Mackay JP. (2011). Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *The Journal of biological chemistry*. 286:11779–11791.
241. Mishra A, La Perle K, Kwiatkowski S, Sullivan LA, Sams GH, Johns J, Curphey DP, Wen J, McConnell K, Qi J, et al. (2016). Mechanism, Consequences, and Therapeutic Targeting of Abnormal IL15 Signaling in Cutaneous T-cell Lymphoma. *Cancer discovery*. 6:986–1005.
242. Moshkin YM, Chalkley GE, Kan TW, Reddy BA, Ozgur Z, van Ijcken WF, Dekkers DH, Demmers JA, Travers AA, Verrijzer CP. (2012). Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Molecular and cellular biology*. 32:675–688.
243. Musselman CA, Gibson MD, Hartwick EW, North JA, Gatchalian J, Poirier MG, Kutateladze TG. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nature communications*. 4:2969.
244. Musselman CA, Kutateladze TG. (2011). Handpicking epigenetic marks with PHD fingers. *Nucleic acids research*. 39:9061–9071.
245. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. (2012). Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology*. 19:1218–1227.
246. Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, O'Leary H, Denu JM, Mackay JP, Kutateladze TG. (2009). Binding of the CHD4 PHD2 finger to

histone H3 is modulated by covalent modifications. *The Biochemical journal*. 423:179–187.
247. Musselman CA, Ramirez J, Sims JK, Mansfield RE, Oliver SS, Denu JM, Mackay JP, Wade PA, Hagman J, Kutateladze TG. (2012). Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proceedings of the National Academy of Sciences of the United States of America*. 109:787–792.
248. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*. 154:490–503.
249. Nitarska J, Smith JG, Sherlock WT, Hillege MM, Nott A, Barshop WD, Vashisht AA, Wohlschlegel JA, Mitter R, Riccio A. (2016). A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell reports*. 17:1683–1698.
250. North JA, Amunugama R, Klajner M, Bruns AN, Poirier MG, Fishel R. (2013). ATP-dependent nucleosome unwrapping catalyzed by human RAD51. *Nucleic acids research*.
251. Oliver SS, Musselman CA, Srinivasan R, Svaren JP, Kutateladze TG, Denu JM. (2012). Multivalent recognition of histone tails by the PHD fingers of CHD5. *Biochemistry*. 51:6534–6544.
252. Patel DJ, Wang Z. (2013). Readout of epigenetic modifications. *Annual review of biochemistry*. 82:81–118.
253. Paul S, Kuo A, Schalch T, Vogel H, Joshua-Tor L, McCombie WR, Gozani O, Hammell M, Mills AA. (2013). Chd5 requires PHD-mediated histone 3 binding for tumor suppression. *Cell reports*. 3:92–102.
254. Pegoraro G, Kubben N, Wickert U, Gohler H, Hoffmann K, Misteli T. (2009). Ageing-related chromatin defects through loss of the NURD complex. *Nature cell biology*. 11:1261–1267.
255. Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. *Nature*. 502:65–70.
256. Ramirez J, Dege C, Kutateladze TG, Hagman J. (2012). MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes. *Molecular and cellular biology*. 32:5078–5088.
257. Reynolds N, Latos P, Hynes-Allen A, Loos R, Leaford D, O'Shaughnessy A, Mosaku O, Signolet J, Brennecke P, Kalkan T, et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell stem cell*. 10:583–594.
258. Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A, Bertone P, Hendrich B. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *The EMBO journal*. 31:593–605.
259. Schuettengruber B, Simboeck E, Khier H, Seiser C. (2003). Autoregulation of mouse histone deacetylase 1 expression. *Molecular and cellular biology*. 23:6993–7004.
260. Simon MD, Chu F, Racki LR, de la Cruz CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell*. 128:1003–1012.
261. Stanley FK, Moore S, Goodarzi AA. (2013). CHD chromatin remodelling enzymes and the DNA damage response. *Mutation research*. 750:31–44.
262. Strahl BD, Allis CD. (2000) The language of covalent histone modifications. *Nature*. 403:41–45.
263. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology*. 14:1025–1040.
264. Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*. 395:917–921.

265. Wade PA, Jones PL, Vermaak D, Wolffe AP. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Current biology:* CB. 8:843–846.
266. Wang X, Hayes JJ. (2007). Site-specific binding affinities within the H2B tail domain indicate specific effects of lysine acetylation. *The Journal of biological chemistry*. 282:32867–32876.
267. Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*. 482:221–225.
268. Wille A, Amort T, Singewald N, Sartori SB, Lusser A. (2016). Dysregulation of select ATP-dependent chromatin remodeling factors in high trait anxiety. *Behavioural brain research*. 311:141–146.
269. Xie W, Ling T, Zhou Y, Feng W, Zhu Q, Stunnenberg HG, Grummt I, Tao W. (2012). The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proceedings of the National Academy of Sciences of the United States of America*. 109:8161–8166.
270. GT, Young MK, Cote J, Wang W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular cell*. 2:851–861.
271. Yamada T, Yang Y, Hemberg M, Yoshida T, Cho HY, Murphy JP, Fioravante D, Regehr WG, Gygi SP, Georgopoulos K, et al. (2014). Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron*. 83:122–134.
272. Yang Y, Yamada T, Hill KK, Hemberg M, Reddy NC, Cho HY, Guthrie AN, Oldenborg A, Heiney SA, Ohmae S, et al. (2016). Chromatin remodeling inactivates activity genes and regulates neural coding. *Science*. 353:300–305.
273. Zhang J, Jackson AF, Naito T, Dose M, Seavitt J, Liu F, Heller EJ, Kashiwagi M, Yoshida T, Gounari F, et al. (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nature immunology* 13:86–94.
274. Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*. 95:279–289.
275. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, *Biological crystallography* 66, 213–221.
276. Avvakumov N, Lalonde ME, Saksouk N, Paquet E, Glass KC, Landry AJ, Doyon Y, Cayrou C, Robitaille GA, Richard DE, et al. (2012). Conserved molecular interactions within the HBO1 acetyltransferase complexes regulate cell proliferation. *Molecular and cellular biology* 32, 689–703.
277. Bamborough P, Barnett HA, Becher I, Bird MJ, Chung CW, Craggs PD, Demont EH, Diallo H, Fallon DJ, Gordon LJ, et al. (2016). GSK6853, a Chemical Probe for Inhibition of the BRPF1 Bromodomain. *ACS Med Chem Lett* 7, 552–557.
278. Chen S, Yang Z, Wilkinson AW, Deshpande AJ, Sidoli S, Krajewski K, Strahl BD, Garcia BA, Armstrong SA, Patel DJ, and Gozani O (2015). The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. *Molecular cell* 60, 319–327.
279. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, and Richardson DC (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica. Section D, *Biological crystallography* 66, 12–21.
280. Demeulenaere S, Beysen D, De Veuster I, Reyniers E, Kooy F, and Meuwissen M (2019). Novel BRPF1 mutation in a boy with intellectual disability, coloboma, facial nerve palsy and hypoplasia of the corpus callosum. *Eur J Med Genet* 62, 103691.

281. Doyon Y, and Cote J (2016). Preparation and Analysis of Native Chromatin-Modifying Complexes. *Methods in enzymology* 573, 303–318.
282. Dyer PN, Edayathumangalam RS, White CL, Bao Y, Chakravarthy S, Muthurajan UM, and Luger K (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in enzymology* 375, 23–44.
283. Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, Plumb M, Graham RL, Gorelick RJ, Hess S, et al. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic acids research* 41, 3924–3936.
284. Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta crystallographica. Section D, *Biological crystallography* 66, 486–501.
285. Gatchalian J, Wang X, Ikebe J, Cox KL, Tencer AH, Zhang Y, Burge NL, Di L, Gibson MD, Musselman CA, et al. (2017). Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nature communications* 8, 1489.
286. Gibson MD, Gatchalian J, Slater A, Kutateladze TG, and Poirier MG (2017). PHF1 Tudor and N-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. *Nucleic acids research*.
287. Hibiya K, Katsumoto T, Kondo T, Kitabayashi I, and Kudo A (2009). Brpf1, a subunit of the MOZ histone acetyl transferase complex, maintains expression of anterior and posterior Hox genes for proper patterning of craniofacial and caudal skeletons. *Developmental biology* 329, 176–190.
288. Huang F, Abmayr SM, and Workman JL (2016). Regulation of KAT6 Acetyltransferases and Their Roles in Cell Cycle Progression, *Stem Cell Maintenance, and Human Disease*. *Molecular and cellular biology* 36, 1900–1907.
289. Igoe N, Bayle ED, Fedorov O, Tallant C, Savitsky P, Rogers C, Owen DR, Deb G, Somervaille TC, Andrews DM, et al. (2017). Design of a Biased Potent Small Molecule Inhibitor of the Bromodomain and PHD Finger-Containing (BRPF) Proteins Suitable for Cellular and in Vivo Studies. *Journal of medicinal chemistry* 60, 668–680.
290. Klein BJ, and Jang e.a. (2019). Histone H3K23-specific acetylation by MORF is coupled to H3K14 acylation. *Nature communications*.
291. Klein BJ, Lalonde ME, Cote J, Yang XJ, and Kutateladze TG (2014). Crosstalk between epigenetic readers regulates the MOZ/MORF HAT complexes. *Epigenetics* 9, 186–193.
292. Klein BJ, Muthurajan UM, Lalonde ME, Gibson MD, Andrews FH, Hepler M, Machida S, Yan K, Kurumizaka H, Poirier MG, et al. (2016). Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation. *Nucleic acids research* 44, 472–484.
293. Klein BJ, Piao L, Xi Y, Rincon-Arano H, Rothbart SB, Peng D, Wen H, Larson C, Zhang X, Zheng X, et al. (2014). The Histone-H3K4-Specific Demethylase KDM5B Binds to Its Substrate and Product through Distinct PHD Fingers. *Cell reports* 6, 325–335.
294. Lalonde ME, Avvakumov N, Glass CK, Joncas FH, Saksouk N, Holliday M, Paquet E, Yan K, Tong Q, Klein BJ, et al. (2013). Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev* in press.
295. Lalonde ME, Avvakumov N, Glass KC, Joncas FH, Saksouk N, Holliday M, Paquet E, Yan K, Tong Q, Klein BJ, et al. (2013). Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes Dev* 27, 2009–2024.
296. Laue K, Daujat S, Crump JG, Plaster N, Roehl HH, Kimmel CB, Schneider R, and Hammerschmidt M (2008). The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development* 135, 1935–1946.
297. Mattioli F, Schaefer E, Magee A, Mark P, Mancini GM, Dieterich K, Von Allmen G, Alders M, Coutton C, van Slegtenhorst M, et al. (2017). Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis. *American journal of human genetics* 100, 105–116.
298. Miller TC, Simon B, Rybin V, Grotsch H, Curtet S, Khochbin S, Carlomagno T, and Muller CW (2016). A bromodomain-DNA interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT. *Nature communications* 7, 13855.
299. Minor W, Cymborowski M, Otwinowski Z, and Chruszcz M (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial

model in minutes. Acta crystallographica. Section D*, Biological crystallography* 62, 859–866.
300. Mishima Y, Miyagi S, Saraya A, Negishi M, Endoh M, Endo TA, Toyoda T, Shinga J, Katsumoto T, Chiba T, et al. (2011). The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis. *Blood* 118, 2443–2453.
301. Musselman CA, Gibson MD, Hartwick EW, North JA, Gatchalian J, Poirier MG, and Kutateladze TG (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nature communications* 4, 2969.
302. Pode-Shakked N, Barel O, Pode-Shakked B, Eliyahu A, Singer A, Nayshool O, Kol N, Raas-Rothschild A, Pras E, and Shohat M (2019). BRPF1-associated intellectual disability, ptosis, and facial dysmorphism in a multiplex family. *Mol Genet Genomic Med* 7, e665.
303. Poplawski A, Hu K, Lee W, Natesan S, Peng D, Carlson S, Shi X, Balaz S, Markley JL, and Glass KC (2013). Molecular Insights into the Recognition of N-Terminal Histone Modifications by the BRPF1 Bromodomain. *Journal of molecular biology*.
304. Tencer AH, Cox KL, Di L, Bridgers JB, Lyu J, Wang X, Sims JK, Weaver TM, Allen HF, Zhang Y, et al. (2017). Covalent Modifications of Histone H3K9 Promote Binding of CHD3. *Cell reports* 21, 455–466.
305. Ullah M, Pelletier N, Xiao L, Zhao SP, Wang K, Degerny C, Tahmasebi S, Cayrou C, Doyon Y, Goh SL, et al. (2008). Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Molecular and cellular biology* 28, 6828–6843
306. van Nuland R, van Schaik FM, Simonis M, van Heesch S, Cuppen E, Boelens R, Timmers HM, and van Ingen H (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics & chromatin* 6, 12.
307. Vezzoli A, Bonadies N, Allen MD, Freund SM, Santiveri CM, Kvinlaug BT, Huntly BJ, Gottgens B, and Bycroft M (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature structural & molecular biology* 17, 617–619.
308. Yan K, Rousseau J, Littlejohn RO, Kiss C, Lehman A, Rosenfeld JA, Stumpel CTR, Stegmann APA, Robak L, Scaglia F, et al. (2017). Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. *American journal of human genetics* 100, 91–104.
309. Yang XJ (2015). MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. *Biochimica et biophysica acta* 1853, 1818–1826.
310. You L, Yan K, Zhou J, Zhao H, Bertos NR, Park M, Wang E, and Yang XJ (2015). The lysine acetyltransferase activator Brpf1 governs dentate gyrus development through neural stem cells and progenitors. *PLoS genetics* 11, e1005034.
311. You L, Zou J, Zhao H, Bertos NR, Park M, Wang E, and Yang XJ (2015b). Deficiency of the chromatin regulator BRPF1 causes abnormal brain development. *The Journal of biological chemistry* 290, 7114–7129.
312. You LY, Li L, Zou JF, Yan KZ, Belle J, Nijnik A, Wang E, and Yang XJ (2016). BRPF1 is essential for development of fetal hematopoietic stem cells. *Journal of Clinical Investigation* 126, 3247–3262.
313. Zhu J, and Caflisch A (2016). Twenty Crystal Structures of Bromodomain and PHD Finger Containing Protein 1 (BRPF1)/Ligand Complexes Reveal Conserved Binding Motifs and Rare Interactions. *Journal of medicinal chemistry* 59, 5555–5561.
314. Zhu J, Zhou C, and Caflisch A (2018). Structure-based discovery of selective BRPF1 bromodomain inhibitors. *Eur J Med Chem* 155, 337–352.
315. Angsuwatcharakon P, Sodsai P, Rerknimitr R, Hirankarn N. (2018). The Polymorphisms at PRSS1-PRSS2 and MORC4 loci and the risk of Post-ERCP pancreatitis. *Gastroenterol. Res, Pr*. 2018:1064783.
316. Derikx MH, et al. (2015). Polymorphisms at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. *Gut*. 64:1426–1433.
317. Duan X, et al. (2019). Baicalin inhibits cell viability, migration and invasion in breast cancer by regulating miR-338-3p and MORC4. *OncoTargets Ther*. 12:11183–11193.

318. Hong G, et al. (2017). The emerging role of MORC family proteins in cancer development and bone homeostasis. *J. Cell Physiol.* 232:928–934.
319. Liggins AP, et al. (2007). MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. *Br. J. Haematol.* 138:479–486.
320. Paliwal S, et al. (2016). Association analysis of PRSS1-PRSS2 and CLDN2-MORC4 variants in nonalcoholic chronic pancreatitis using tropical calcific pancreatitis as model. *Pancreas.* 45:1153–1157.
321. Soderman J, et al. (2013). Analysis of single nucleotide polymorphisms in the region of CLDN2-MORC4 in relation to inflammatory bowel disease. *World J. Gastroenterol.* 19:4935–4943.
322. Weiss FU, et al. (2018). Common variants in the CLDN2-MORC4 and PRSS1-PRSS2 loci confer susceptibility to acute pancreatitis. *Pancreatology*. 18:477–481.
323. Yang Z, Zhuang Q, Hu G, Geng S. (2019). MORC4 is a novel breast cancer oncogene regulated by miR-193b-3p. *J. Cell Biochem*. 120:4634–4643.
324. Li DQ, Nair SS, Kumar R. (2013). The MORC family: new epigenetic regulators of transcription and DNA damage response. *Epigenetics*. 8:685–693.
325. Dutta R, Inouye M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* 25:24–28.
326. Li X, et al. (2012). Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J. Am. Chem. Soc.* 134:1982–1985.
327. Andrews FH, et al. (2016). Multivalent chromatin engagement and inter-domain crosstalk regulate MORC3 ATPase. *Cell Rep*. 16:3195–3207.
328. Li S, et al. Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *Proc. Natl Acad. Sci. USA*. 2016;113:E5108–E5116.
329. Zhang, Y. et al. (2019). Mechanism for autoinhibition and activation of the MORC3 ATPase. *Proc. Natl Acad. Sci. USA* 10.1073/pnas.1819524116.
330. Liu Y, et al. (2016). Family-wide characterization of histone binding abilities of human CW domain-containing proteins. *J. Biol. Chem.* 291:9000–9013.
331. Douse CH, et al. (2018). Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nat. Commun*. 9:651.
332. Tchasovnikarova IA, et al. (2017). Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat. Genet*. 49:1035–1044.
333. Zhang Y, et al. (2019). MORC3 is a target of the influenza A viral protein NS1. *Structure.* 27:1029–1033.e3.
334. Zhang Y, et al. (2019). MORC3 forms nuclear condensates through phase separation. *iScience.* 17:182–189.
335. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* 19:1218–1227.
336. Andrews FH, Strahl BD, Kutateladze TG. (2016). Insights into newly discovered marks and readers of epigenetic information. *Nat. Chem. Biol.* 12:662–668.
337. Li G, Widom J. (2014). Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* 11:763–769.
338. Luo Y, North JA, Rose SD, Poirier MG. (2014). Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res*. 42:3017–3027.
339. Kim H, et al. (2019). The gene-silencing protein MORC-1 topologically entraps DNA and forms multimeric assemblies to cause DNA compaction. *Mol. Cell*. 75:700–710 e6.
340. Li DQ, et al. (2012). MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response. *Cell Rep*. 2:1657–1669.
341. Liu Y, et al. (2016). Family-wide characterization of histone binding abilities of human CW domain containing proteins. *J. Biol. Chem*. 291:9000–9013.
342. Emsley P, Lohkamp B, Scott WG, Cowtan K. (2010). Features and development of Coot. Acta Crystallogr. *D. Biol. Crystallogr*. 66:486–501.
343. Adams PD, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. *D. Biol. Crystallogr*. 66:213–221

344. Chen VB, et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. *D. Biol. Crystallogr*. 66:12–21.
345. Petell CJ, Pham AT, Skela J, Strahl BD. (2019). Improved methods for the detection of histone interactions with peptide microarrays. *Sci. Rep*. 9:6265.
346. Klein BJ, et al. (2016). Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation. *Nucleic Acids Res*. 44:472–484.
347. Tencer AH, et al. (2017). Covalent modifications of histone H3K9 promote binding of CHD3. *Cell Rep*. 21:455–466.
348. Gibson MD, Gatchalian J, Slater A, Kutateladze TG, Poirier MG. (2017). PHF1 Tudor and N-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. *Nucleic Acids Res* 45:3767–3776.

Appendices

A. Detailed Protocols

A.1 Measuring protein to nucleosome binding with anisotropy

Adapted from: Klein, B. J., Cox, K. L., Jang, S. M., Singh, R. K., Côté, J., Poirier, M. G., & Kutateladze, T. G. (2021). Structural and biophysical characterization of the nucleosome-binding PZP domain. *STAR Protocols*, 2(2).

Anisotropy is used to measure binding of a molecule to another fluorescently labeled molecule while exciting the fluorescently labeled molecule with polarized light. Typically this is a good method to use when you have a large molecule binding to a much smaller molecule labeled with a long excitation lifetime fluorophore like fluorescein. The long lifetime fluorophore helps ensure the initial emitted light is unpolarized/isotropic, such that when the larger molecule binds to the smaller labeled one, the resulting emitted light will then become polarized/anisotropic from the reduction in rotational diffusion of the labeled molecule. For nucleosomes, typically we have a smaller protein molecule binding to the larger 200 kDa nucleosome instead, so in our case it appears that binding of a protein causes anisotropy by reducing the rotational freedom of the fluorescein placed on the nucleosomal DNA in the entry-exit region. Supporting this theory, proteins that do not bind the nucleosomal DNA and only bind the histone tails or other sites, are difficult to measure binding to the nucleosome with anisotropy [347].

Measurements are taken using a Tecan Infinite with built in fluorescence polarization functionality. Reactions are mixed (outside of) and then placed into Costar black 98-well plates for ensemble measurements.

For the below steps, store all dilutions on ice

1. Prepare reaction buffer.
   a. This can be altered to suit the protein or desired physiological conditions (this thesis used 25 mM Tris-HCl pH 7.5, 75 mM NaCl, 0.00625% (v/v) Tween-20, and 5 mM DTT)

2. Using reaction buffer, prepare concentration dilutions of the binding protein over 2 orders of magnitude around the dissociation constant of the protein binding to nucleosomes in order to observe the inflection point and binding saturation during the titration. If the dissociation constant is 10 nM then make 0.3, 1, 3, 10, 30, 100 nM final concentration dilutions for example.
   a. Also prepare one set of dilutions for background measurements
   b. Prepare each dilution in triplicate in addition to the dilution for background, (~ 110 µL total each for a triplicate titration with background using 30 µL reaction volume)
   c. Mixing the dilutions with nucleosomes to start the reactions will cause further dilution, upscale the dilution concentrations accordingly to make sure the final concentrations are correct in each reaction

Note: If the dissociation constant for the reaction can not be found in published literature, then try to go as high as supply allows, or first attempt an EMSA to save protein.

3. Prepare a dilution of nucleosomes in 0.5x TE to 50 nM, again, enough for a triplicate (~ 21 µL using 30 µL reaction volume).

4. Mix 3 µL of nucleosome with 27 µL of each protein dilution to start the reactions with 5 nM final nucleosome concentration. Keep on ice for 30 mins.

a. For zero protein concentration, 27 μL of reaction buffer is used instead

b. For background measurements use 3 μL of 0.5x TE instead of the NCP sample.

5. Pipette reaction mixtures into wells of the Costar black plate slowly to avoid creating bubbles.

6. Load plate into the Tecan plate reader.

7. Initialize i-control and use the fluorescence module to design a protocol that will measure fluorescence polarization of each reaction by exciting the samples in the wells with 470 nm polarized light and measuring emission at 520 nm (for fluorescein) using the following settings:

Parameter	**Setting**
Bandwidth	5 nm
Gain	optimal
G-factor	To be measured for each instrument (1.02 for the Tecan used in this thesis)
Lag time	0 ms
Number of flashes	50
Settle time	0 ms
Z-position	Calculated from well

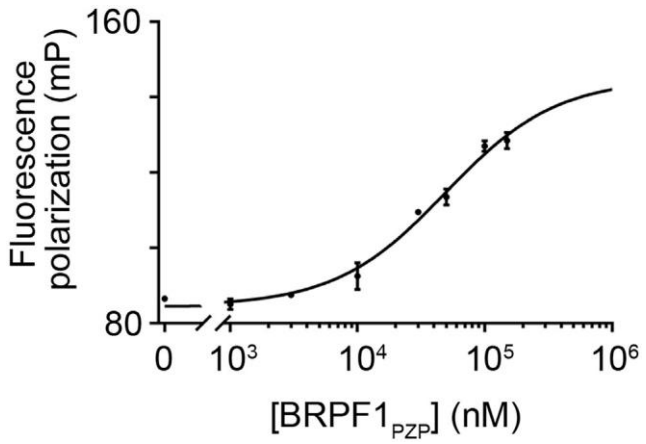


Figure A.1: Anisotropy binding curve obtained for the interaction of wild-type $BRPF1_{PZP}$ with NCP_{207}, titration done in triplicate and fit with equation (2.15)

8. Run the protocol. The instrument will automatically calculate anisotropy and polarization values, but the output options can be configured to also output raw data of the polarized emission light intensity parallel and perpendicular to the plane of excitation for each well, allowing calculation of anisotropy and polarization manually using equations (2.6) and (2.8).
   a. If using a new instrument or changing fluorophores, a new G-factor will need to be measured. This involves a simple measurement of the fluorophore alone against a blank. Refer to the instrument manual for detailed steps on how to perform a G-factor calibration.

Note: If the fluorescence polarization binding assay is outputting data with large deviations between data points, consider increasing the reaction volume. Stirring the reaction mixtures in wells with a pipette tip to make sure they are sitting evenly in the wells may help as well.

9. Export the data.

10. Analyze the data using preferred graphing software (GraphPad Prism, Origin, Gnuplot, etc) (Figure A.1).

A.2 Site-Specific Deposition of Trimethyl-lysine Analog

This protocol is based on Simon et al. [260]. It was adapted for use in the Poirier lab by Justin North in 2010 and further adapted to improve labeling efficiency by Matthew Gibson in 2012. It should be noted that this reaction does not create the true trimethylation PTM. Instead, the lysine γ-methylene is substituted with a sulfide in the side chain of the residue.

Before performing the labeling, this protocol is designed for labeling of 5-10 mg histones, and the histones should be mutated to isolate a single Cys residue at the desired labeling site of the MLA.

A.2.1 Materials

• (2-bromoethyl)-trimethylammonium bromide (Sigma 117196)

• Alkylation Buffer

– 1 M HEPES

Note: 310 mM HEPES (Sigma H3034) and 690 mM glshepes sodium salt (Sigma H7006) should yield the correct pH of 7.8. Adjust with HEPES (down) and HEPES sodium salt (up).

– 10 mM D/L-Methionine (Sigma M9500)

– 4 M Guanidine-HCl (MP Biomedicals 820539)

– Flash freeze and store at -80°C in 1.5 mL aliquots

• 1M DTT (Sigma 43816-50ML)

• 14.3 M BME (Sigma M6250-100ML)

A.2.2 Labeling

1. Resuspend 5-10 mg histone to be labeled in 980 µL of alkylation buffer and allow to unfold for 1 h.

2. While unfolding the histones, set up a heated stir plate with a 1.5 mL tube block. Place a 1.5 mL tube containing ~1mL water with a stir bar, thermocouple, and parafilm cap to secure the thermocouple probe inside of the tube containing water on the block. Cover from light and air currents. Heat the block to 50°C while stirring at maximum speed (typically temperature set to ~55-60°C results in 50°C as measured with the thermocouple). Check the temperature often over the next few hours to ensure a stable 50°C.

3. Mix the unfolded histone and transfer to a clean 1.5 mL tube.

4. Add 20 µL 333 mm DTT (diluted with water from 1M stock) and allow to reduce at 37°C for 1 h.

5. Transfer the histones to a new 1.5 mL tube containing. 100 mg (2-bromoethyl) tri-methylammonium bromide and a flea stir bar. Place in the heated block and replace the light/air cover. Keep using the tube containing water to measure temperature changes. Avoid placing the thermocouple probe directly into the alkylation buffer with histones.

6. Allow the reaction to proceed for 2.5 h.

7. Add another 20 µL 333 mm DTT to the reaction.

8. Allow the reaction to proceed an additional 2.5 h.

9. Quench the reaction with 50 µL 14.3 m BME.

10. Dialyse against at least 4 1.8 L changes of 3 mm BME.

11. Lyophilize the protein.

12. Resuspend the protein in dH2O and measure the concentration via UV-VIS spectrophotometry absobance.

13. Aliquot into 1–10×(0.1–1 mg for H3, equal molar for other histones) aliquots.

14. Vacuum concentrate the samples to dryness and store at -80°C

A.2.3 Storage

Histones harboring methyl-lysine-analog (MLA) labels appear to be stable for years at -80°C. However, once resuspended they appear to lose their methyl groups on the order of weeks. It has been determined that storage of the MLA-labeled sample in 0.5 mm potassium phosphate pH 7.5 mitigates this loss of label and extends the lifetime of the sample to typical levels. It appears the free amines in Tris attack the methyl groups and remove them so any buffer used to store MLA-labeled histones should be free of amines.

A.2.4 Mass Spectrometry Verification

Labeling efficiency must be verified by mass spectrometry. The trimethyl label theoretically adds 86 Da to the protein, so a comparison can be made between the mutated histone before and after labeling via examination of the protein mass peak (Figure A.2).

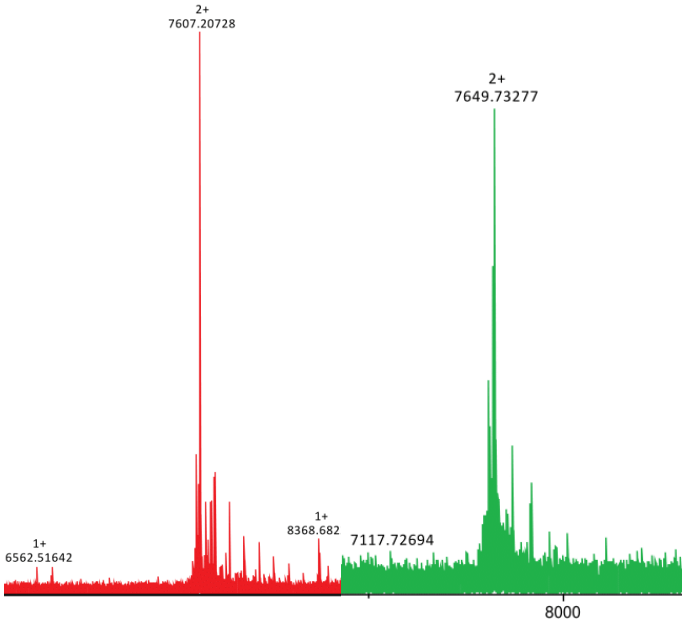


Figure A.2: Liquid chromatography-mass spectrometry done with H3K36C histone before (red) and after (green) trimethylation. Pictured is the 2+ peak (protein has charge $2e^+$) of the H3 histone with the x-axis being mass over charge. Examination of the peak confirms successful labeling with a mass shift of ~40 Da for the 2+ peak, which is very close to the theoretical added value. This chromatogram was provided by the Mass Spectrometry and Proteomics Facility at The Ohio State University.

A.3 Refolding and Purification of Multiply Labeled Histone Octamers

Since both the Cy5 and trimethylation labeling methods used in this thesis target cysteines, it is necessary to separate the labeling sites for each labeling reaction instead of performing both on a fully formed octamer with multiple cysteines. As such, for octamers with multiple label types it is recommended to design the octamer with one cysteine on the H3/H4 tetramer and one on the H2A/H2B dimer. Another consideration is labeling is usually most efficient on histone complexes, especially when the labeling site is located on a histone tail, and so we first fold H2A into H2A/H2B dimers before Cy5 maleimide labeling while depositing the methyl-lysine-analog on H3 histones in unfolding conditions.

1. Refold H2A/H2B and H3/H4 separately. The MLA should already be deposited, but the fluorophore should be deposited while folded, if possible.

   Note: The H3/H4 tetramer can be refolded using typical Tris-based buffers as the loss of the MLA label to Tris is a slow process as mentioned above in A.2.3.

2. Label the H2A/H2B dimer with the fluorescent label by the typical protocol.

3. Purify the Cy5-H2A/H2B dimer from free dye by size exclusion as if it were octamer.

4. Combine tetramer and dimer to a final molar ratio of 1:2.2 tetramer:dimer.

   Note: The tetramer can be in typical refolding buffer during this step, again, as mentioned above.

5. Allow to rotate at 4°C on a lab rotisserie overnight.

6. Purify the octamer by size exclusion replacing refolding buffer with 5 mM PIPES pH 6.1, 2 M sodium chloride.

7. Store the MLA labeled octamer in 5 mM PIPES pH 6.1, 2 M sodium chloride. Bring to 40% glycerol final concentration for storage at -20°C.

   Note: Nucleosomes containing the MLA label should be stored in 0.5 mM potassium phosphate pH 7.5 as mentioned in A.2.3 with 0.5 mM EDTA added in order to avoid Tris based buffers.