ABSTRACT

Title of dissertation: MOLECULAR DYNAMIC SIMULATIONS OF NUCLEOSOMES AND HISTONE TAILS: THE EFFECTS OF HISTONE VARIANCE AND POST-TRANSLATIONAL MODIFICATION

David N. Winogradoff, Doctor of Philosophy, 2015

Dissertation directed by: Professor Garegin Papoian
Department of Chemistry and Biochemistry

The packaging of genomic information and the regulation of gene expression are both fundamentally important to eukaryotic life. Meters of human DNA must fit inside the micron-diameter nucleus while still rapidly becoming available for templated processes such as transcription, replication, and repair. Therefore, the DNA-protein complex known as chromatin must dynamically transition between more compact, closed states and more accessible, open ones. To fully understand chromatin structure and dynamics, it is necessary to employ a multifaceted approach, integrating different general philosophies and scientific techniques that include experiment and computation. Since the DNA in chromatin is organized into arrays of nucleosomes, we take a bottom-up approach in this dissertation, striving first to understand the structure and dynamics of an individual nucleosome and subdomains thereof. Atomistic computational methods have provided useful tools to study DNA and protein dynamics at the nanosecond, and recently microsecond, timescale. In

this dissertation, we present recent developments in the understanding of the nucleosome though atomistic molecular dynamics (MD) simulations. By applying different all-atom MD computational techniques, we demonstrate that replacing the canonical H3 histone with the centromere-specific variant CENP-A translates to greater structural flexibility in the nucleosome, that replacing H3 with CENP-A increases the plasticity of an individual histone dimer, and that the effects of acetylation on the H4 histone tail are cumulative and specific to lysine 16 mono-acetylation.

MOLECULAR DYNAMIC SIMULATIONS OF NUCLEOSOMES AND HISTONE TAILS: THE EFFECTS OF HISTONE VARIANCE AND POST-TRANSLATIONAL MODIFICATION

by

David N. Winogradoff

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Garegin Papoian, Chair/Advisor
Professor David Fushman
Professor Devarajan Thirumalai
Professor John Weeks
Professor Sergei Sukharev, Dean's Representative

To my lovely wife, Jong-In

Acknowledgments

Science does not happen in a vacuum. This thesis was possible because of all the people who surrounded me and my work, the work colleagues, friends, and family who all contributed greatly to this graduate experience.

First and foremost, thank you to my advisor Garyk Papoian for the opportunity to undertake this work and for all the support in completing it. Thank you for all the great conversations, and for everything you taught me about approaching scientific questions. This thesis would not have been possible without your confidence and vision.

Thank you to Davit Potoyan for all your guidance early on with specific details on setting up simulations. Your work in our lab paved the road for the research in this thesis, and I hope current and future members of the lab can build on my work in the years to come. Thank you to Ignacia Echeverria for all your help with the histone tail project. Your expertise and work ethic provide an inspiring example of success. Thank you to Haiqing Zhao for all your contributions to the CENP-A projects and for your creativity.

To Yamini Dalal, thank you for all your guidance and help with the nucleosome project. Your experience and biological insight helped ground this thesis and place it in the context of current scientific debate.

To the previous and current members of the Papoian lab, you have all contributed to this work and made the experience enjoyable. Thanks for all the lunch conversations and activities outside of the lab.

To all my friends at Maryland, whether we bonded over common research interests or playing basketball and soccer, thank you for all the good times and for all the fun. We did a great job of creating a community within the Biophysics and Chemical Physics programs.

Thank you to Michael Coplan for your personal mentorship and encouragement. The Chemical Physics program here at Maryland is truly unique. To Debbie Jenkins, thank you for your patience and organization, and for all the attention to important details navigating my time here.

Thank you to everyone responsible for supercomputers Deepthought (I and II) at the University of Maryland, Kraken at the University of Tennessee, and Stampede at the University of Texas, and the National Science Foundation XSEDE project for access to great computational resources.

To my family, thank you for all your love and encouragement. To my mom, dad, and brother Michael, thank you for always believing in me, for cheering me on and for lifting me up. To my extended family and in-laws, my success is our success.

To my wife Jong-In, I love you baby. This thesis is a reflection of your love and all your little pushes in the right direction. Your interpretations of my figures will always make me smile.

Table of Contents

List of Figures vii

List of Abbreviations ix

Glossary of Terms x

1 Introduction 1
- 1.1 Nucleosome Structure . . . 2
- 1.2 The Nucleosome Core Particle . . . 6
- 1.3 Histone Tails . . . 11
- 1.4 Overview . . . 15

2 Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome 16
- 2.1 Introduction . . . 16
- 2.2 Methods . . . 17
  - 2.2.1 Simulation protocol . . . 17
  - 2.2.2 Analysis of the trajectories . . . 22
- 2.3 Results . . . 24
  - 2.3.1 The CENP-A octamer and nucleosome display greater local fluctuations than their corresponding H3 systems . . . 25
  - 2.3.2 The CENP-A octamer and nucleosome exhibit greater global fluctuations than their H3 counterparts . . . 28
  - 2.3.3 Greater global fluctuations are underpinned by weaker contacts at the CENP-A:CENP-A' dimerization interface . . . 30
  - 2.3.4 Greater plasticity at the CENP-A:CENP-A' dimerization interface is a dominant mode of motion . . . 31
  - 2.3.5 H2A patch mobility is sensitive to surrounding local interactions 34
  - 2.3.6 CENP-A nucleosomal DNA near the pseudo-dyad is relatively unstable compared to the corresponding H3 nucleosomal DNA 36
  - 2.3.7 The CENP-A nucleosome features a more rugged free energy landscape than the canonical H3 nucleosome . . . 42

2.4 Discussion . . . 43
2.5 Summary and Biological Implications . . . 48

3 CENP-A/H4 is more plastic than H3/H4 49
3.1 Introduction . . . 49
3.2 Methods . . . 51
3.2.1 Simulation protocol . . . 51
3.2.2 Analysis of the trajectories . . . 52
3.3 Results and Discussion . . . 54
3.3.1 CENP-A/H4 exhibits greater structural plasticity than H3/H4 55
3.3.2 More regions in CENP-A/H4 display local mobility than in H3/H4 . . . 55
3.3.3 HJURP plays a stabilizing role in CENP-A/H4 . . . 59
3.3.4 H4 adopts more native-like conformations than CENP-A or H3 62
3.4 Conclusion . . . 62
3.5 Summary and Biological Implications . . . 64

4 The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects. 65
4.1 Introduction . . . 65
4.2 Methods . . . 70
4.2.1 Molecular dynamics simulations. . . . 70
4.2.2 Analysis of the trajectories. . . . 73
4.2.3 Clustering Analysis. . . . 74
4.3 Results and Discussion . . . 76
4.3.1 Acetylation of the H4 tail reduces the conformational heterogeneity of the sampled ensemble. . . . 76
4.3.2 Acetylation of K16, but not other lysines, leads to more extended conformations. . . . 79
4.3.3 Acetylation of H4 tails induces increased helical propensities. . 80
4.3.4 Acetylation of H4 tails increases the long-range contact occupancies. . . . 84
4.3.5 Proposed model for the recognition of acetylated H4 tails. . . 87
4.4 Conclusion . . . 90
4.5 Summary and Biological Implications . . . 93

A Chapter 2 Supplementary Information 95

B Chapter 3 Supplementary Information 107

C Chapter 4 Supplementary Information 111

Bibliography 119

List of Figures

1.1 Symmetries and structural motifs of the nucleosome core particle. . . 3
1.2 The CENP-A nucleosome occupies a more rugged free energy landscape than its canonical H3 counterpart. 8
1.3 Location and specific residues of the acidic patch. 10
1.4 Histone tails extend beyond nucleosomal DNA. 12

2.1 The CENP-A octamer and nucleosome structures display greater local flexibility than their canonical H3 counterparts. 26
2.2 The CENP-A nucleosome exhibits greater global flexibility than the canonical H3 nucleosome. 29
2.3 The CENP-A dimerization interface forms fewer and weaker contacts than the corresponding H3 interface. 32
2.4 Increased plasticity of the CENP-A dimerization interface is a dominant mode of motion. 33
2.5 H2A' acidic patch mobility depends on local, electrostatic interactions. 37
2.6 CENP-A nucleosomal DNA is relatively unstable near the pseudo-dyad compared to the corresponding H3 nucleosomal DNA. 41
2.7 The characteristic free energy landscape is more rugged for the CENP-A nucleosome than for the H3 nucleosome. 44

3.1 The CENP-A/H4 dimer exhibits greater structural plasticity than H3/H4. 56
3.2 More regions in CENP-A/H4 display local mobility than in H3/H4. . 57
3.3 HJURP stabilizes the CENP-A/H4 through electrostatic interactions. 60
3.4 H4 adopts conformations closer to the native state than CENP-A or H3. 61

4.1 The H4 N-terminal histone tail sequence and acetylation sites. 71
4.2 Conformational clustering analysis of each of the H4 models. 77
4.3 Heterogeneity of the conformational space of WT and acetylated H4 tails. 79
4.4 Probability distributions of radius of gyration R_g. 81
4.5 Helical propensity per residue. 82

4.6 Hydrogen-bond occupancies of lysines. 85
4.7 Contact maps between different peptide segments. 88
4.8 Ramachandran plots of residues K16 and R17. 89
4.9 Average distance between lysine side-chains. 91

A.1 Amino acid sequences of the H3 and CENP-A systems. 96
A.2 H3 and CENP-A systems reach stable equilibrium after 600 ns. . . . 97
A.3 CENP-A/H4 is more compact, on average, than H3/H4 in the context of the nucleosome and octamer structures. 98
A.4 The CENP-A octamer exhibits greater global flexibility than the H3 octamer 99
A.5 The CENP-A octamer dimerization interface is formed by fewer, and weaker, contacts than the corresponding interface in the H3 octamer. 100
A.6 Greater distortion at the CENP-A dimerization interface, relative to the H3 octamer, is a major mode of motion of the CENP-A octamer. 101
A.7 Asymmetric variability in H2A acidic patch mobility. 102
A.8 Key residue substitutions in CENP-A contribute to DNA instability near the pseudo-dyad. 103
A.9 The asymmetric role of CENP-A H59 in DNA flexibility. 104
A.10 The characteristic free energy landscape is more rugged for the CENP-A octamer than for the H3 octamer. 105
A.11 The effects of DNA length and sequence on CENP-A nucleosome dynamics. 106

B.1 Amino acid sequences of the H3 and CENP-A within dimers. 108
B.2 H3 and CENP-A dimers reach stable equilibrium after 400 ns. 109
B.3 HJURP forms electrostatic interactions with the CENP-A C-terminus. 110

C.1 Convergence of the replica exchange simulations. 112
C.2 Heterogeneity of the conformational ensemble 113
C.3 H4-K16$_{ac}$ helical content by simulation time 114
C.4 Percentage of helical residues. 114
C.5 Ion Association. 115
C.6 The H4-K8$_{ac}$K16$_{ac}$ model. 116
C.7 Hydrogen bonds occupancies. 116
C.8 Stabilizing interactions and salt bridge occupancies. 117
C.9 Ramachandran plots for selected residues. 118

List of Abbreviations

in vivo	"within the living," in living organisms
in vitro	"within the glass," in a laboratory environment
in silico	"within silicon," performed on a computer
AFM	Atomic Force Microscopy
bp	base pair
CENP-A	CENtromere Protein-A
HJURP	Holliday Junction Recognition Protein
IDP, IDR	Intrinsically Disordered Protein, Intrinsically Disordered Region
PTM	Post-Translational Modification
WT	Wild Type
K_{ac}	acetylated lysine, or acetyl-lysine
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuations
R_g	Radius of gyration
COM	Center Of Mass
PCA	Principal Component Analysis
MD	Molecular Dynamics
REMD	Replica-Exchange Molecular Dynamics
Å	Ångstrom (10^{-10} m)
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
TIP3P	Three-point explicit water model
IPST	Institute for Physical Science and Technology

Glossary of Terms

1. **Root-mean-square deviation (RMSD)** is a measure of difference between two structures after alignment:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}, \tag{1}$$

where δ is the distance between N pairs of equivalent atoms (Cα atoms, for example).

2. **Root-mean-square fluctuation (RMSF)** is a measure of the deviation, over time, between a particle (i) and some reference position:

$$\sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(x_i(t_j) - \langle x_i \rangle\right)^2}, \tag{2}$$

where T is the total time (i.e. total number of simulation snapshots considered) and $\langle x_i \rangle$ is the time-averaged position of that particle, serving as a reference point.

3. **Center-of-mass (COM)** is the geometric average location of a distribution of mass in space. For a system of particles P_i, $i = 1, 2, ..., n$, with masses m_i, summing to M, located in space with coordinates r_i,

$$COM = \frac{1}{M}\sum_{i=1}^{n} m_i r_i. \tag{3}$$

4. **Principal Component Analysis (PCA)** is a statistical method that transforms a set of observations (e.g. Cα positions) of possibly correlated variables into a set of orthogonal, uncorrelated variables known as principal components. The first principal component accounts for the greatest possible variance, the second principal component accounts for the second-most, and so-forth.

5. **Chromatin** is the DNA-protein macromolecular complex within eukaryotic nuclei, a dynamic array of nucleosomes forming the first level of organization.

6. **Nucleosomes** are the fundamental structural units of chromatin. The nucleosome core particle includes ∼1.7 super-helical turns of DNA surrounding a protein octamer core, containing two copies of the four core histones (H2A, H2B, H3 and H4).

7. **Pseudo-dyad** (or dyad axis) of the nucleosome core particle is a plane of rotational symmetry, dividing the nucleosomal DNA and histone octamer in two (each containing one copy of H3, H4, H2A, and H2B proteins). This plane is perpendicular to the superhelical axis of DNA, and pseudo-dyad can also refer to the central base pair of nucleosomal DNA.

8. **Acidic patch** is a cluster of eight acidic residues (H2A E56, E61, E64, D90, E91, E92 and H2B E102, E110) that forms a pocket at the surface of the nucleosome.

9. **Histones** are positively charged globular proteins found in the nuclei of eukaryotic cells, which order and package DNA into nucleosomes, the fundamental structural units of chromatin.

10. **CENP-A** CENtromere Protein-A is a histone variant of H3 found at the centromere, defining the location to which microtubules bind during mitosis.

11. **Lysine acetylation** is a common chemical modification, whereby an acetyl group is transferred to the epsilon-amino group, neutralizing the lysine residue's positive charge. Histone acetylation is known to play an important role in regulating gene expression.

Chapter 1: Introduction

The packaging of genomic information and the regulation of gene expression are both fundamentally important to eukaryotic life. Meters of human DNA must fit inside the micron-diameter nucleus while still rapidly becoming available for templated processes such as transcription, replication, and repair. Therefore, the DNA-protein complex known as chromatin must dynamically transition between more compact, closed states and more accessible, open ones. To fully understand chromatin structure and dynamics, it is necessary to employ a multifaceted approach, integrating different general philosophies and scientific techniques that include experiment and computation. Since the DNA in chromatin is organized into arrays of nucleosomes, we take a bottom-up approach in this dissertation, striving first to understand the structure and dynamics of an individual nucleosome and subdomains thereof. Atomistic computational methods have provided useful tools to study DNA and protein dynamics at the nanosecond, and recently microsecond, timescale. In this dissertation, we present recent developments in the understanding of the nucleosome though atomistic molecular dynamics (MD) simulations. By applying different computational all-atom MD techniques, we investigate the effects of histone variance and post-translational modification on the overall thermodynamics and specific

mechanistic details of nucleosome dynamics. In the remainder of this section, we will discuss the motivation behind our work and provide a review of previous and current all-atom MD computational studies of the nucleosome, including the results of this thesis.

1.1 Nucleosome Structure

The nucleosome is the fundamental packaging unit of DNA within the nucleus and, therefore, serves as the key determinant of DNA accessibility. The nucleosome structure was originally proposed in the 1970s [2, 3], then determined to high-resolution by X-ray diffraction near the turn of the century (PDB ID: 1AOI [1]), and further refined to the near atomic resolution (1.9 Å) model used today (PDB ID: 1KX5 [4]). Experimental techniques continue to develop in order to investigate chromatin folding and individual nucleosomes, including single-molecule techniques such as atomic force microscopy (AFM).

The structure of the canonical nucleosome is approximately 1.7 left-handed super-helical turns of double-stranded DNA wrapped around a protein octamer core (Figure 1.1.A,B,C). The protein octamer contains two copies of each of the four core histones – H2A, H2B, H3, H4 – divided into four dimers. Together, the dimers form a positively charged helical ramp, around which the negatively charged DNA wraps. More specifically, the octameric histone core consists of four heterodimers, two of H2A/H2B and two of H3/H4. The two H3/H4 dimers form a tetramer, primarily interacting through a four-helix bundle, two helices from H3 and two from H3',

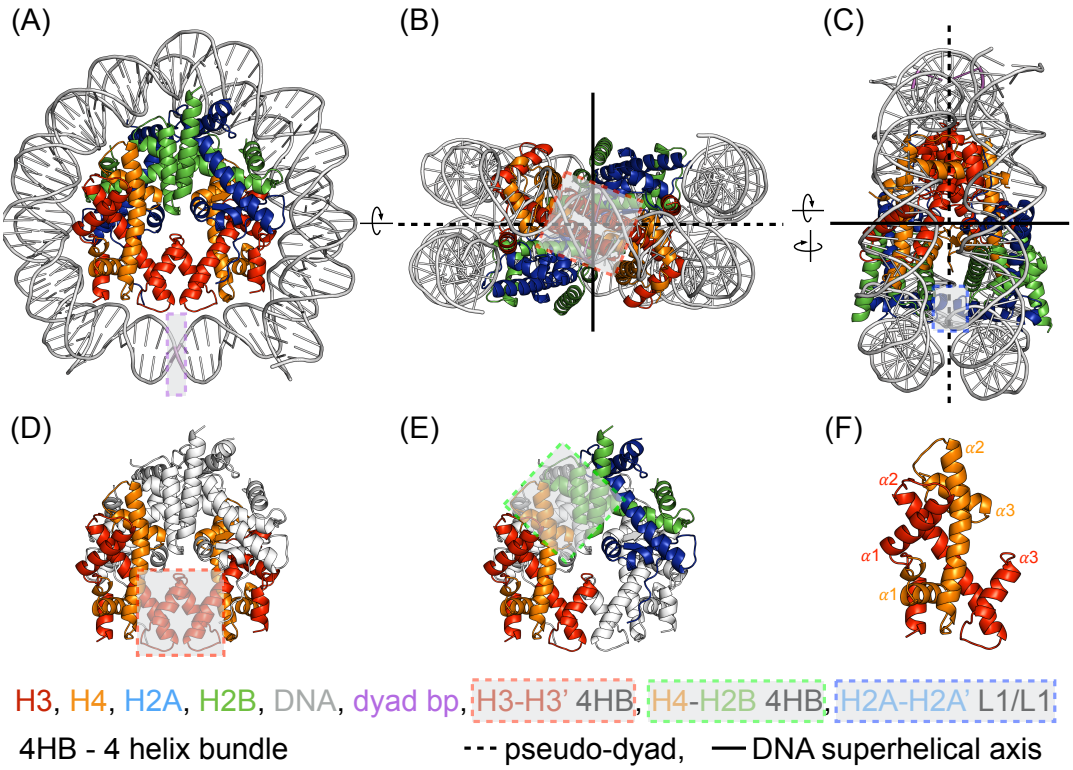


Figure 1.1: **Symmetries and structural motifs of the nucleosome core particle.** The nucleosome core particle (PDB ID: 1KX5 [1]), without tails shown, aligning the DNA superhelical axis (A) into the page, (B) vertically, and (C) horizontally. The pseudo-dyad is a plane defined by the black, dashed line, and going into the page from viewpoints (B) and (C). The nucleosome is rotational symmetric about the pseudo-dyad. The purple shaded box highlights the base pair located at the dyad, the central base pair in the DNA sequence. Most evident from viewpoint (B), the histone octamer core forms a ramp matching the pitch of the superhelically-wound DNA. (D) Two H3/H4 dimers associate into a tetramer, interacting through an H3-H3' four-helix bundle, denoted by a red shaded box. (E) The $(H3/H4)_2$ tetramer is flanked on both sides by H2A/H2B dimers, binding through H4-H2B four-helix bundle interactions, shown as a green shaded box. This means that each H3/H4 dimer is more stably associated within the core, bracketed on each side by a four-helix bundle, than the H2A/H2B dimers, which each only form a four-helix bundle on one side. (F) An individual H3/H4 dimer, enlarged to identify structural features. Each core histone has three central helices ($\alpha1$, $\alpha2$, $\alpha3$) connected by two loops, giving the protein a broad U-shape. H3 and H4 proteins slot together to form a pair, creating a stable handshake motif.

close to the central base pair of DNA at the pseudo-dyad (Figure 1.1.D). Similar four-helix bundles, between H2B and H4, define the interactions between the H3/H4 tetramer and the two H2A/H2B dimers [1] (Figure 1.1.A,E). Following the standard nomenclature convention, we use a prime superscript, or absence thereof (e.g. H3' vs H3), to distinguish between the two identical heterotypic halves, rotationally symmetric about the pseudo-dyad (Figure 1.1.B), each containing one copy of histones H3, H4, H2A and H2B. Among the most evolutionarily conserved proteins in eukaryotes, the core domain of each histone follows the same structural motif: three long α-helices connected by two loops (Figure 1.1.F). Assembly of the nucleosome occurs in discrete steps *in vitro* depending on the ion concentration [5], suggesting the possibility of intermediate dimer, tetramer, or hexamer structures *in vivo*.

Even without considering intermediate subsets of the entire assembly, the nucleosome is a structural ensemble because of the variability in DNA sequence and chemical modifications to the DNA and histones. DNA methylation and histone post-translational modifications influence the structure and dynamics of chromatin, playing an important role in the regulation of gene expression *in vivo*. Furthermore, every core histone except H4 has its own repertoire of variants, containing different amino acid sequences, which can replace the conventional and most-common canonical histones.

The family of published X-ray nucleosome structures recently expanded to include nucleosomes containing CENP-A (CENtormere Protein-A) [6]. CENP-A is an essential H3 histone variant specific to eukaryotic centromeres, which dictate the single location per chromosome to which microtubules bind during mitosis. How

CENP-A containing nucleosomes contribute structurally to chromosome function is a fundamental question to chromosome biology. The CENP-A nucleosome occupies a range of structures *in vivo* [6–23], and experimental studies continue to debate how the CENP-A nucleosome differs from the canonical H3 nucleosome, with respect to rigidity [15,24], height [25], and distortability. CENP-A is the most-distant known relative of the canonical H3, sharing only ∼60% amino acid sequence similarity, yet the structural alignment of CENP-A and H3 proteins demonstrates that they are nearly identical, aside from the 2-residue longer loop 1 region of CENP-A (Figure 1.2.A). Overall, the global description of the centromeric octameric nucleosome matches that of the canonical nucleosome [6], with CENP-A (and CENP-A') replacing the canonical histone H3 (and H3').

The standard high-resolution nucleosome X-ray crystal structures containing the canonical H3 (PDB ID 1KX5 [4]) and the centromere-specific H3 variant CENP-A (PDB ID 3AN2 [6]), both include palindromic α-satellite DNA – highly repetitive, non-transcribed sequences rich in A-T base pairs and associated with centromeric regions – wrapped around histone octamers. However, 147 base pairs are visible in the crystal structure of the canonical H3 nucleosome, whereas only the central 121 base pairs are visible in the centromeric nucleosome containing CENP-A when starting from a similar 147 base pair sequence [6]. Aside from the difference in DNA length, the octameric CENP-A and H3 nucleosome crystal structures are superimposable [6], and they serve as the starting points for our computational modeling of the entire nucleosome.

The dynamics of nucleosomal DNA depend heavily on interactions with the

histone core. Critically situated arginine side chains play a general (i.e. non-specific to DNA sequence) stabilizing role in the nucleosome, similar to the spokes of a wheel. There are 14 regularly spaced major points of contact between protein and DNA, occurring once every 10.5 bp, first observed in the original high-resolution crystal structure of the nucleosome [1].

For a complete picture of the nucleosome, we must accompany the structural description with its dynamic characteristics. Through computational biophysical methods, we aim to guide and complement experiments by making predictions about the nucleosome and to provide physical insight by modeling details not easily accessible to experimental techniques, including regions of intrinsic disorder. Here we present the contributions of different molecular dynamics studies to our overall understanding of the dynamic characteristics of the nucleosome and subdomains thereof.

1.2 The Nucleosome Core Particle

The nucleosome core particle includes the histone octamer core and surrounding super-helical DNA. Each histone contains a more structured, globular core as well as intrinsically disordered tails, which extend beyond the surrounding DNA. We will separately consider the contributions of atomistic computational studies to our overall understanding of (1) the nucleosome core particle, and (2) histone terminal tails.

Canonical nucleosomal DNA is approximately one persistence length under

physiological conditions (50 nm), yet it wraps around the octameric histone core in ∼1.7 left-handed superhelical turns, thus adopting a highly bent conformation. Previous all-atom MD studies investigated the local and global dynamics of nucleosomal DNA by examining the helical parameters, including α-parameter analysis [26], and by characterizing overall DNA modes of motion [27]. The super-helical DNA exhibits greater structural flexibility than the histone octamer core, while still remaining surprisingly stable [27]. Variations in local DNA dynamics, and of the α-parameter specifically, are periodic with peaks correlated to the sites of strong DNA-histone contacts [26]. Through principal component analysis, all-atom MD simulations on the 20 nanosecond timescale identified breathing-type fluctuations of the entry and exit super-helical DNA as a global mode of motion [27]. Indeed, several MD simulations noted increased flexibility at the entry and exit DNA [27–29].

A recent atomistic study on the 100 nanosecond timescale demonstrated that entry and exit DNA regions are more flexible in the CENP-A nucleosome than in its canonical H3 counterpart due to the substitution of two specific arginines in H3 with lysines in CENP-A [29]. On the microsecond timescale, we observed asymmetric unwinding of the entry and exit DNA of the CENP-A and canonical H3 nucleosomes. Furthermore, CENP-A nucleosomal DNA displays asymmetric instability near the pseudo-dyad relative to the corresponding H3 nucleosomal DNA. DNA traces a continuous path *in vivo* with linker DNA sections bridging the gaps between nucleosomes in an array. Therefore, nucleosomal endpoint DNA flexibility could be important for higher-order chromatin structure. In addition to the surrounding DNA, interactions within the histone core, and between histones and

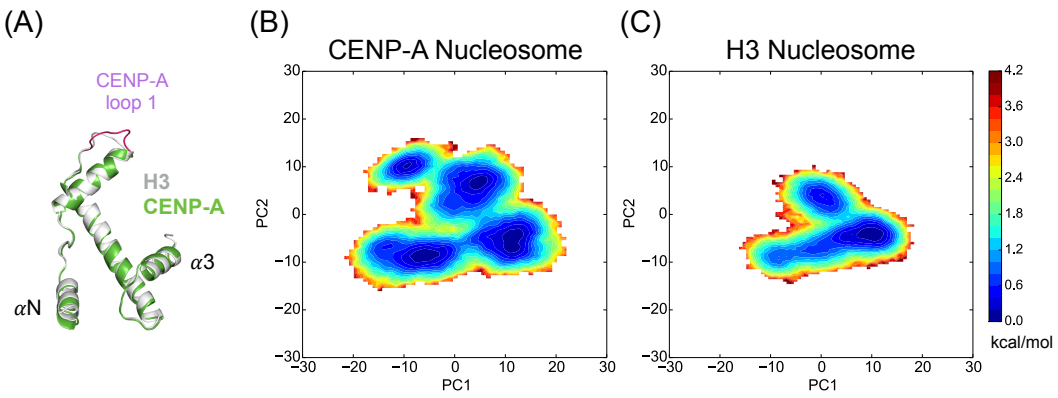


Figure 1.2: **The CENP-A nucleosome occupies a more rugged free energy landscape than its canonical H3 counterpart.** (A) Structural alignment of histone proteins CENP-A and H3 reveals remarkable similarity, except for the longer loop 1 region of CENP-A. Projections of free energy from all-atom MD simulations along the first two principal components of (B) the CENP-A nucleosome, and (C) the canonical H3 nucleosome. The CENP-A nucleosome explores greater configurational space and more distinct conformational basins than the H3 nucleosome.

DNA, contribute to the overall nucleosomal dynamics.

All-atom MD simulations observed structural deviations of the histone core domains to be relatively small in comparison to nucleosomal DNA. However, atomistic studies revealed that the core domain of H2A displays greater mobility than the core domains of other histones [27, 30], specifically identifying $\alpha 2$ and $\alpha 3$ of H2A [30]. In the absence of histone tails, a specific residue in H2A $\alpha 3$ (Arg88) interacts with a specific residue of H3 $\alpha 2$ (Glu105), increasing the overall structural fluctuations of H2A [30]. Using principal component analysis and free energy landscape theory on the microsecond timescale, we found that the histone core of the CENP-A nucleosome is more flexible than that of the H3 nucleosome, underpinned by greater plasticity at the CENP-A:CENP-A' interface than at H3:H3'. Overall, the CENP-A nucleosome adopts a more rugged free energy landscape than the H3 nucleosome (Figure 1.2.B,C). Our results indicate that the CENP-A induces greater flexibility in the octameric nucleosome, suggesting that the CENP-A nucleosome is less rigid and more distortable than the canonical H3 nucleosome. This may lead to the adoption of alternate conformations under biological force.

The surrounding solvent and electrostatics of the histone core are important for interactions within and between nucleosomes. A previous 200 nanosecond long MD simulation revealed that explicit waters and ions permeate the space between histones within the nucleosome core, creating a channel coinciding with the axis of super-helical DNA [28]. The electrostatic surface of the histone core is mainly positive (Figure 1.3.A), offsetting the negative DNA, except for a highly negatively charged region near the center of the protein core surface exposed to solvent, known

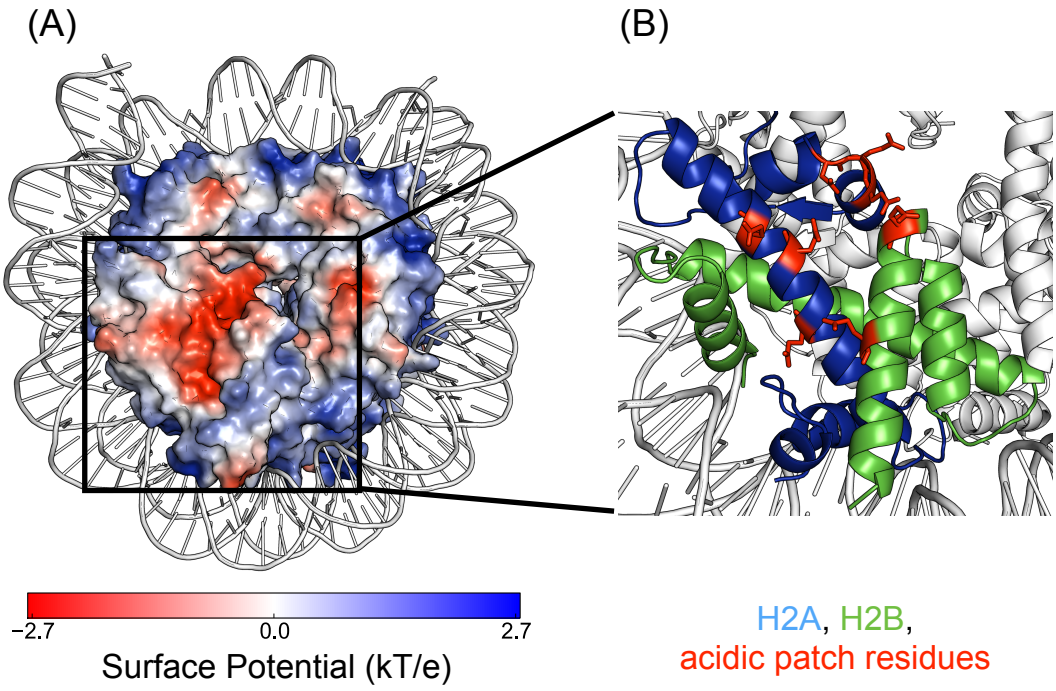


Figure 1.3: **Location and specific residues of the acidic patch.** (A) The electrostatic surface potential of the canonical histone core (PDB ID: 1KX5), rendered in Pymol. The acidic patch is a region of highly negative charge density exposed to the surrounding solvent. (B) Specific, negatively-charged residues of histones H2A (E56, E61, E64, D90, E91, E92) and H2B (E102, E110) define the acidic patch, shown as sticks.

as the H2A/H2B acidic patch (Figure 1.3.B). Materese et al. [28] provided a theoretical model explaining how nucleosomal DNA is more neutralized than DNA free in solution and observed a high condensation of positive ions near the acidic patch, previously identified by experiment to bind a highly positively charged region of the H4 tail [1]. We continue to investigate the important regulatory role that histone tails play in chromatin dynamics, and the underlying physical principles governing their function require further explanation.

1.3 Histone Tails

In addition to their globular cores, each histone extends beyond the surrounding DNA with N- and C-terminal tails, protruding outwards from the nucleosomal surface (Figure 1.4). Histone tails are intrinsically disordered regions (IDRs) that play a key regulatory role in chromatin higher-order structure and dynamics directly by interacting within and between nucleosomes and indirectly by serving as recognition sites for chromatin remodelers. These tails also feature multiple possible sites for post-translational modification (PTM), providing another layer of regulation. Lysine acetylation is one of the most common PTMs, whereby an acetyl group is transferred to the epsilon-amino group, neutralizing the residue's positive charge. Atomistic simulations have studied histone tails within the context of an entire nucleosome and isolated in solution. These studies propose that histone tails play an active and essential regulatory role through structural flexibility, conformational heterogeneity, and PTM diversity.

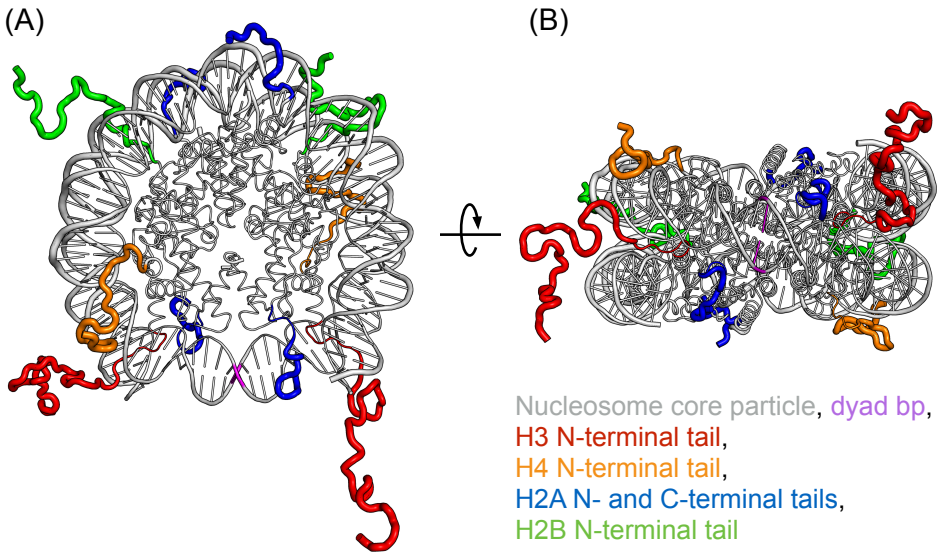


Figure 1.4: **Histone tails extend beyond nucleosomal DNA.** Histone N- and C-terminal tails highlighted in different colors, displayed within the context of the nucleosome core particle (PDB ID: 1KX5 [4]) viewed by aligning the DNA superhelical axis (A) into the page and (B) vertically. Every core histone features an N-terminal tail, and H2A also includes a C-terminal tail. Furthermore, the histone tails vary in length, location with respect to the DNA sequence, and whether they extend around or through the gyres of the surrounding DNA. Histones and DNA are represented by tubes, whose width corresponds to the B-factor determined by experiment. B-factor is directly proportional to mean-squared displacement of each atom, or groups of atoms in residues and base pairs in this case, and therefore represents a coarse measure of local mobility. The B-factors from the crystal structure illustrate that the surrounding DNA is, on average, more mobile than the histone core; that the DNA local mobility depends both on the location in the sequence and proximity to the protein core; and lastly, that the histone terminal tails are much more mobile and structurally flexible than the core domains.

Histone tails were originally thought to act simply as unstructured electrostatic mediators, and PTMs were thought to function solely through modulating these interactions. More recent studies challenge this assertion, demonstrating the essential role of histone tails in higher-order chromatin structure through *in vitro* experiment, and revealing the propensity of histone tails to form secondary structure in computational studies. Early *in vitro* experiments determined that histone tails are essential for the formation of the 30-nm fiber [31–33], and another landmark *in vitro* study concluded that the mono-acetylation of the H4 histone tail at lysine 16 alone was sufficient to inhibit the formation of higher-order chromatin structures [34]. A region of high positive charge in the H4 tail including lysine 16 must bind to the H2A acidic patch of another nucleosome to ensure the proper nucleosome stacking in condensed states [1, 35–37]. The mono-acetylation of lysine 16 clearly disrupts the electrostatic aspect of these interactions by reducing the H4 tail's positive charge [38]. However, the effects of histone tails and lysine acetylation extend beyond electrostatics, and recent studies demonstrate that histone tails contain elements of structural order, and, furthermore, that acetylation could induce structural transitions.

Several recent all-atom MD studies investigate the propensity of histone tails to form secondary structure. One study generated conformational ensembles of each of the four histone tails isolated in explicit solvent by performing all-atom replica exchange molecular dynamics (REMD). This revealed transient β-hairpin formation in the H4 tail and α-helical elements in H3 and H2B tails while the H2A tail remained fully disordered [39]. This study classified the H4, H3, and H2B tails

as molten globular, between ordered and disordered states [39]. Within the context of the entire nucleosome, H3 and H2B tails also displayed the propensity to form α-helices on the 100 nanosecond timescale [30]. This study also outlined an allosteric mechanism whereby specific arginines in H2A mediate nucleosome stability upon H3 tail truncation [30].

Because experimental studies have specifically identified the H4 tail as important, atomistic studies have focused on this tail and on the structural effects of acetylation. In 20 nanosecond long MD simulations, Lins and Rothlisberger demonstrated that the H4 tail helical content increases upon tetra-acetylation, forming helices near the C-terminus [40]. Using replica exchange MD to enhance the ensemble sampled, totaling 3 microseconds for each system, Potoyan and Papoian found that the un-acetylated H4 tail contains a transient β-hairpin [39] and, furthermore, that the mono-acetylation of lysine 16 induces a partial ordering of the H4 tail, leading to the formation of elements of helical secondary structure and increasing the binding affinity to DNA [41]. Korolev et al. used all-atom MD to model fragments of the H4 tail in the presence of DNA, demonstrating that a highly charged region, residues 16-23, adopts a stretched conformation when unacetylated [42], similar to the results of Potoyan and Papoian.

Our work continued this thread by exploring how different levels of acetylation modulate the conformational preferences of the H4 tail isolated in explicit solvent through REMD [43], totaling 6 microseconds of simulation time for each system. We demonstrated that progressive acetylation largely has a cumulative effect, decreasing conformational heterogeneity and increasing helical propensity, and that lysine 16

mono-acetylation has a specific effect, rigidifying the tail near lysine 16 and leading to a specific helix and extended conformations [43]. Furthermore, we found that increased acetylation results in spatially clustered acetyl-lysines, which could serve as recognition patches for chromatin regulating proteins [43], such as bromodomains.

1.4 Overview

This dissertation investigates the effects of histone variance and post-translational modification on the structure and dynamics of the entire nucleosome and subdomains thereof using different molecular dynamics techniques. We order the results chapters from more general to more specific, focusing first on the entire nucleosome, then on an individual dimer, and finally, on a specific histone tail. In Chapter 2, we use all-atom MD to examine the effects of replacing the canonical H3 histone with the centromere-specific variant CENP-A on the structure and dynamics of the octameric nucleosome through principal component analysis and free energy landscape theory. In Chapter 3, we employ all-atom MD simulations to determine the effects of replacing H3 with CENP-A in the structure and dynamics of an individual heterodimer with H4, and to examine the role of HJURP in stabilizing the CENP-A/H4 dimer. In Chapter 4, we use replica-exchange molecular dynamics (REMD) to investigate how increasing the level of acetylation modifies the conformational preferences of the H4 histone tail, distinguishing between the cumulative effects of progressive acetylation and the specific effects of lysine 16 mono-acetylation.

Chapter 2: Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome

2.1 Introduction

Histone variants are key players in the epigenetic process, encoding identity to specialized regions of the genome. CENP-A/CENH3 is a centromere specific histone H3 variant present in all eukaryotes, whose role is to dictate the single location per chromosome to which microtubules bind every mitosis. CENP-A is thought to specify not just the location, but also structural identity to centromeres. CENP-A nucleosomes are currently shown to occupy a range of structures *in vivo* [6–23]. Furthermore, even *in vitro*, crystallographic and biophysical analyses conflict, with data suggesting that CENP-A octamers are more rigid than H3 octamers [15, 24]; that CENP-A octameric nucleosomes are more compact [25]; that CENP-A nucleosomes are more unstable than H3 nucleosomes [44]; that Drosophila CENP-A can be assembled into hemisomes [20]; that yeast CENP-A nucleosomes possess a more elongated "open" conformation [16]; and finally, that human CENP-A and H3 octameric nucleosomes are essentially indistinguishable [6, 45, 46]. These divergent experimental observations made us curious whether subtle contributions from

internal regions of CENP-A might translate to changes in the overall dynamics of the CENP-A nucleosome. To explore this possibility, we modeled CENP-A and H3 nucleosomes, as well as their respective DNA-free protein octamers, using all-atom molecular dynamics (MD) in explicit solvent.

To our surprise, despite the fact that H3 and CENP-A octameric nucleosomal crystal structures are superimposable, within 2 Å [6], and appear almost identical by AFM [46], computational modeling reveals distinct differences in their dynamics. CENP-A nucleosomes and octamers demonstrate greater local and global structural fluctuations than their canonical H3 counterparts. The CENP-A nucleosome reveals intrinsic local flexibility at the loop 1 region of CENP-A and increased plasticity of the CENP-A:CENP-A' dimerization interface relative to H3:H3'. Furthermore, the CENP-A nucleosomal DNA near the pseudo-dyad is more unstable than the corresponding H3 nucleosomal DNA. Finally, our simulations demonstrate that the CENP-A nucleosome explores more conformational space than the H3 nucleosome. Taken together, our data support the possibility that pliability is an intrinsic feature of CENP-A nucleosomes, which may have implications for their structure and function *in vivo*.

2.2 Methods

2.2.1 Simulation protocol

We performed all-atom molecular dynamics (MD) using the gromacs 4.5.7 MD software [47], the amber99SB*-ILDN [48, 49] force field for proteins, the parmbsc0

[50] force field for DNA, the ions94 [51] force field for ions, and the TIP3P water model. Starting from crystal structures for the canonical H3 nucleosome (PDB ID: 1KX5 [4]) and the CENP-A nucleosome (PDB ID: 3AN2 [6]), we built models for four systems: (1) the canonical H3 nucleosome; (2) the CENP-A nucleosome; (3) the canonical H3 histone protein octamer; and (4) the CENP-A histone protein octamer.

The following modifications were made to the experimentally determined crystal structures for a fair comparison. Canonical histone lengths were redefined to match the tailless histones found in the CENP-A crystal structure (3AN2). The missing section for CENP-A' chain E, Thr 79 to Asp 83, was generated with MODELLER, using the corresponding region in CENP-A chain A as a homologous structure. Lastly, non-standard Mse residues found in the crystal structures were replaced with Met (a single atom substitution, Se to S).

We used the *pdb2gmx* tool in gromacs to set the Lys and Arg residues to +1e, the Asp and Glu residues to -1e, the Gln residues to neutral, and to protonate the His residues solely at NE2. Each system was solvated in a rectangular water box, ensuring a minimum buffer length of 15 Å between the system and the edges of the box. We introduced Na^+ and Cl^- ions to neutralize the charge and represent the physiological 0.150 M NaCl environment. The systems were minimized using steepest descent, until reaching a maximum force less than 100 kJ/mol. Periodic boundary conditions were employed throughout all the simulations, and long-range electrostatics were treated with the Particle Mesh Ewald method [52]. Non-bonded Coulomb and Lennard-Jones interactions were truncated at 10 Å, and all bonds involving

hydrogen were constrained using the LINCS [53] algorithm. After minimization, the systems were heated to 300 K by 500 ps of protein and DNA restrained NVT MD simulation followed by 500 ps of NVT MD simulation with weak harmonic restraints on protein and DNA atoms (K = 2.5×10^{-1} kJ/(mol nm^2)). Weak restraints were used throughout the simulations after the initial protein and DNA restrained NVT simulation in order to prevent large-scale translation and rotation because of the non-cubic water boxes. After reaching thermal equilibrium, the systems were equilibrated at 300 K and 1.0 bar for 1.5 ns in the NPT ensemble.

To characterize the structure and dynamics of the canonical and CENP-A nucleosomes and octamer cores, we performed production all-atom MD simulations in the NPT ensemble at 1.0 bar and 300 K with a 2 fs time-step, saving coordinates, velocities, and energies every 2 ps for further analysis. We updated the list of non-bonded neighbors every 10 steps. Using the V-rescaled, modified Berendsen thermostat [54] with a 1.0 ps time-constant and the Parrinello-Rahman barostat [55] with a relaxation time of 2.0 ps, we performed one microsecond of MD simulations, only considering the final 400 ns for analysis, with weak position restraints on heavy atoms (K = 2.5×10^{-1} kJ/(mol nm^2)) for each system. These restraints do not interfere with common internal motions, both local and collective, but could repress unusually large-scale displacements or major structural disruptions. Hence, we ran an additional 400 ns of MD simulations with significantly reduced position restraints (K = 5.0×10^{-5} kJ/(mol nm^2)), which do not interfere with internal collective dynamics at any length scale. Performing the same analysis for the two above-mentioned restraint values, CENP-A and canonical H3 structures were found

to undergo the same characteristic local and global dynamics except for the H2A' acidic patch. Therefore, we separately addressed the sensitivity of H2A acidic patch mobility to the surrounding local interactions.

It is important to consider the choices made in designing our models, and future directions moving forward with all-atom MD simulations. Here, we do not include the tail domains in our models because they are not resolved in the octameric CENP-A nucleosome (PDB ID: 3AN2 [6]), and we want to make the comparison between the CENP-A and canonical H3 structures as fair as possible. The intrinsically disordered tail regions will have important effects, however, these effects will depend heavily on the initially determined conformations. From a simulation standpoint, each tail adopts an ensemble of conformations, and 1 μs of constant T all-atom MD simulations in explicit solvent would not provide sufficient sampling to reach equilibrium. The most recent canonical H3 nucleosome crystal structure (PDB ID: 1KX5 [4]) used special chemical additives to stabilize the tail regions enough for x-ray diffraction. Different computational techniques, such as replica exchange molecular dynamics, are often used to investigate histone tail regions in order to enhance sampling, overcoming the many barriers in the rugged free energy landscapes such tails occupy [39, 41–43]. Examining the specific role of histone tails within an individual nucleosome is an important future direction, upon the publication of an experimentally resolved octameric CENP-A nucleosome crystal structure that includes histone tail domains.

The possible effects of DNA length and sequence on nucleosome dynamics are important to consider as well. For the results shown in the main text, we only

consider the central 121 base pairs of DNA for the CENP-A nucleosome, since only these base pairs, from a 147 base-pair long sequence, were visible in the CENP-A nucleosome crystal structure (PDB ID: 3AN2 [6]). The arrangement of these additional base pairs of CENP-A nucleosomal DNA is not known for sure, i.e. it is not clear that they follow the same superhelical path observed experimentally for the canonical H3 nucleosome. Indeed, Tachiwana et al. [6] hypothesize that the terminal regions of DNA could be more flexible than the corresponding regions in the canonical H3 nucleosome due to structural differences between CENP-A and H3.

We performed another simulation transplanting the experimentally resolved 147 base pairs of canonical H3 nucleosomal DNA (from PDB ID: 1KX5 [4]) onto the CENP-A nucleosome. In the Supplementary Information, we compare CENP-A nucleosome dynamics with both the 121 bp from 3AN2 [6] ("CENP-A^{121}") and with the 147 bp from 1KX5 [4] ("CENP-A^{147}") to the canonical H3 nucleosome [4] ("$H3^{147}$"), considering the CENP-A dimerization interface (Supplementary Figure A.11.A), the distances between histone dimers (Supplementary Figure A.11.B), and nucleosomal DNA flexibility (Supplementary Figure A.11.C). These data show a potential increase in the stability of interactions between CENP-A/H4 and H2A/H2B in the first heterotypic half and in the stability of CENP-A nucleosomal DNA near the pseudo-dyad when 13 bp are added to entry and exit regions of DNA (Supplementary Figure A.11.B,C). However, the interactions of the internal CENP-A/H4 homotypic core remained similar (Supplementary Figure A.11.A,B).

2.2.2 Analysis of the trajectories

We first determined the root-mean-square fluctuations (RMSF) of every Cα atom for all of the studied systems. Cα RMSF serves as a measure of the local structural fluctuations, as compared to the geometric average structure, identifying specific residues, and protein regions, that exhibit greater flexibility. We compared the distances between the centers-of-mass (COM) of the H3 (and CENP-A) α1 helix and H4 α2 helix to provide insight into the structural fluctuations within one protein dimer. Furthermore, we examined the global structure and dynamics of the canonical and CENP-A systems by comparing the distances between dimers within histone tetramers. Four dimers provided coarse-grained definitions for the protein component of the canonical, H3 systems (H3/H4, H2A/H2B, H3'/H4', H2A'/H2B') and of the CENP-A systems (CENP-A/H4, H2A/H2B, CENP-A'/H4', H2A'/H2B'). We analyzed inter-residue contact preferences at the interfaces of H3:H3' and CENP-A:CENP-A'. A contact was determined to exist when the distance between two non-hydrogen atoms from different residues was less than 3.6 Å. Contacts were calculated as fractions of time of their respective entire trajectories. This contact definition will include hydrogen bonds, electrostatic interactions, and hydrophobic interactions between residues, without distinguishing the specific type of contact. In general, the length scales defining electrostatic and hydrophobic interactions are longer than those defining hydrogen bonds. For example, distances between oppositely charged heavy-atoms from different residues under 4.0 Å is a common salt-bridge definition [56]. We investigated DNA structural fluctuations by calculating basepair RMSF

for the CENP-A and H3 nucleosomal DNA, and the associated standard deviations were determined by the contributions of each third of the trajectory to the structural variation in each basepair.

Principal component analysis (PCA) was performed to extract the dominant modes of motion of the nucleosomal and octameric structures from the MD simulation trajectories [57]. Overall translational and rotational motion in the trajectories were eliminated by a translation to the average geometric center and by alignment to the energy-minimized structure. Using the Cartesian coordinates of all the Cα atoms (N = the number of Cα atoms), we generate a $3N \times 3N$ covariance matrix. The diagonalization of this matrix provides a set of eigenvectors that give a vectorial description of each component of motion. Every eigenvector has a corresponding eigenvalue that represents the contribution of that component of motion to the total variance of the data set. For visualization, we projected the top two principal components, times the square roots of their corresponding eigenvalues (since the eigenvalues are variances in Å^2) multiplied by an array of unitless scalars between -5 and +5 (to facilitate easier observation), onto the corresponding representative structure, saving modified structures and compiling them into a movie. The representative structure for each system was defined by the simulation snapshot with the lowest RMSD with respect to the geometric average of the entire analyzed trajectory. Lastly, we projected the CENP-A and H3 trajectories onto the first two principal components, calculated by Cα PCA, to reveal the conformational space explored by each system studied.

2.3 Results

Here, we explore the effects of an important histone variant, CENP-A, on the dynamics of the nucleosome. Using all-atom molecular dynamics (MD) simulations in explicit solvent on a microsecond timescale, requiring over 400,000 cpu hours in total, we investigated the structural fluctuations and dominant modes of motion of the CENP-A nucleosome in comparison to the canonical H3 nucleosome, based on the crystal structures for the octameric nucleosomes containing CENP-A (PDB ID 3AN2 [6]) and H3 (PDB ID 1KX5 [4]). Four systems were studied in total: (1) the canonical H3 nucleosome; (2) the octameric CENP-A nucleosome; (3) the canonical H3 octamer; and (4) the CENP-A octamer. Two copies of each histone exist within the octameric nucleosome protein cores. Following the standard nomenclature, we use a prime superscript, or absence thereof (e.g. CENP-A' vs CENP-A), to distinguish between the two identical heterotypic halves, rotationally symmetric about the pseudo-dyad (Fig. 2.1.A), each containing one copy of histones H3 (or CENP-A), H4, H2A and H2B.

Our analysis reveals dynamics that deviate significantly from the nearly identical crystal structures of the CENP-A and H3 nucleosomes (Fig. 2.1.A), which are due, in part, to the significant dissimilarities in the amino acid sequences of the CENP-A and H3 proteins (Supplementary Figure A.1). Sequence and structural alignment highlights the longer loop 1 region in CENP-A compared to H3 (Supplementary Figure A.1 and Fig. 2.1.B). We observed that it took several hundred nanoseconds for all systems to reach equilibrium. Thus, in a technical advance, we

ran our simulations for a full microsecond each, in order to harvest the stable final 400 ns of each trajectory for subsequent analyses (Supplementary Figure A.2).

2.3.1 The CENP-A octamer and nucleosome display greater local fluctuations than their corresponding H3 systems

We identified local structural variations within each system by computing root-mean-squared fluctuations (RMSF) of C-alphas within each system compared to their respective trajectory averages. To determine a threshold for significant differences between the local fluctuations of different systems, we calculated the RMSF of each third of the trajectories separately and obtained the standard deviation for each fluctuation. The maximum standard deviation in RMSF is $\sim$0.3 Å, meaning any difference greater than 0.6 Å is very significant. The CENP-A loop 1 region, a region spanning the H2A' acidic patch, and H4 αN helix of the CENP-A octamer and nucleosome exhibit greater structural variability than the corresponding regions for the H3 octamer and nucleosome, respectively (Fig. 2.1.C,D).

The CENP-A loop 1 region (residues 78-84) is two residues longer than the same region of H3 and predicted to display relatively greater structural variation [6]. This region is important because of its exposure to solvent and, structurally, because it provides the connection between two major helices in H3 and CENP-A, α1 and α2. The loop 1 region is the only significance difference in local flexibility, at the individual residue level, between H3 and CENP-A proteins (Fig. 2.1.C,D). The maximum differences in C-alpha RMSF between the loop 1 regions of CENP-

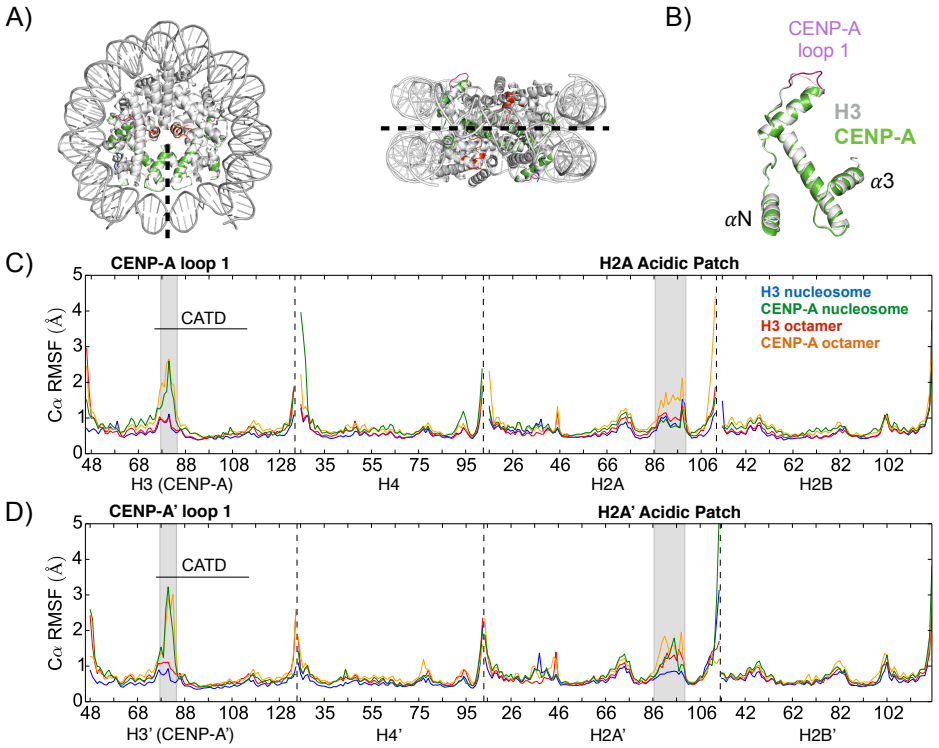


Figure 2.1: **The CENP-A octamer and nucleosome structures display greater local flexibility than their canonical H3 counterparts.** (A) The CENP-A and H3 nucleosomes are nearly identical by crystal structure alignment. Colors label CENP-A and CENP-A' (green), CENP-A loop 1 (purple), and a region spanning the H2A acidic patch (red). The dashed line represents the pseudo-dyad. (B) Structural alignment of CENP-A and H3 proteins highlights the longer CENP-A loop 1 as a major difference. (C) Cα root mean square fluctuations (RMSF) of the first heterotypic half of the CENP-A and H3 structures displays greater local flexibility in the CENP-A systems at several specific regions. (D) Cα RMSF of the second heterotypic half demonstrates specific asymmetries in local flexibility. Dashed lines separate protein segments. Differences in Cα RMSF greater than 0.6 Å are considered very significant. Structure figures rendered in Pymol.

A and H3 is +1.5 Å (comparing CENP-A Gly 81 and H3 Asp 81) and +2.3 Å (comparing CENP-A' Gly 81 and H3' Asp 81). Although they contain the same amino acid sequences, a region spanning the H2A' acidic patch (residues 87-100) and the H4 αN helix (residues 26-29) display greater structural variation in the CENP-A octamer and nucleosome than in the H3 octamer and nucleosome (Fig. 2.1.C,D). For the nucleosomal systems, we observed increased structural fluctuations only for the region spanning the H2A' acidic patch (a maximum difference of +0.93 Å at Lys 95) and for H4 αN helix (a maximum difference of +2.5 Å at Ile 26), and not in their reciprocals. The observed asymmetries may occur because of the different starting points for H4 (Asn 25) and H4' (Asp 24), and different ending points for H2A (Gln 112) and H2A' (Val 114). The additional negatively charged residue of H4' could discourage interactions with negatively charged DNA.

Finally, the CENP-A targeting domain (CATD, residues 75-114) has been predicted to rigidify the CENP-A/H4 interface because of experimentally measured relative increase in hydrophobicity in comparison to the H3/H4 interface [15]. However, consistent with prediction from the crystallographic evidence [6], the only major difference in local flexibility within the CATD we observed is located at the loop 1 region. Overall, CENP-A systems appear to have greater flexibility on the local level compared to the H3 systems in three specific regions: CENP-A loop 1, a region spanning the H2A' acidic patch, and the H4 αN helix.

2.3.2 The CENP-A octamer and nucleosome exhibit greater global fluctuations than their H3 counterparts

Beyond individual residues, structural fluctuations occur at multiple spatial scales for octamers and nucleosomes. Within CENP-A/H4 and H3/H4 dimers, our analysis demonstrate that individual CENP-A/H4 dimer pairs are on average more compact, but not more rigid, than H3/H4 in the context of either the protein octamer, or the entire nucleosome (Supplementary Figure A.3). These data are partially consistent with the prediction from experimental evidence [15, 24, 58]. To further investigate potential rigidity in the CENP-A nucleosome on a global scale, we determined the distances between the centers-of-mass (COM) of dimers in the homotetramer (Fig. 2.2.A), and between dimers in the two heterotetramers (Fig. 2.2.B,C, Supplementary Table A.1, and Supplementary Figure A.4). We find that the distance between CENP-A/H4 dimers is greater than between H3/H4 dimers in both the octamer structures (34.5 $\pm$ 0.29 Å in the CENP-A octamer; and 33.9 $\pm$ 0.21 Å in the H3 octamer) and nucleosome structures (34.7 $\pm$ 0.22 Å in the CENP-A nucleosome; and 34.3 $\pm$ 0.18 Å in the H3 nucleosome). This difference is more evident in the absence of DNA. The average distance between H3 (CENP-A)/H4 and H2A/H2B dimers is virtually identical, with a greater standard deviation, for the CENP-A octamer (33.6 $\pm$ 0.39 Å) than for the H3 octamer (33.7 $\pm$ 0.25 Å). The same average distance is greater for the CENP-A nucleosome (33.7 $\pm$ 0.34 Å) than for the H3 nucleosome (33.4 $\pm$ 0.27 Å). The distance between dimers for the second heterotetramer, between H3' (CENP-A')/H4' and H2A'/H2B', is slightly greater for

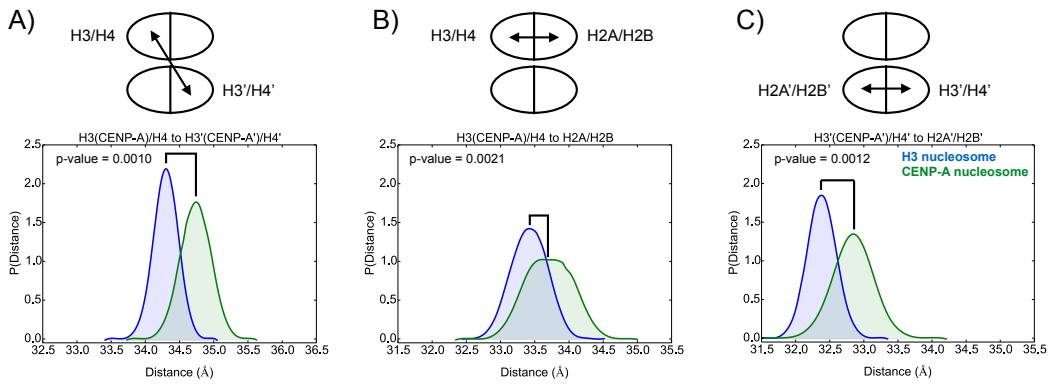


Figure 2.2: **The CENP-A nucleosome exhibits greater global flexibility than the canonical H3 nucleosome.** (A) The distance between the centers-of-mass (COMs) of the two dimers within the homotetramer are greater, on average, in the CENP-A nucleosome than in the H3 nucleosome. The distances between the COMs of the two dimers within the first (B) and second (C) heterotetramers are greater, on average, in the CENP-A nucleosome than in the H3 nucleosome. Specific average distances and standard deviations are included in Supplementary Table A.1.

the CENP-A octamer (33.0 $\pm$ 0.35 Å in the CENP-A octamer; and 32.8 $\pm$ 0.27 Å in the H3 octamer) and significantly greater for the CENP-A nucleosome (32.8 $\pm$ 0.30 Å in the CENP-A nucleosome; and 32.4 $\pm$ 0.21 Å in the H3 nucleosome) compared to their canonical counterparts. In total, these data indicate that the CENP-A octamer and nucleosome display greater fluctuations on a global level than the H3 octamer and nucleosome. The greater distances observed between histone dimers within the CENP-A nucleosome, compared to the canonical H3 nucleosome, are consistent with recently published work [59], which, using fluorescent probes to measure the distance between specifically defined regions of H2B histones, reports that histones in the centromeric nucleosomes are more loosely packed than in the canonical H3 nucleosome (between H2B and H2B').

2.3.3 Greater global fluctuations are underpinned by weaker contacts at the CENP-A:CENP-A' dimerization interface

One plausible reason for the global fluctuations seen above might be increased flexibility in the CENP-A dimerization interface relative to the corresponding interface of H3. To test this hypothesis, we performed contact map analysis to compare the specific contacts that form the H3:H3' and CENP-A:CENP-A' interfaces (Fig. 2.3.A,B). Quantitatively, 4 contacts out of the 15 formed by H3:H3' are lost in the CENP-A:CENP-A' interface (Fig. 2.3.B), and 6 are weakened. For instance, His113/115 is a key residue in the crystal structures of both octameric nucleosomes, binding to Asp123/125 [1]. In our simulations, this contact is still prominent, however, His113/115 is promiscuous, making multiple contacts at the dimerization interface. These contacts in the CENP-A:CENP-A' interface are, on average, weaker than the corresponding set in H3:H3' (Fig. 2.3.A), and several contacts are lost altogether. Another clear difference is the contacts formed by H3 His 113 and H3' Arg 116, and by the reciprocal set of interactions. The corresponding contacts are not present between CENP-A and CENP-A'. This difference is due, in part, to shorter, less well defined, helices composing the four helix bundle of the CENP-A:CENP-A' interface compared to H3:H3' (Fig. 2.3.C) and a longer minimum distance between CENP-A Arg 118 and CENP-A' His 115 (5.3 Å) than between H3 Arg 116 and H3' His 113 (3.7 Å). Furthermore, two hydrophobic interactions present between H3 and H3' (Ile 130 to Ala' 137, and its reciprocal) are lost between CENP-A and CENP-A'. We observed a similar contrast between the H3:H3' and CENP-A:CENP-

A' dimerization interfaces in the octameric structures (Supplementary Figure A.5). Generally, our analysis demonstrates that the CENP-A:CENP-A' dimerization interface is weaker than that of H3:H3', suggesting that the CENP-A:CENP-A' interface could exhibit greater plasticity.

2.3.4 Greater plasticity at the CENP-A:CENP-A' dimerization interface is a dominant mode of motion

We tested the possibility of increased plasticity in the CENP-A dimerization interface by performing principal component analysis (PCA), which determines the dominant modes of motion in each system. Projections of the top principal components from Cα PCA onto representative structures in 2D plots (Fig. 2.4 and Supplementary Figure A.6) and 3D movies (Supplementary Movies 1 and 2) illustrate the magnitude and direction of dominant motions for the H3 and CENP-A structures. These principal components clearly contrast the natures of the motion associated with the H3:H3' and the CENP-A:CENP-A' dimerization interfaces. Interestingly, in the top two principal components for the H3 nucleosome, H3 and H3' move together as a single unit in a concerted manner, maintaining the integrity of their interface (Fig. 2.4.A and Supplementary Movie 1). In contrast, in the first principal component of the CENP-A nucleosome, CENP-A and CENP-A' visibly move separately from each other in a shearing motion (Fig. 2.4.B and Supplementary Movie 2). Furthermore, in their respective top principal components, the CENP-A nucleosome exhibits greater breathing motion (a global opening and closing of the central

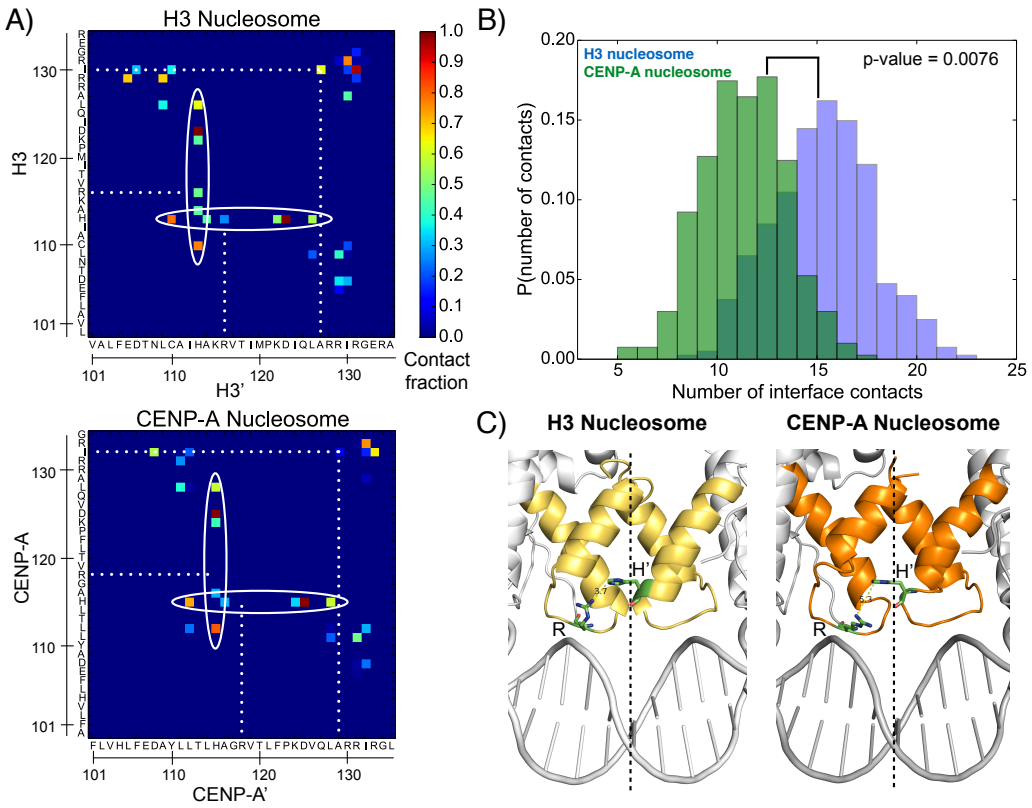


Figure 2.3: **The CENP-A dimerization interface forms fewer and weaker contacts than the corresponding H3 interface.** (A) Contact map analysis illustrates the promiscuous interactions of His113/115 (circled) and contacts present in the dimerization interface of the H3 nucleosome but not in the corresponding interface of the CENP-A nucleosome (dotted lines). (B) Histograms demonstrate that the number of contacts formed at the CENP-A dimerization interface is fewer, on average, than at the corresponding H3 interface. (C) Crystal structure dimerization interfaces for the CENP-A (dark orange) and H3 (light orange) nucleosomes highlight that the four helices composing the CENP-A:CENP-A' interface are less well-defined than the four helices composing H3:H3'. Residues Arg 116 and His' 113 of H3:H3', as well as Arg 118 and His' 115 of CENP-A:CENP-A', are shown as sticks. The dashed line represents the pseudo-dyad. Structure figures rendered in Pymol.

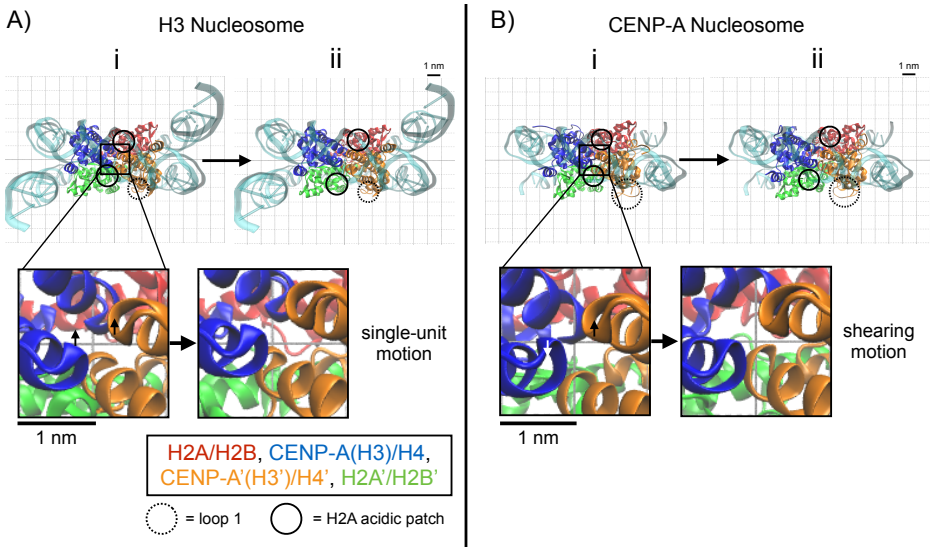


Figure 2.4: **Increased plasticity of the CENP-A dimerization interface is a dominant mode of motion.** (A) 2D projection of the top principal component of the H3 nucleosome onto the representative structure, viewed from the side of the DNA pseudo-dyad, highlights mobility in the H2A' acidic patch restrained by H3' and a lack of movement in H3' loop 1. The inset displays the single-unit motion of the H3 dimerization interface. (B) 2D projection of the top principal component onto the representative structure for the CENP-A nucleosome highlights mobility in the H2A' acidic patch unrestrained by CENP-A' and enhanced fluctuations in CENP-A' loop 1. The inset displays the shearing and pinching motion of the CENP-A dimerization interface. Structure figures rendered in VMD.

void) than the H3 nucleosome (Supplementary Movies 1 and 2), and the four helices defining the CENP-A:CENP-A' interface display anti-correlated pinching motions (Supplementary Movie 2). We observed a similar contrast between the dimerization interfaces of the H3 and CENP-A octamers. In their respective first two principal components in the octamer structures, H3 and H3' move together at their interface (Supplementary Figure A.6.A) and CENP-A and CENP-A' move away and towards each other at their interface (Supplementary Figure A.6.B). Overall, the PCA analysis indicates that a weakened interface drives greater plasticity at the CENP-A:CENP-A' dimerization interface within the CENP-A nucleosome.

2.3.5 H2A patch mobility is sensitive to surrounding local interactions

The acidic patch is a unique structural motif of the nucleosome surface, carrying the greatest net charge of the solvent-exposed region of the histone octamer surface [1, 60]. Formally defined by eight residues (H2A E56, E61, E64, D90, E91, E92 and H2B E102, E110), the acidic patch forms a complex interface, with a high negative charge density and distinct groove shape. Furthermore, the acidic patch is topologically poised to interact with multiple types of chromatin factors, including the H4 N-terminal tail of an adjacent nucleosome [1], Sir3 [61], LANA [62], and the CENP-A specific CENP-C [63], essential in forming the inner kinetochore. From our investigation of local flexibility, we identified an important region of H2A (residues 87-100) spanning the H2A C-terminal extension part of the acidic patch (H2A D90,

E91, and E92). This region contains a short α-helix bracketed between H2A α3 and the long H2A C-terminal tail. Importantly, the region spanning the H2A acidic patch undergoes similar structural fluctuations in the first heterotypic half of the CENP-A and H3 nucleosomes (Fig. 2.1.C) and greater structural variation in the CENP-A nucleosome than in the H3 nucleosome in the second heterotypic half (Fig. 2.1.D). We first examined the results from principal component analysis to investigate this asymmetric behavior.

The PCA analysis revealed differences in the modes of motion of the CENP-A and H3 systems in addition to those found at the dimerization interfaces. We observed structural fluctuation in the H2A' acidic patch of the H3 and CENP-A nucleosomes in the top principal components, but the natures of these fluctuations are very different (Fig. 2.4). For the H3 nucleosome, the H2A' acidic patch moves together with the H3' C-terminus, a salt bridge between H3' Arg 134 and H2A' Glu 91 playing a contributing factor (Fig. 2.4.A and Supplementary Movie 1). In contrast, in the first principal component of the CENP-A nucleosome, the H2A' acidic patch moves up and down the DNA supercoil axis, and into and out of the central void, independently of the CENP-A' C-termini (Fig. 2.4.B and Supplementary Movie 2), which lacks the positively charged Arginine seen in H3 (i.e. Arg 134). Therefore, PCA analysis provides insight into the mobility of the region spanning the H2A' acidic patch in the CENP-A nucleosome. The top principal components of octameric structures indicate a similar increase in the overall mobility of the region spanning the H2A' acidic patch for the CENP-A system compared to H3 in the absence of DNA (Supplementary Figure A.6). Because of the asymmetry between

structural fluctuations of the regions spanning the H2A and H2A' acidic patches (Fig. 2.1.C vs 2.1.D), we decided to investigate further by extending our MD simulation by 400 ns with reduced position restraints (reduced from K = 2.5×10^{-1} kJ/(mol nm^2) to K = 5.0×10^{-5} kJ/(mol nm^2)). We concluded that the H2A acidic patch occupies a rugged conformational landscape, and the overall mobility depends on transient, local interactions. We focused on the H2A' acidic patch, where a significant and variable difference exists between the CENP-A and H3 nucleosomes (Supplementary Figure A.7). We identified specific, local electrostatic interactions that contribute to H2A' acidic patch mobility in the H3 nucleosome (between H3' R134 and H2A' E91, Fig. 2.5.A) and in the CENP-A nucleosome (between H2A' R99 and E91, Fig. 2.5.B). The stochastic formation and disruption of these interactions, among others, can alternatively restrain and relax the H2A' acidic patch, demonstrating that its mobility is highly sensitive to the dynamic rearrangement of local interactions.

2.3.6 CENP-A nucleosomal DNA near the pseudo-dyad is relatively unstable compared to the corresponding H3 nucleosomal DNA

We were curious whether the local and global fluctuations in the CENP-A nucleosome noted above (Fig. 2.1-2.4) had an impact on the DNA wound about the CENP-A nucleosome. Past experiments have shown that CENP-A protects ∼120bp of DNA relative to H3, which almost exclusively protects 147bp of DNA. The difference in binding has been attributed to missing Arginines in the αN he-

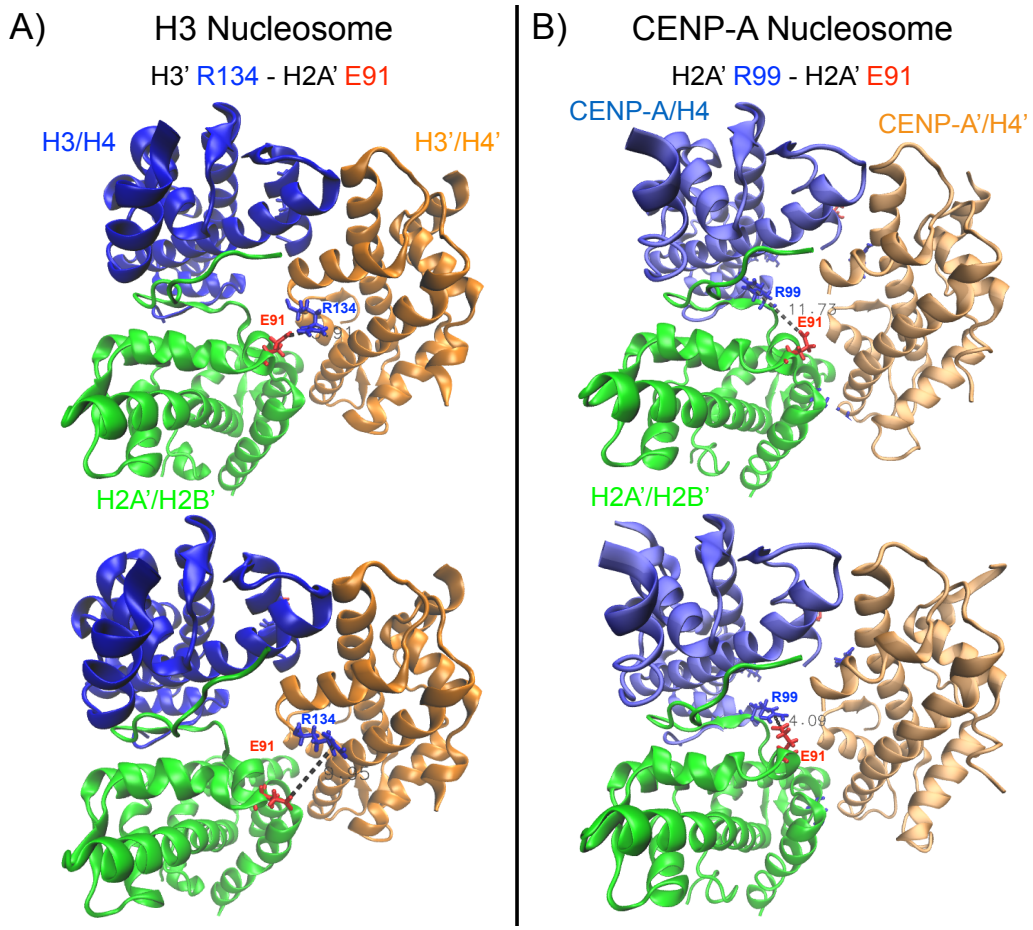


Figure 2.5: **H2A' acidic patch mobility depends on local, electrostatic interactions.** Representative simulation snapshots highlight electrostatic interactions contributing to the mobility of a region spanning the H2A' acidic patch in (A) the H3 nucleosome, between H3' Arg134 and H2A' Glu91, and in (B) the CENP-A nucleosome, between H2A' Arg99 and Glu91. Colored sticks represent positive residues (in blue) and negative residues (in red). Dashed lines measure the distances between the center of positive charge on the arginines, CZ, and the center of negative charge on the glutamic acid, CD. The surrounding superhelical DNA and the H2A/H2B dimers are removed to facilitate observation. Structure figures rendered in VMD.

lices of CENP-A [44]. Recent computational short-timescale simulations have suggested the exit/entry helices of DNA have weaker interactions with the CENP-A protein core [29]. Exit and entry DNA "site exposure" is an important biological feature [64], and in the context of the CENP-A nucleosomes, could potentially destabilize H2A/H2B binding, resulting in losing this dimer pair from the octamer, thus destabilizing the nucleosome.

In our long-timescale simulations, we noted that the entry/exit fluctuations of DNA are similar around the H3 and CENP-A nucleosome, with roughly equal propensity for local and asymmetric unwinding. The H3 nucleosome displays asymmetric site exposure, freeing up ∼15bp of the exit nucleosomal DNA consistent with recent biochemical experiments suggesting asymmetric behavior in the unwinding of H3 nucleosomal DNA *in vivo* [65] and *in vitro* [66]. ∼15bp of the entrance nucleosomal DNA becomes completely exposed in the CENP-A nucleosome, the detachment occurring at the opposite side to the H3 nucleosome. It should be noted that this type of entry/exit detachment event could be stochastic in nature, because *in vivo*, DNA traces a continuous path via the linker DNA to adjacent nucleosomes in the array, and the presence or absence of proteins such as H1 (for the H3 nucleosome), CENP-C and CENP-B (for the CENP-A nucleosome), will likely alter the stability and crossing-over of exit and entry DNA. Thus, asymmetric unpeeling of the palindromic (i.e. symmetric) α-satellite based DNA sequences resolved in the crystal structures could be a probabilistic event, in part, governed by structural motifs that may exist within the centomeric DNA sequences *in vivo* [65, 66]. Furthermore, the asymmetric unpeeling observed in our simulations could be an example of sponta-

neous symmetry breaking that is stochastically induced by structural fluctuations, either locally or at a large scale.

Significantly, in our long timescale simulations above (Fig. 2.6), we observed that CENP-A nucleosomal DNA exhibits greater structural fluctuations near the pseudo-dyad, from SHL +1 to +3, which extrude away from the octameric core surface. Throughout these 20 bp of DNA, the CENP-A DNA basepair fluctuations were on average ∼1.0 Å greater than the corresponding base pairs in the H3 nucleosome (Fig. 2.6.A inset). We performed event coincidence analysis (Fig. 2.6.B,C,D and Supplementary Figure A.8) to unveil factors contributing to the release of ∼20bp pseudo-dyad proximal DNA. We tracked several key residues in CENP-A and H4, finding an increase in the distance between CENP-A H59 (a substitution for E59 in H3) and the entry helix of nucleosomal DNA at bp -60 (Supplementary Figure A.8.A) upon detachment. This stochastic event coincides with CENP-A' H59 swinging inwards to DNA SHL +1 or +2. This swing disrupts the local interactions between CENP-A' and DNA, including CENP-A' K64 and bp +19, causing the ripple of DNA moving away from the protein core. H59 (a positive residue) in CENP-A replaces E59 (a negative residue) of H3. H59 is located near the N-terminus of CENP-A, and is important for the stability of CENP-A nucleosomal entry DNA, forming electrostatic interactions with the negatively charged DNA (Fig. 2.6.B). Once this section of DNA detaches from the CENP-A protein octamer, CENP-A H59 is available to form alternating interactions between two pseudo-dyad proximal turns of DNA (Fig. 2.6.C and Supplementary Figure A.9). These interactions alone can increase the base pair structural fluctuations because they are not found in the

native CENP-A crystal structure. H59 and K64 compete for the same interaction with DNA, both positive residues naturally repel each other, and both favor similar electrostatic interactions with the highly negatively charged DNA. The replacement of H3 E59 with CENP-A H59 is definitely an important factor in CENP-A nucleosomal DNA flexibility near the pseudo-dyad, however, it is not the only factor. This ripple is amplified by an intrinsically weak interaction of this section of DNA because of a missing R83 present in H3, but not in CENP-A. Finally, a substitution of L82 (present in H3) by a F84 in CENP-A, creates a hydrophobic hotspot with CENP-A R80 and H4 K79 (Supplementary Figure A.8.B,C), weakening K79's affinity to DNA. Common contacts between two arginine residues, counterintuitive from an electrostatic perspective, have been investigated both from the vast number of experimentally determined crystal structures within the Protein Data Bank [67], and by molecular dynamics simulation [68, 69]. Arginine and lysine residues both have long side chains, which feature a positively charged head group and a hydrophobic tail. Examining Fig. 2.6.D in close detail, we observe that F84 (a hydrophobic residue) interacts with the long tail regions of R80 and K79, both of which are hydrophobic. The accumulation of CENP-A DNA endpoint detachment, CENP-A H59 replacing H3 E59, and the formation of a hydrophobic hotspot within the CENP-A nucleosome conspire to release ∼20bp of DNA from the pseudo-dyad proximal region of CENP-A nucleosome.

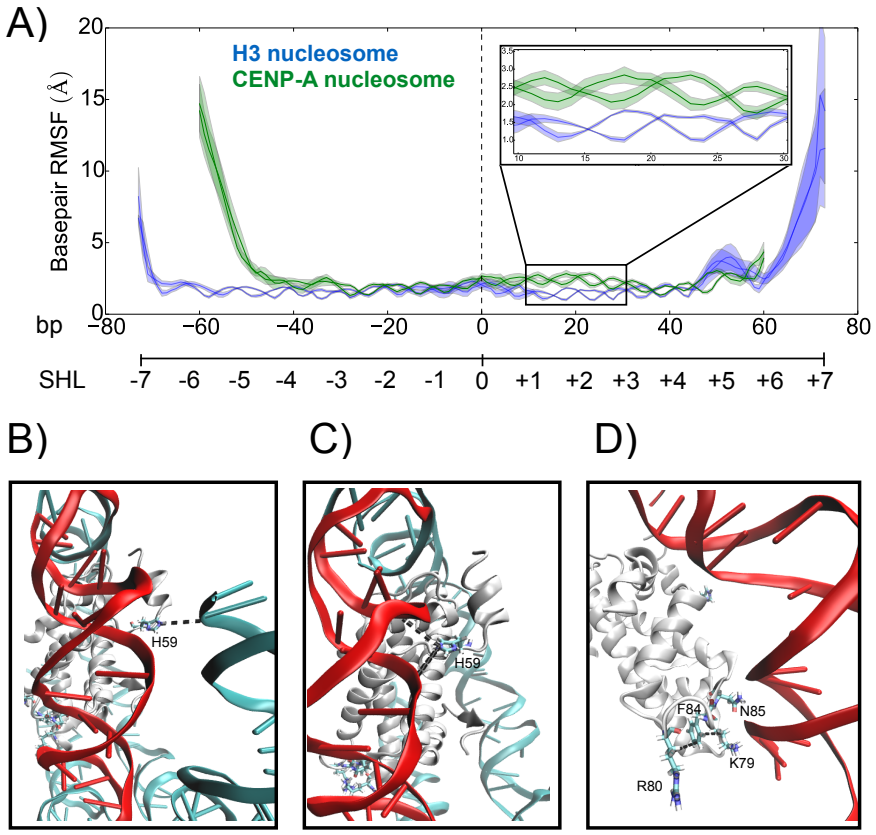


Figure 2.6: **CENP-A nucleosomal DNA is relatively unstable near the pseudo-dyad compared to the corresponding H3 nucleosomal DNA.** (A) Root mean square fluctuations (RMSF) per basepair for the CENP-A and H3 nucleosomes demonstrates asymmetric entry/exit unwinding for both systems and further highlights significantly increased fluctuations for CENP-A basepairs +10 to +30 (inset). Two lines are shown for each system since the DNA is double-stranded, and the shaded areas represent the RMSF ± one standard deviation. (B) His 59, unique to CENP-A, has electrostatic interactions with the entry terminal of DNA before it detaches. (C) Coinciding with the detachment of entry DNA, His59 turns towards DNA SHL +1 or +2, disrupting the local stability of DNA. (D) N85 in CENP-A, substituted for R83 in H3, has a weaker binding affinity to DNA, and two other substitutions, F84 and R80, develop a hydrophobic hotspot with K79 of H4, further contributing to DNA instability near the pseudo-dyad. Basepairs 10-30 are highlighted in red, key residues are shown as sticks, and CENP-A interactions not found in the H3 nucleosome are represented by dashed lines. Structure figures rendered in VMD.

2.3.7 The CENP-A nucleosome features a more rugged free energy landscape than the canonical H3 nucleosome

We compared the conformational space explored by the nucleosome systems by projecting the trajectories onto their corresponding top two principal components, thereby mapping two-dimensional free energy landscapes (Fig. 2.7) [70, 71]. The landscape topography of the canonical H3 nucleosome contains broad and well-connected basins, whereas the number of basins, and the barriers between basins, is greater for the CENP-A nucleosome. When examining representative structures of the free energy basins for the H3 nucleosome, we observe different conformations for the H2A' acidic patch, concordant with differences in the H3' C-terminus (Fig. 2.7.A). In contrast, there are multiple conformations for the CENP-A:CENP-A' dimerization interface in the distinct basins for the CENP-A nucleosome, featuring different arrangements of the four helices defining this interface (Fig. 2.7.B). Representative structures for the CENP-A nucleosome also display different conformations for CENP-A loop 1 and the H2A' acidic patch. Specifically, the H2A' acidic patch is more disordered in CENP-A representatives 1 and 2, and more ordered in representatives 3 and 4 (Fig. 2.7.B). Globally, the CENP-A nucleosome free energy landscape covers more conformational space (i.e. greater overall area) than the H3 nucleosome free energy landscape. From our two-dimensional free energy landscapes of the CENP-A and H3 nucleosomes, we calculated the average free energy barriers between distinct conformational basins. For the CENP-A nucleosome, we determined the free energy barriers between adjacent, and accessible,

conformational basins to be ∼2 to 4 k_BT, and for the single free energy barrier for the H3 nucleosome to be ∼2 k_BT. This calculation illustrates that the CENP-A nucleosome occupies a more rugged free energy landscape, and that both CENP-A and H3 nucleosomes can overcome free energy barriers to convert between different conformational basins. Overall, the CENP-A nucleosome has more distinct conformational basins than the H3 nucleosome, and the structural differences between free energy basins correspond to the contrasts in local and global fluctuations we observed through other modes of analysis.

2.4 Discussion

Here, we present the first microsecond timescale all-atom computational investigation of the conformational dynamics of canonical H3 and CENP-A variant octamers and nucleosomes in explicit solvent. This significantly extended timescale relative to previous explicit solvent all-atom studies of nucleosomes containing H3 [26, 28, 30] and CENP-A [29] allowed our analyses to reveal the internal protein dynamics of the H3 and CENP-A nucleosomes at stable equilibrium (Supplementary Figure A.2). For H3, we observe consistency with previously published experimental results on the stability of the DNA near the pseudo-dyad (Fig. 2.6) [1,4,72,73], and asymmetric unwinding of the entry and exit DNA (Fig. 2.6) [66], and with previous computational studies identifying structural mobility in H2A (Fig. 2.4) [30] and flexibility at the entrance and exit DNA (Fig. 2.6) [28]. For CENPA, we also observe consistency with previously published results and predictions of flexible DNA

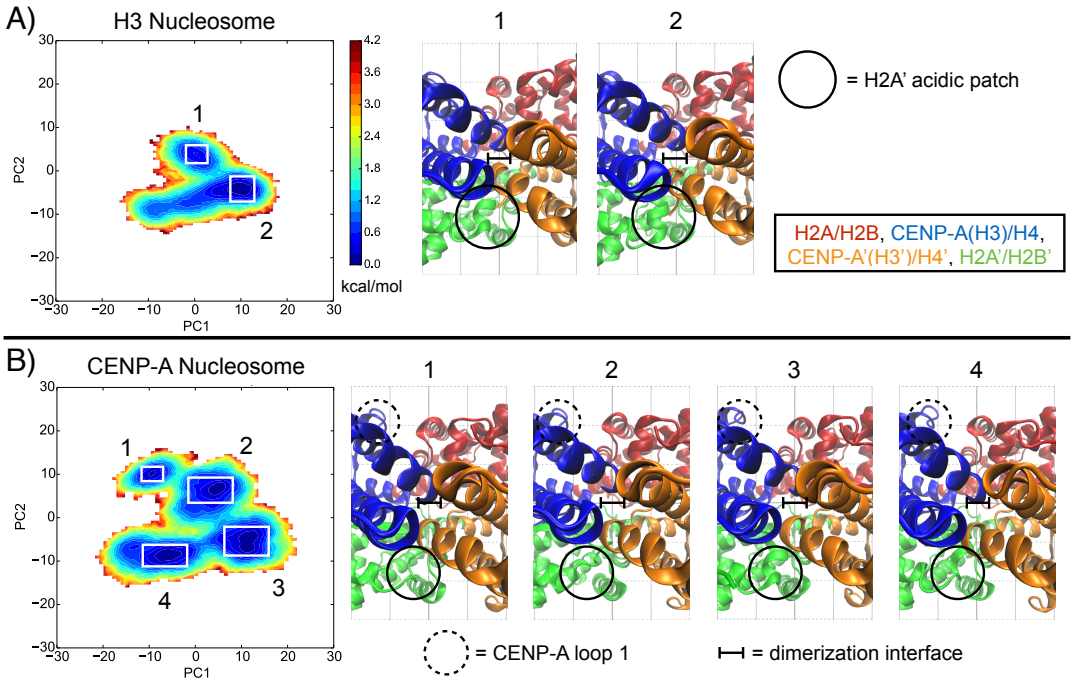


Figure 2.7: **The characteristic free energy landscape is more rugged for the CENP-A nucleosome than for the H3 nucleosome.** A) Free energy projection of the H3 nucleosome onto its first two principal components reveals two distinct conformational basins. B) Free energy projection of the CENP-A nucleosome reveals four distinct conformational basins. White boxes highlight distinct free energy basins. Insets highlight key differences between representative structures for the numbered basins, consistent with the rest of our analysis. Structure figures rendered in VMD.

edges (Fig. 2.6) [6, 11, 16, 29, 44, 46, 74–79], greater structural variation in CENP-A loop 1 relative to the same region of H3 (Fig. 2.1), predicted experimentally [6] and observed computationally [29], as well as increased CENP-A/H4 dimer compactness relative to H3/H4 (Supplementary Figure A.3) [58]. However, using long timescale simulations coupled to principal component analyses and free energy landscape theory, our data suggest that the CENP-A nucleosome is structurally flexible on the local and global scale.

Most importantly, the overall increase of global flexibility in the CENP-A nucleosome (Fig. 2.2) is underpinned by a reduced number of contacts in the CENP-A:CENP-A' dimerization interface (Fig. 2.3). This results in the CENP-A:CENP-A' four helix bundle undergoing a distinctive and unique shearing motion coupled to a pinching motion (Fig. 2.4 and Supplementary Movie 2). This motion translates to a flexing of the entire CENP-A nucleosomal and octameric particles, such that CENP-A containing complexes explore significantly more rugged energy landscapes (Fig. 2.7 and Supplementary Figure A.10). A major consequence of this rugged landscape is additional free energy minima are available to the CENP-A nucleosome, which can spontaneously visit these basins with near equal probability. Plasticity in the CENP-A dimerization interface is driven by the dynamic rearrangement of mostly weak contacts between the $\alpha 2$ and $\alpha 3$ helices of the CENP-A proteins (Fig. 2.3). In addition to the CENP-A dimerization interface, several other specific regions contribute to the greater overall flexibility of the CENP-A nucleosome compared to the H3 nucleosome.

First, we note that the mobility of the region spanning the acidic patch of the

H2A depends on the surrounding local interactions in the CENP-A and H3 nucleosomes (Fig. 2.1 and 2.5). The stochastic formation and disruption of electrostatic interactions at the interface of the H2A acidic patch and other histones, and within the H2A acidic patch, can alternatively restrain and relax this region (Fig. 2.5). Variable mobility of H2A in the CENP-A could effect its ability to interact freely with non-histone proteins, such as the kinetochore protein CENP-C, which has experimentally been shown to bind to the acidic patch of H2A in the context of a hybrid CENP-A fusion nucleosome [80]. In addition, the highly flexible CENP-A loop 1 (Fig. 2.1 and 2.4.B) could play an important regulatory role in the formation of higher order CENP-A chromatin [81]. Furthermore, CENP-A nucleosomal DNA displays asymmetric instability near the pseudo-dyad, relative to H3 nucleosomal DNA (Fig. 2.6 and Supplementary Figure A.9). One predicted outcome of such instability near the pseudo-dyad is increased access to the protein octameric core. These data could provide a potential explanation for experimentally observed acetylation seen within the CENP-A nucleosomal core at H4 K79 *in vivo* [19].

Lastly, these *in silico* findings have important implications for the behavior of CENP-A nucleosomes *in vitro* and *in vivo* [11, 12, 14, 16, 24]. Specifically, our data shows the loss of DNA contacts near the pseudo-dyad (Fig. 2.6.D), a weaker CENP-A dimer interface (Fig. 2.3), coupled to shearing of the four-helix bundle (Fig. 2.4). Even in the DNA-free CENP-A octameric core, we observe global shearing motions between the two CENP-A/H4/H2A/H2B heterotetramers (Supplementary Figure A.4.A).

So far, there is no direct experimental evidence testing whether or how oc-

tameric nucleosomes can unpeel at the pseudo-dyad to generate two hemisomes [14, 82]. However, based on these results, it is plausible that CENP-A hemisomal intermediates might reflect increased distance and weaker contacts between two heterotypic halves of the octameric CENP-A nucleosome, coupled to looser DNA contacts at pseudo dyad proximal region, both of which could be exaggerated by specific biological conditions *in vivo*. Our simulations of DNA-free octameric CENP-A particles do not appear to show striking disruptions of H2A/H2B in the octameric core, as would be predicted from a hexameric structure [12]. However, we note that H2A/H2B have slightly increased distance from CENP-A/H4 in the CENP-A nucleosome (Fig. 2.2). Thus, as has recently been noted for H3 nucleosomes [65, 83], it is possible that stochastic asymmetric entry/exit DNA site exposure, coupled to chromatin remodelers such as RSF [84], could effect the eviction of one or both H2A/H2Bs in the CENP-A nucleosome.

It has already been demonstrated that *in vitro*, and *in vivo*, internal covalent modifications of H3 can alter canonical nucleosomal conformation and stability [85–88], which may utilize alternative modes of internal motion relative to the motions described for CENP-A in this study. The data presented above support the possibility that *in vivo* CENP-A nucleosomes subjected to pulling, pushing or twisting mechanical forces may geometrically adapt to extrinsically imposed deformations by exploiting internal pliability. Such conformational changes in the CENP-A nucleosome are likely to be promoted or prohibited by specific inner kinetochore proteins like CENP-C, or specific modifications, which may predominantly favor one conformational basin of CENP-A over another at specific points of the

cell cycle, or in response to one or more binding partners [63]. Consequently, interrogating the effects of biological forces, covalent modifications, point and domain mutations, and the binding of kinetochore and non-kinetochore proteins on the stability of CENP-A nucleosomes are critical and exciting future avenues of research awaiting investigation. It is feasible that internal flexibility within the CENP-A octameric nucleosome may permit exploration of multiple conformations, contributing to CENP-A's structural and epigenetic signature *in vivo*.

2.5 Summary and Biological Implications

In summary, when the centromere-specific histone variant CENP-A replaces H3, the nucleosome is more distortable on both local and global scales. This means that a nucleosome containing CENP-A could conform better to mechanical forces experienced during mitosis than nucleosomes containing the canonical H3 histone. Furthermore, the CENP-A nucleosome could be more likely to adopt different conformations under specific biological conditions *in vivo* than the canonical H3 nucleosome. Overall, we find that CENP-A encodes greater distortability to the nucleosome, which may allow for enhanced flexing of the histone core during mitotic tension.

Chapter 3: CENP-A/H4 is more plastic than H3/H4

3.1 Introduction

The DNA of chromatin must organize into an array of nucleosomes, and histone dimers assemble together in order to form the protein component of nucleosomes. Variants of the most common, canonical histones encode identity to specialized regions of the genome, and play an important role in the epigenetic process. CENP-A/CENH3, a centromere specific histone H3 variant present in all eukaryotes, dictates the single location of the centromere per chromosome, to which microtubules bind during mitosis. CENP-A is also thought to confer structural identity to the centromere. The CENP-A/H4 dimer is an essential subdomain of CENP-A nucleosomes, which are currently shown to occupy a range of structures *in vivo* [6–23]. Subtle differences between CENP-A and H3 could lead to changes in dimer dynamics, even though the CENP-A/H4 and H3/H4 dimers are structurally superimposable (Figure 3.1.A). By interacting directly with highly positively charged histone monomers and dimers and by providing electrostatic shielding, histone binding proteins play an essential role in nucleosome assembly. It was recently discovered that the chaperone HJURP (Holliday junction recognition protein) is required for the deposition of CENP-A at centromeres [89, 90], an analog to the CENH3-specific

chaperone Scm3 in budding yeast [10, 12, 13, 91–93]. To investigate dimer dynamics and the role of CENP-A specific HJURP, we modeled CENP-A/H4 and H3/H4, in the absence and presence of the chaperone HJURP, using all-atom molecular dynamics (MD) in explicit solvent.

The CENP-A/H4 dimer demonstrates greater structural plasticity, spread throughout more local regions, than the canonical H3/H4. In the absence of chaperone HJURP, the C-terminal helices of CENP-A and H4 become partially unwound. Upon the introduction of HJURP, these regions maintain their helical structure due to the formation of an electrostatic network between the C-termini of CENP-A and H4 and the α domain of HJURP. Finally, H4 adopts configurations closer to the native state than CENP-A or H3, except for in the artificial complex of H3/H4 in conjunction with HJURP. The underlying physical principles determining the conformational ensembles of histone dimers in isolation are not necessarily the same as those governing the conformational preferences within the histone octamer core, or in complexes with chaperone proteins. However, understanding the different dynamics of CENP-A/H4 and H3/H4 dimers provides a structural characterization baseline that can inform future research of centromeric and canonical nucleosomal structures.

3.2 Methods

3.2.1 Simulation protocol

We performed all-atom molecular dynamics (MD) in explicit solvent using the gromacs 4.5.7 MD software [47], the amber99SB*-ILDN [48, 49] force field for proteins, the ions08 [94] force field for ions, and the TIP3P water model. Starting from the crystal structures for canonical H3 nucleosome (PDB ID: 1AOI [1]) and the CENP-A/H4 heterodimer with chaperone HJURP (PDB ID: 3R45 [95]), we built models for four systems: (1) the H3/H4 heterodimer; (2) the CENP-A/H4 heterodimer; (3) the H3/H4 heterodimer with the CENP-A specific chaperone HJURP (as a control); and (4) the CENP-A/H4 heterodimer in a complex with the chaperone HJURP.

Using the *pdb2gmx* tool in gromacs, we set the Lys and Arg residues to +1e, the Asp and Glu residues to -1e, the Gln residues to neutral, and protonated the His residues solely at NE2. Each system was solvated in a cubic water box, ensuring a minimum buffer length of 15 Å between the system and the edges of the box. We introduced Na^+ and Cl^- ions to neutralize the charge and represent the physiological 0.150 M NaCl environment. The systems were minimized using steepest descent, until reaching a maximum force less than 100 kJ/mol. Periodic boundary conditions were employed throughout all the simulations, and long-range electrostatics were treated with the Particle Mesh Ewald method [52]. Non-bonded Coulomb and Lennard-Jones interactions were truncated at 10 Å, and all bonds involving hy-

drogen were constrained using the LINCS [53] algorithm. After minimization, the systems were heated to 300 K by 500 ps of protein-restrained NVT MD simulation followed by 500 ps of NVT MD simulation without restraints. After reaching thermal equilibrium, the systems were equilibrated at 300 K and 1.0 bar for 1.5 ns in the NPT ensemble.

To characterize the structure and dynamics of the canonical and CENP-A heterodimers, we performed unrestrained production all-atom MD simulations in the NPT ensemble at 1.0 bar and 300 K with a 2 fs time-step, saving coordinates, velocities, and energies every 2 ps for further analysis. We updated the list of non-bonded neighbors every 10 steps. One microsecond of MD simulations was performed for each system using the V-rescaled, modified Berendsen thermostat [54] with a 1.0 ps time-constant and the Parrinello-Rahman barostat [55] with a relaxation time of 2.0 ps. We only considered the final 400 ns of the trajectories for analysis to account for further temperature and pressure equilibration.

3.2.2 Analysis of the trajectories

We first determined the root-mean-square deviations (RMSD) of all the Cα atoms of the CENP-A/H4 and H3/H4 dimers with respect to their corresponding crystal structures. Then, we identified specific residues, and protein regions, that exhibit greater local flexibility by calculating the root-mean-square fluctuations (RMSF) of Cα atoms with respect to the corresponding geometric average structures for all of the studied systems. To visualization local flexibility, we calcu-

lated the B-factor from the RMSF, ($B = (8\pi^2/3)$*(RMSF)^2), and displayed dimer structures as tubes whose widths correspond to these values. Furthermore, we determined the distances between the C-termini of CENP-A (or H3) and H4 in dimeric structures, and investigated the effects of a specific salt-bridge (between H3 E133 and H4 R95) on this distance for H3/H4. We considered a salt bridge to form between an arginine and a glutamic acid when the center of positive charge of the arginine, CZ, and one of the negatively charged oxygen atoms of the a glutamic acid, OE1 or OE2, were within 4.0 Å. We also analyzed contacts at the interface of CENP-A and H4, in the absence and presence of HJURP. A contact was determined to exist when the distance between two non-hydrogen atoms from different residues was less than 3.6 Å. Contacts were calculated as fractions of time of their respective entire trajectories.

To analyze our data from a more global perspective, we calculated a specific measure of structural similarity, Q, of all the snapshots in our computational trajectories to the experimentally determined crystal structures. Q is a normalized order parameter, with higher values indicating greater similarity between two structures.

$$Q = \frac{2}{(N-2)(N-3)} \sum_{i>j+2} \exp\left[-\frac{\left(r_{ij} - r_{ij}^n\right)^2}{2\sigma_{ij}^2}\right], \tag{3.1}$$

where N is the total number of Cα atoms, r_{ij} is the instantaneous distance between the Cα atoms of residues i and j, r_{ij}^n is the same distance in the native state obtained from experiment, and σ_{ij} is a resolution parameter where $\sigma_{ij} = (1 + |i - j|)^{0.15}$. We applied this definition to histone monomers CENP-A (or H3)

and H4 separately.

3.3 Results and Discussion

Here, we explore the effects of an important histone variant on the structure and dynamics of a histone dimer. Using all-atom molecular dynamics (MD) simulations in explicit solvent on the microsecond timescale, requiring over 300,000 cpu hours in total, we investigated the structural flexibility and dominant local interactions of the CENP-A/H4 dimer in comparison the canonical H3/H4, based on the crystal structures for the CENP-A/H4 heterodimer with chaperone HJURP (PDB ID: 3R45 [95]) and canonical H3 nucleosome (PDB ID: 1AOI [1]). We studied four systems in total: (1) the H3/H4 heterodimer; (2) the CENP-A/H4 heterodimer; (3) the H3/H4 heterodimer with the CENP-A specific chaperone HJURP (as a control); and (4) the CENP-A/H4 heterodimer in a complex with the chaperone HJURP. Our analysis reveals dynamics that deviate significantly from the nearly identical crystal structures of the CENP-A/H4 and H3/H4 dimers (Figure 3.1.A), which are due, in part, to the significant dissimilarities in the amino acid sequences of the CENP-A and H3 proteins (Supplementary Figure B.1). Sequence and structural alignment highlights the longer loop 1 region and $\alpha 3$ in CENP-A compared to the corresponding regions of H3 (Figure 3.1.A and Supplementary Figure B.1). We observed that it took several hundred nanoseconds for all systems to reach equilibrium. We ran our simulations for a full microsecond each and harvested the stable final 600 ns of each trajectory for subsequent analyses (Supplementary Figure B.2).

3.3.1 CENP-A/H4 exhibits greater structural plasticity than H3/H4

We compared the overall structural flexibility of the CENP-A/H4 and canonical H3/H4 dimers by calculating the root-mean-square deviations (RMSD) of Cα atoms with respect to their corresponding experimentally resolved crystal structures. The CENP-A/H4 dimer displays greater RMSD, on average, than H3/H4 (Figure 3.1.B). However, both CENP-A/H4 and H3/H4 explore native state dynamics on the microsecond time scale, adopting conformations within 3.0 Å of their respective crystal structures. Unlike CENP-A/H4, the H3/H4 RMSD probability distribution features two prominent peaks, suggesting the rearrangement of specific interactions in the H3/H4 dimer. Greater plasticity in the CENP-A/H4 dimer, compared to H3/H4, could translate to greater plasticity within and between nucleosomes containing CENP-A than in their canonical H3 counterparts. Furthermore, CENP-A and H3 interact with different proteins in the nucleosome and when binding to chaperones. Therefore, it is important to investigate the local flexibility of specific residues of the CENP-A and H3 dimers.

3.3.2 More regions in CENP-A/H4 display local mobility than in H3/H4

We next examined the local structural mobility of the CENP-A and H3 dimers by determining the root-mean-square fluctuations (RMSF) of Cα atoms, with respect to their time-average positions. For visualization, we calculated the B-factor,

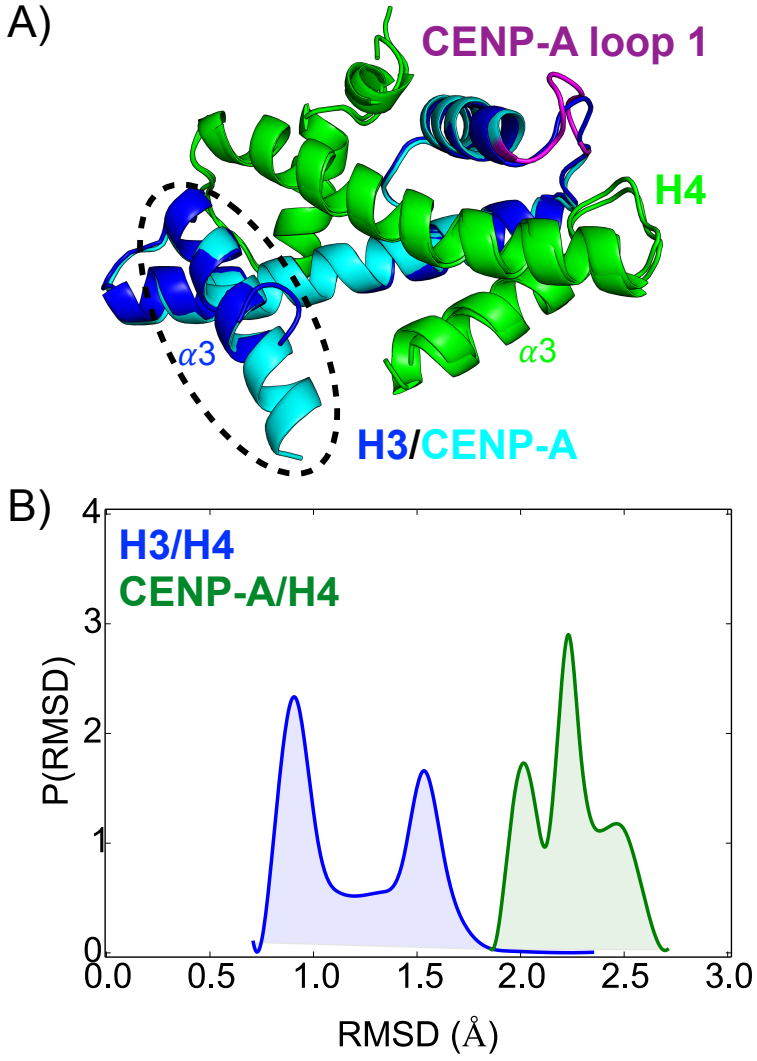


Figure 3.1: **The CENP-A/H4 dimer exhibits greater structural plasticity than H3/H4.** (A) Structural alignment of CENP-A/H4 and H3/H4 highlights the longer loop 1 and $\alpha 3$ of CENP-A than the corresponding regions of H3 as major differences. Colors label H3 (dark blue), CENP-A (light blue), CENP-A loop 1 (purple), and H4 (green). The dashed oval identifies the C-terminal regions of CENP-A and H3. (B) Cα root mean square deviations (RMSD) with respect to the crystal structures of CENP-A/H4 and H3/H4.

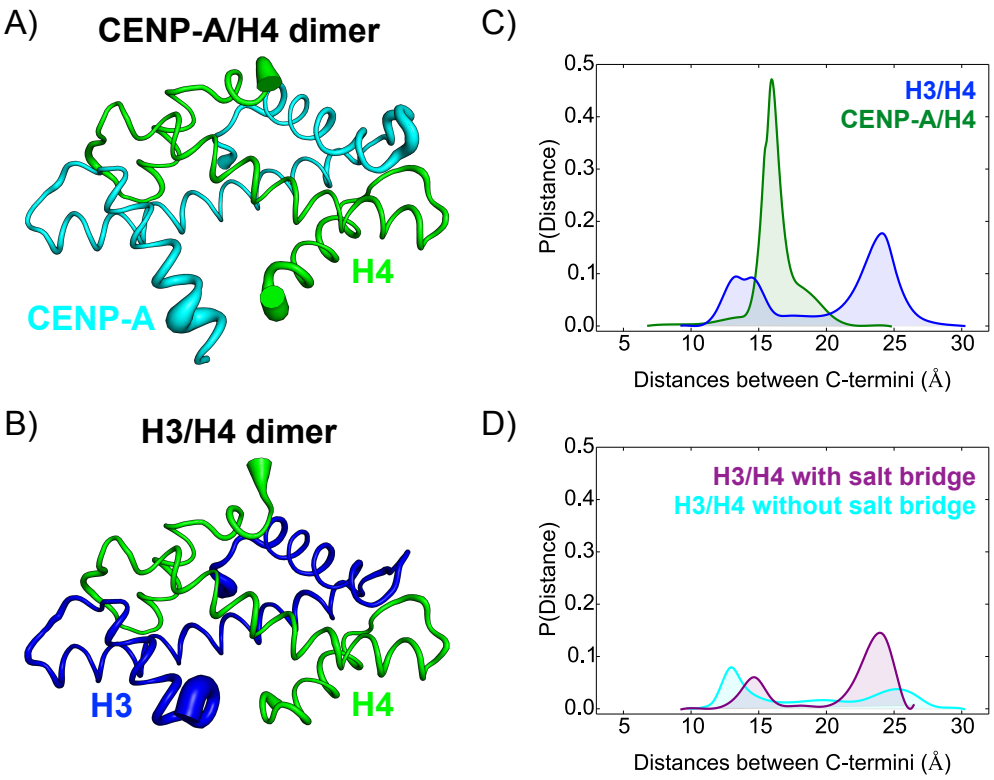


Figure 3.2: **More regions in CENP-A/H4 display local mobility than in H3/H4.** (A) CENP-A/H4 and (B) H3/H4 dimers represented by tubes, whose width corresponds to the B-factor, directly proportional to mean squared deviations, calculated in gromacs from computational trajectory. The CENP-A/H4 dimer exhibits significant local mobility in multiple different regions, while most of the H3/H4 local mobility is confined to the H3 C-terminus. (C) Probability distributions for the distances between the C-termini of CENP-A (or H3) and H4. (D) A specific salt bridge (H3 E133 to H4 R95) plays an important role in determining the distance between the C-termini of H3 and H4.

directly proportional to the mean squared fluctuation of each atom, and displayed dimer structures as tubes whose widths correspond to these values (Figure 3.2.A,B). CENP-A loop 1, a region near the C-terminus of CENP-A (residues 130-136), and the C-terminus of H4 exhibit increased local mobility in the CENP-A dimer (Figure 3.2.A). The H3 dimer is less locally mobile than the CENP-A dimer, except for the final several residues at the C-terminus of H3 (Figure 3.2.B). CENP-A $\alpha 3$ is several residues longer than the corresponding helix of H3 (Figure 3.1.A, dashed oval), and the $\alpha 3$ regions of CENP-A and H4 become partially unraveled in the absence of HJURP. However, the C-termini of CENP-A and H4 can interact with each other in spite of partial unwinding. We measured the distances between the C-termini of individual proteins of the CENP-A and H3 dimers to investigate further. The distance between the C-termini of CENP-A and H4 remains relatively consistent, however, the probability distribution of the distance between C-termini features two distinct peaks (Figure 3.2.C). We identified a specific salt-bridge (between H3 E133 and H4 R95) as important to the local mobility of H3/H4. When this salt-bridge is present, H3/H4 the H3 and H4 C-termini more likely to be farther apart, facing away from each other (Figure 3.2.D). On the other hand, the distance between the C-termini of CENP-A and H4 remains relatively short. Differences in the dynamics of the C-termini of CENP-A and H3 are partially due to electrostatics. The C-terminal tail of CENP-A (residues -LEEGLG) has an overall net charge of -2e, while the corresponding, shorter tail of H3 (residues -ERA) is neutral. HJURP is specific to CENP-A, in part, because the C-terminal tail of CENP-A is more acidic than the corresponding tail of H3. Overall, electrostatics are clearly important to

the direct interactions between HJURP and the CENP-A dimer.

3.3.3 HJURP plays a stabilizing role in CENP-A/H4

HJURP interacts directly with the CENP-A dimer, localizing CENP-A to the centromere and enabling CENP-A to become deposited into the centromeric nucleosome. To investigate the effects of HJURP on CENP-A dimer structure and dynamics, we mapped specific contacts between the C-termini of CENP-A and H4 (Figure 3.3.A,B). In the absence of HJURP, the $\alpha 3$ regions of CENP-A and H4 become partially unraveled and a contact forms between the oppositely charged H4 R95 and CENP-A E137 $\sim 40\%$ of the time (Figure 3.3.A). Upon the introduction of HJURP, an electrostatic network develops between the C-termini of CENP-A and H4 and the α domain of HJURP, wherein the $\alpha 3$ regions of CENP-A and H4 retain their helical structure, and the contact between H4 R95 and CENP-A E137 increases in likelihood to $\sim 70\%$ (Figure 3.3.B). Our results indicate that several charged residues – including HJURP R23, R26, CENP-A E136, E137, and H4 R95 – form contacts at the interface between the C-termini of CENP-A and H4 and the α domain of HJURP (Supplementary Figure B.3). In the absence of HJURP, the C-terminal end of CENP-A $\alpha 3$ partially unravels, preventing contacts between CENP-A R130 and H4 T96. Upon the introduction of HJURP, CENP-A $\alpha 3$ remains ordered, and contacts form between CENP-A R130 and H4 T96. Differences in specific contacts are important for characterizing the dynamics of CENP-A and H3 dimers, however, we must also compare CENP-A and H3 from a more global perspective. Sequence

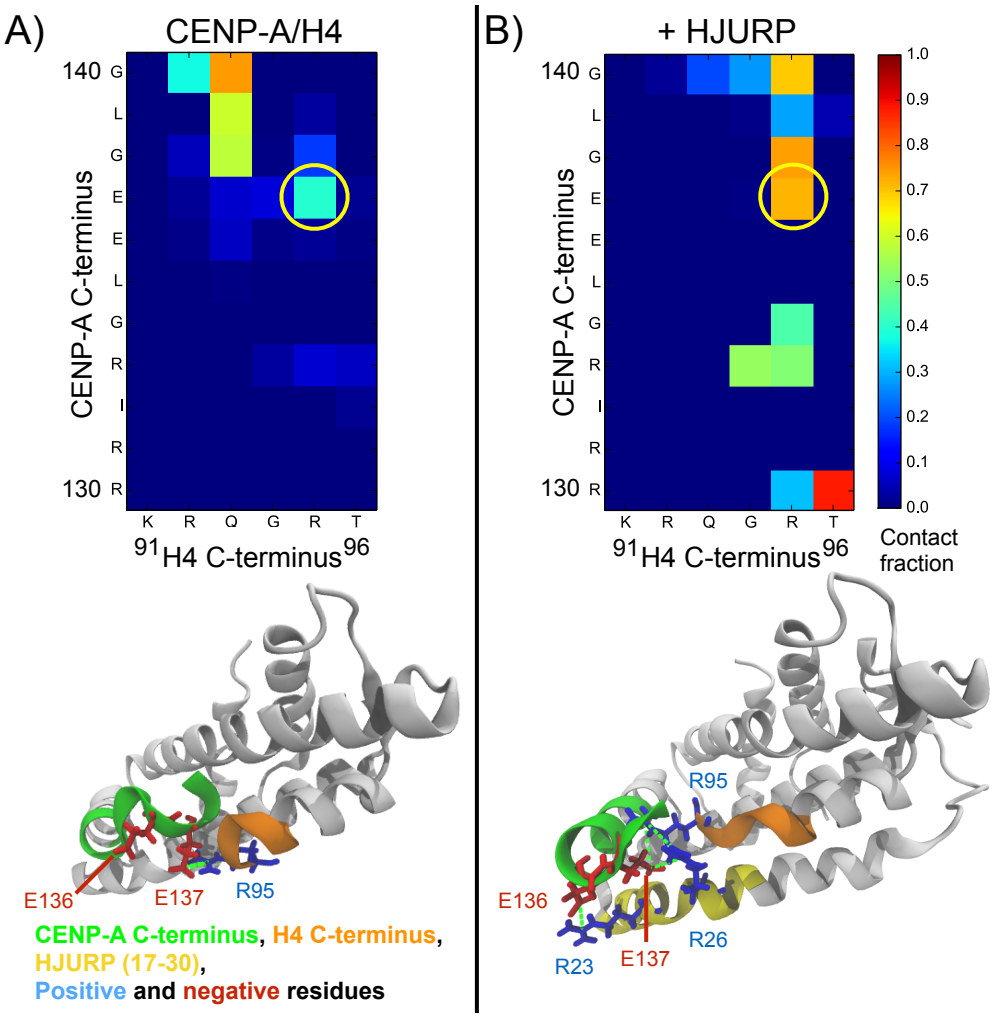


Figure 3.3: **HJURP stabilizes the CENP-A/H4 through electrostatic interactions.** Contact maps between the C-termini of CENP-A and H4, and representative simulation snapshots, in (A) the CENP-A/H4 dimer, and in (B) the CENP-A/H4 dimer in conjunction with HJURP, a CENP-A specific chaperone protein.

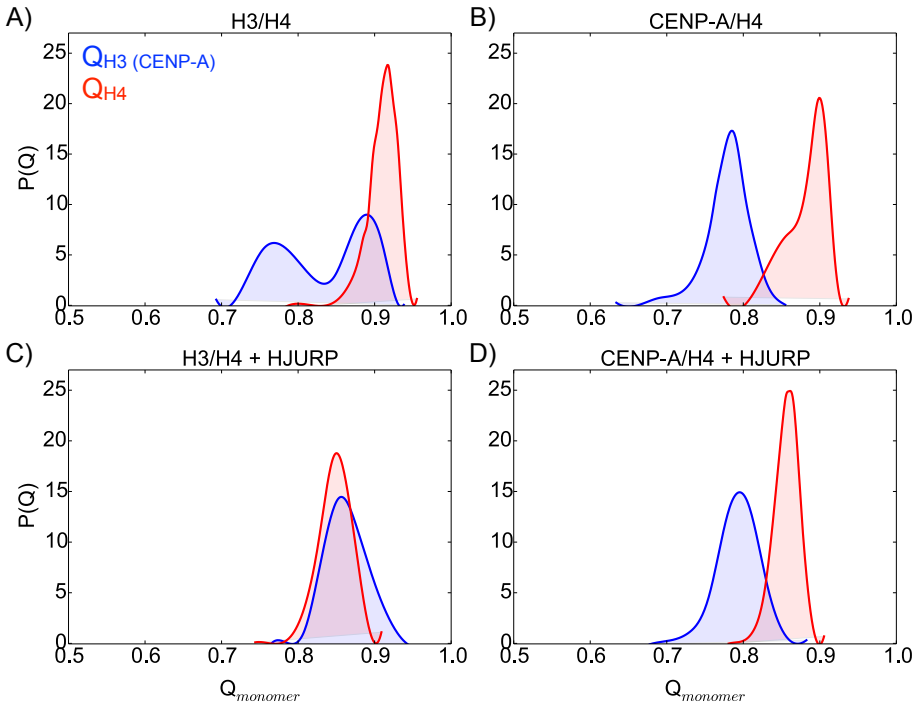


Figure 3.4: **H4 adopts conformations closer to the native state than CENP-A or H3.** $Q_{monomer}$ probability distributions for H3 or CENP-A and H4 for (A) the H3/H4 dimer, (B) CENP-A/H4, (C) H3/H4 in a complex with HJURP and (D) CENP-A/H4 in a complex with HJURP. Q is a normalized measure of structural similarity to the native state, greater values indicating stronger similarity. Native states were obtained from the experimentally resolved crystal structures. For each system the average monomer Q value for H4 is greater than the average for CENP-A or H3, except for H3/H4 in conjunction with HJURP.

analysis demonstrates that CENP-A and H3 are related proteins. Therefore, H4 could play a fundamentally different role than CENP-A or H3 in the context of an individual nucleosome and in the formation of higher-order chromatin structure.

3.3.4 H4 adopts more native-like conformations than CENP-A or H3

We compared H4 to CENP-A and H3 globally by calculating the Q values, a measure of structural similarity, of the histone monomers with respect to their corresponding experimentally determined structures. Q is a normalized order parameter, with higher values indicating greater similarity between two structures (see Methods). For all the systems studied, except for H3/H4 in conjunction with HJURP, the average monomer Q value for H4 (Q_{H4}) is greater than the average for CENP-A, or H3 ($Q_{CENP-A(H3)}$) (Figure 3.4). The H4 proteins explore configurations closer to the native state than either CENP-A or H3 except for in the complex of H3/H4 in conjunction with HJURP (Figure 3.4.C), which is not observed *in vivo*. This is consistent with our biological understanding, because variants exist for all of the core histones except H4. Moving vertically down the figure, we can assess the effects of HJURP binding on structural similarity with respect to the native state. Upon the introduction of HJURP, H4 adopts less native conformations, and CENP-A or H3 adopts more native conformations (Figure 3.4). The H3 family, including canonical H3 and the centromere-specific CENP-A, could provide structural variability while H4 provides greater consistency.

3.4 Conclusion

Our analysis of all-atom MD simulations of CENP-A and H3 dimers revealed that CENP-A/H4 is more structurally plastic than H3/H4 on a global level, where more regions in CENP-A/H4 display enhanced local mobility than in H3/H4. We

demonstrated that HJURP plays a stabilizing role in CENP-A/H4 by forming an electrostatic network of interactions with the C-termini helices of CENP-A and H4, which become partially unwound in the absence of HJURP. We also found H4 adopts more native-like conformations than CENP-A or H3 in all the systems studied, except for H3/H4 in conjunction with HJURP.

We illustrated that the replacement of canonical H3 with CENP-A translates to increased plasticity in histone dimer dynamics, and that HJURP plays a stabilizing role for CENP-A/H4 but not for H3/H4. These effects of CENP-A are due, in part, to its longer C-terminal tail, which maintains helical structural integrity when in a complex with HJURP. The increased acidity of CENP-A C-terminus is also important for electrostatic interactions between CENP-A and H4, compared to between H3 and H4. Furthermore, the higher negative charge of this region compared to H3 could contribute to HJURP's specificity to CENP-A [63].

We also considered the differences between the H4 protein and the H3 family from a more global view. H4 could adopt conformations closer to the native state than CENP-A or H3 because H4 plays a different fundamental role in chromatin organization. H4 and CENP-A or H3 associate with H2B and another copy of CENP-A or H3 through four-helix bundles, respectively, within the octameric protein core of a nucleosome. Therefore, the primary interactions of CENP-A or H3 beyond the dimer level are closer to self-association, while H4 must interact with another family of proteins. This pattern could hold for the interactions between nucleosomes as well. CENP-A and H3 could mainly form interactions within one nucleosome while H4 must form interactions between nucleosomes. Indeed, the highly basic H4 N-

terminal tail directly interacts with the acidic patch of an adjacent nucleosome [1]. Overall, the different characteristic dynamics of the CENP-A and H3 dimers could translate to different dynamics of individual nucleosomes and nucleosomal arrays where CENP-A replaces canonical H3 *in vitro* and *in vivo*.

3.5 Summary and Biological Implications

In summary, histone dimers containing the centromere-specific variant CENP-A are more structurally flexible than canonical H3 dimers on both global and local levels. Furthermore, the chaperone HJURP plays an essential role in stabilizing CENP-A/H4 through electrostatic interactions. In general, H4 adopts conformations closer to the native state than CENP-A or H3, meaning that the H4 histone could play a critical role in the interactions between all variations of nucleosomes and in the formation of higher order chromatin structure. Overall, we find that CENP-A confers greater flexibility to an individual histone dimer, and that HJURP is necessary for structurally stabilizing the CENP-A dimer.

Chapter 4: The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects.

Results presented in this chapter have been published by Winogradoff, Echeverria, Potoyan and Papoian [43].

4.1 Introduction

Many proteins do not form well defined three dimensional structures in cells of higher organisms, yet they are biologically active and involved in a variety of biological processes [96–104]. These intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are characterized *via* heterogeneous ensembles of chain conformations, where the potentially complex conformational landscapes are regulated by tuning both the non-specific interactions and the overall chain entropy, as well as specific inter-residue interactions [39, 100]. Disordered proteins play a key role in signaling and transcription regulation by interacting with each other or with more structured proteins. To accomplish their activities, some IDPs or IDRs undergo disorder-to-order transitions or bind to their biological partners

by conformational selection [105–111], though some are known to function without ever becoming structured [112–114]. Furthermore, IDPs and IDRs are richly regulated by a combinatorial variety of post-translational modifications (PTMs) that can significantly change their conformational and binding preferences [115,116]. The structural flexibility and conformational heterogeneity of IDPs and IDRs are considered advantageous, and even essential to the biological complexity found in higher organisms, where the potential for highly intricate biological regulation emerges from the sophisticated and flexible interaction networks formed by the inclusion of these proteins [104].

Histone tails, the terminal segments of histone proteins, are key IDRs that regulate the structure and dynamics of the genomic DNA-protein fibers, called chromatin, where the latter are central to many template directed processes, including DNA replication, repair and transcription [117–119]. Histone tails are highly flexible, highly positively charged, low in hydrophobicity and feature multiple sites for potential PTMs, such as acetylation, methylation and phosphorylation [120]. It was initially thought that histone tails acted mainly as unstructured electrostatic mediators, and that PTMs simply function as modulators of these interactions. This implies, for example, that lysine acetylation, by neutralizing the positive charge of the lysine amino, reduced the electrostatic interactions between the histones and the DNA phosphates, making the DNA more accessible for active processes such as transcription. However, the realization that PTMs are highly diverse, acting individually or in various combinations, led to the hypothesis that the PTMs could form the *histone-code* where highly specific PTM combinations specify different chromatin

states [118, 121, 122].

The combinatorial effect of different PTMs of histone tails can be interpreted by the distinct, yet overlapping, *direct* and *effector mediated* mechanisms. In the *direct* mechanism, histone tails interact with the neighboring nucleosome, such that these interactions regulate inter-nucleosomal structure (e.g. histone-tail bridging). For example, histone tails are known to participate in intra- and inter-nucleosome interactions with both protein and DNA in condensed chromatin structures [35, 123]. In contrast, the *effector mediated* mechanism postulates that PTMs serve as recognition sites for macromolecular complexes involved in chromatin remodeling activities [120, 124, 125], which, in turn, can alter the chromatin architecture. A salient example of these effectors are bromodomains, a large family of proteins which recognize acetyl-lysine motifs, found in chromatin remodeling complexes [120, 126].

The H4 histone tail (Figure 4.1) has been identified to be of great importance for chromatin structure formation and stability. For example, early *in vitro* experiments of nucleosomal arrays determined that histone tails are necessary for the stability of higher order chromatin structure [31–33], and, furthermore, that the acetylation of histone tails plays a key role in regulating chromatin structure by, for example, disrupting the formation of the 30-nm fiber [127–129]. In a landmark *in vitro* study, the homogenous mono-acetylation of the H4 histone tail at lysine 16 alone was enough to inhibit the formation of higher order chromatin structures and impede the interactions between chromatin and non-histone proteins [34]. *In vitro* analysis of the cation-induced nucleosome-nucleosome association determined that the nucleosome stacking is mainly governed by electrostatics interactions that

are modulated by ion-ion correlations and histone-tail bridging [38]. The latter is mediated by a region of high charge density of the H4 tail, which includes lysine 16 (Figure 4.1), that interacts with the acidic patch on the H2A-H2B dimer of an adjacent nucleosome [1, 35–37]. Consequently, acetylation of lysine 16 directly alters these interactions, disrupting the nucleosome-nucleosome stacking [38]. Besides the charge reduction associated with acetylation, recent studies have found that the mono-acetylation of lysine 16 induces the partial ordering of the H4 tail, increases the affinity to DNA [41], and is associated with the formation of transient elements of secondary structure [39–42, 130].

The unique role of lysine 16 acetylation has been further confirmed by systematic genetic studies which show that single lysine H4 mutations do not result in defects in chromatin assembly or DNA replication, with the exception of the single mutation of lysine 16. Similarly, the mutation of all four H4 tail lysines is lethal, but not triple mutations. These studies suggest that the different H4 tail lysines are partially redundant and that acetylation of the histone H4 tails may be mediated through two distinct mechanisms: a cumulative and non-specific effect for lysines 5, 8 and 12 and a specific mechanism for lysine 16 [131, 132].

Our understanding of the effects of different levels of H4 acetylation has been obtained, mainly, from functional [34, 133], biochemical [134], and genetic analysis [131, 132], especially for the proteins that recognize these modifications [120], as well as from the conformational effects in chromatin. The *histone-code* hypothesis suggests that PTMs, such as lysine acetylations, may yield rich combinatorial outputs. However, it is unclear how different levels of acetylation change the con-

formational preferences of the H4 tail and if these effects are combinatorial or cumulative, in particular from a structural viewpoint. Further understanding of how acetylation changes the conformational landscape of the H4 tails would allow us to understand, for example, the role of *conformational selection* or *induced fit* mechanisms in H4 tail recognition. That is, binding-competent conformations, which are well-visited already in the unbound state are selected (conformational selection) or binding partner induces the required binding conformations (induced fit). For example, recent experimental and computational studies illustrate that histone tails display transient elements of secondary structure [39, 41, 135, 136], suggesting that histone tails could participate in specific interactions that tune chromatin structure. In this context, PTMs would add another important layer of control by regulating the histone tail conformations.

Here, we explore the effects of different levels of acetylation on the conformational preferences of an isolated H4 histone tail, and how these preferences are modulated by different levels of acetylation. Using all-atom replica exchange molecular dynamics (REMD) simulations in explicit solvent, we determined how various combinatorial acetylation patterns affect the peptide's conformational landscapes. We found that progressive acetylation reduces the conformational heterogeneity of the sampled states, altering both short and long-range interactions. Increased acetylation results in greater helical propensities and hydrogen bond occupancies without significantly changing the overall radii of gyration. These cumulative effects of acetylation highlight how the charge reduction and increased hydrophobicity associated with adding the acetyl groups enhance cohesive interactions within the peptide. Our

results also show that the sole acetylation of lysine 16 has structural effects that are unique to this case, including a significant reduction in the number of states sampled and the formation of a specific 3_{10} helix corresponding generally to more elongated structures and specifically to a unique positioning of the lysine 16 residue. At the microscopic level, the specific effects of the sole acetylation of lysine 16 include effectively rigidifying the peptide, setting an entropic constraint on the accessible conformations and leading to the formation of elongated structures. Results presented here provide, therefore, a structural characterization baseline to understand the effects of the histone cores and/or nucleosomal DNA on the conformations of the H4 tail. These interactions might have important implications under physiological conditions.

4.2 Methods

4.2.1 Molecular dynamics simulations.

We performed all-atom molecular dynamics (MD) using the Amber12 MD software [137], the amber99SB* [48] force field for proteins, the ions94 [51] force field for ions, and the TIP3P water model. Starting from the wild type (H4-WT) or unacetylated H4 N-terminal histone tail model from an earlier work [39], we used the *xleap* tool in AmberTools12 to prepare WT, mono-, di-, tri-, and tetra-acetylated H4 N-terminal histone tails. We modeled two mono-acetylated tails with acetyations at lysine 16 (H4-K16$_{ac}$) and lysine 5 (H4-K5$_{ac}$); for di-acetylated, we acetylated lysines 8 and 16 (H4-K8$_{ac}$K16$_{ac}$) and lysines 5 and 8 (H4-K5$_{ac}$K8$_{ac}$); for tri-acetylated,

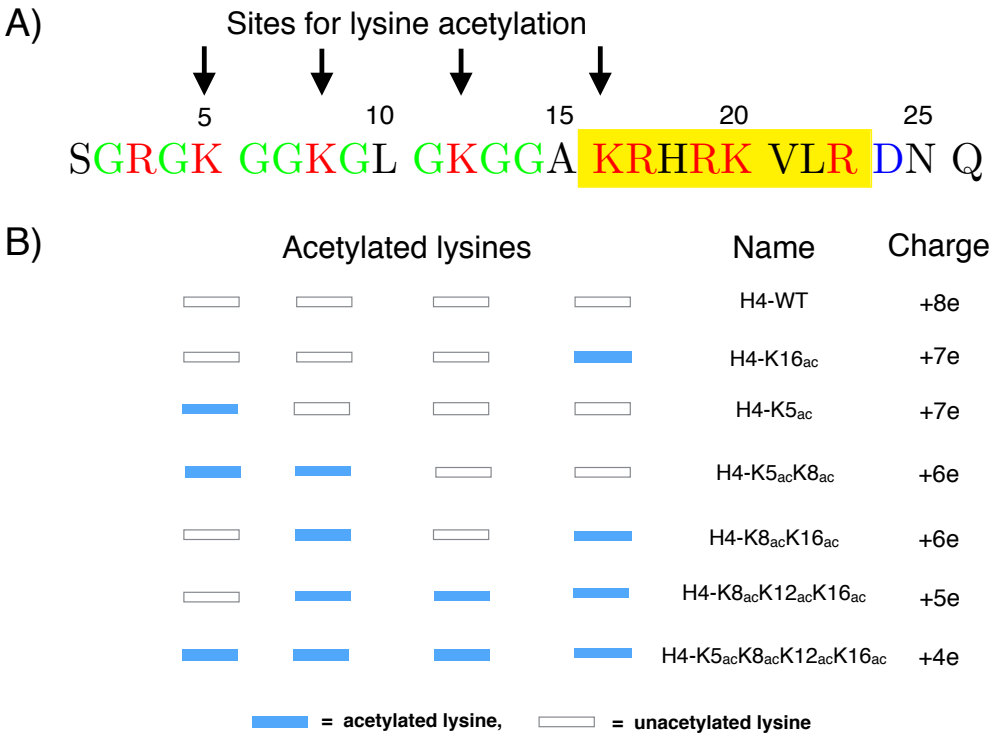


Figure 4.1: **The H4 N-terminal histone tail sequence and acetylation sites.** (A) The H4 histone tail sequence, including residue numbers, possible sites for acetylation, and residue types: glysines (green), positive residues (red), and negative residues (blue). A region of high positive charge density, residues 16-23 (also known as the basic patch), is highlighted in yellow. (B) Studied models with different levels of acetylation. Solid blue boxes highlight the specific sites of lysine acetylation. The given name and net charge for each of the studied models is provided as well.

we acetylated lysines 8, 12, and 16 (H4-K8$_{ac}$K12$_{ac}$K16$_{ac}$); for tetra-acetylated, we acetylated lysines 5, 8, 12, and 16 (H4-K5$_{ac}$K8$_{ac}$K12$_{ac}$K16$_{ac}$), every possible site. We did not consider lysine 20 given that it is mostly found methylated [138]. We determined the specific sites for lysine acetylation by their physiological abundance [139]. For further details about convergence tests, see Supplementary Figure C.1.

To characterize the conformational ensemble of the H4 tail at different levels of acetylation, we performed replica exchange molecular dynamic (REMD) simulations [140]. Exchanges between replicas at different temperatures enhances the conformational sampling relative to standard MD simulation, creating an ensemble that includes both high and low energy configurations. First, each system was copied to generate a total of ~ 60 replicas. The temperatures used in REMD simulations, ranging from 300 K to 450 K, were determined by T-REMD [141], an REMD temperature online server, with a target exchange probability of 30%. Then, each replica was heated to the desired temperature over 500 ps in the NVT ensemble. REMD production runs were performed in the NVT ensemble, attempting exchanges every 5 ps with a 2 fs time-step, saving coordinates and energies every picosecond for further analysis. 100 ns of REMD simulations were performed for each system using the Langevin thermostat with a 2-ps time constant, totaling 6 μs of simulation each. The exchange probability observed for each system was $\sim 50\%$. For analysis we only considered only the final 90 ns of trajectories set to 300 K. This allowed us to account for further thermal equilibration.

4.2.2 Analysis of the trajectories.

We determined the radius of gyration (R_g) along each trajectory for all the studied systems. The obtained values were first compared to the prediction for a globular protein of the same length ($R_{g,\text{globular}} = 2.2N^{0.38}$), a relation based on a power law best fit of R_g as a function of sequence length for a subset of proteins in the PDB [142]. Similarly, the R_g values were compared to the prediction for a thermally denatured random coil of the same length ($R_{g,\text{denatured}} = 2.02N^{0.60}$), a relation proposed by Flory's theory and confirmed by power fitting R_g values determined by computation and experiment [143].

The secondary structure present for each simulation snapshot was determined with the Amber 12 *secstruct* tool, which uses the DSSP program [144] to identify hydrogen bond motifs through backbone amide (N-H) and carbonyl (C=O) atom positions. By definition, a 3_{10} helix spans at least three consecutive residues requiring two hydrogen bonds between residues (i, i+3), and an α-helix spans at least four consecutive residues requiring two hydrogen bonds between residues (i, i+4). For each residue, we determined the percentage of simulation snapshots where the residue is part of a 3_{10} or α-helix, which we will refer to as the *helix propensity* per residue. Also, we identified all of the protein's hydrogen bonds. A geometric definition of a hydrogen bond was used: two heavy atoms are considered to be bonded if (1) their donor-acceptor distance is less than 3.5 Å, and (2) the acceptor-donor-hydrogen angle is less than 30°. Furthermore, we analyzed inter-residue contact preferences. We identified contacts between residues, excluding (i, i $\pm$ 1) residue

pairs, for all levels of acetylation. A contact was determined to exist when the distance between two non-hydrogen atoms from different residues was less than 3.6 Å. Contacts were first calculated as percentages of their respective entire trajectories. Then, we divided the H4 tail into five segments: residues 1-5, residues 6-10, residues 11-15, residues 16-23 (the aforementioned basic patch) and residues 24-26. Then, contact pairs between, and within, segments were considered together as sums. We present contacts between and within segments for all acetylated tails as ratios relative to the corresponding inter-segmental contact sums for the unacetylated WT H4 tail. Lastly, we used Ramachandran plots of key residues to graphically display specific backbone dihedral angle preferences.

4.2.3 Clustering Analysis.

We performed clustering analysis to characterize the conformational ensemble sampled during the REMD simulations. For this purpose we defined the dissimilarity metric as the pairwise RMSD, after proper alignment, between the backbone atoms of simulation snapshots selected every 1 ps (N $\sim$ 88,000 structures per trajectory). Following a bottom-up approach [145], clustering was performed as follows: we computed the RMSD between structures i and j, if the RMSD was smaller that a given cutoff ($\text{RMSD}_{\text{cutoff}}$), the structure j was added to cluster i and the center of cluster i was defined as the average between structures i and j.

Following, we computed the RMSD between the average structures of clusters i and $j+1$, and merged the clusters if the RMSD $<\text{RMSD}_{\text{cutoff}}$. Conversely, if RMSD

$>\text{RMSD}_{\text{cutoff}}$, $j+1$ is defined as its own cluster. We repeated this procedure until we compared the structure $i = N$ to all clusters. In cases where a structure could be assigned to two or more distinct clusters, we assigned it to the cluster to which its RMSD with respect to the cluster's center was the smallest. The $\text{RMSD}_{\text{cutoff}}$ was set to 2.8 Å.

After the first round of clustering, we re-defined the center of each cluster by identifying the reference structure. To determine the reference structure, we computed the average position of the backbone atoms over all the simulation snapshots in the cluster, and then identified the structure with the lowest RMSD with respect to the average backbone. A second round of clustering followed considering the obtained reference structures. We found that after two rounds of clustering we obtained a converged set of clusters. Among the advantages of this clustering approach is that the $\text{RMSD}_{\text{cutoff}}$ is the only free parameter and that it is not necessary to define the desired number of clusters beforehand. The distance between the clusters was defined as the RMSD between the representative structures (i.e. structures at the center of the cluster).

To visualize the results obtained from the clustering analysis, we computed the principal components of the dissimilarity matrix $\mathbf{R} = R_{ij}$, where R_{ij} is the pairwise RMSD between the reference structures of clusters i and j. By computing the first two principal components, we obtained a set of points in two-dimensional space such that the distances between the points are approximately equal to their dissimilarities (Figure 4.2). The size of each cluster is proportional to the number of structures in each cluster.

To assess the heterogeneity of conformational ensemble sampled by the WT and acetylated H4 histone tails, we determined the cumulative percentage of structures represented in the clusters, after ordering clusters by size. Additionally, we determined the distances of all clusters to the center-of-mass. The center-of-mass was determined as $COM = \sum_{i=1}^{M} S_i X_i / M$, where S_i is the size of cluster i, X_i is the position of clusters i, defined by the principal components, and M is the number of clusters. We analyzed the representative structures of the most populated clusters to determine the stabilizing interactions and the molecular determinants of prominent structural features.

4.3 Results and Discussion

4.3.1 Acetylation of the H4 tail reduces the conformational heterogeneity of the sampled ensemble.

To elucidate the effects of acetylation on the conformational preferences of the H4 histone tail, we performed conformational clustering analysis on each of the studied models. This analysis allowed us to represent the peptides' conformational space by a set of discrete microstates, one corresponding to each cluster, where the most prominent structures are identified as those belonging to the largest clusters. Figure 4.2 illustrates that the acetylation of the H4 tails has a significant effect on the number of sampled microstates and their dissimilarity. For example, the WT H4 explores a sparse set of conformations, which are shown as small clusters (i.e. with only few simulation snapshots per cluster) that are far apart structurally from each

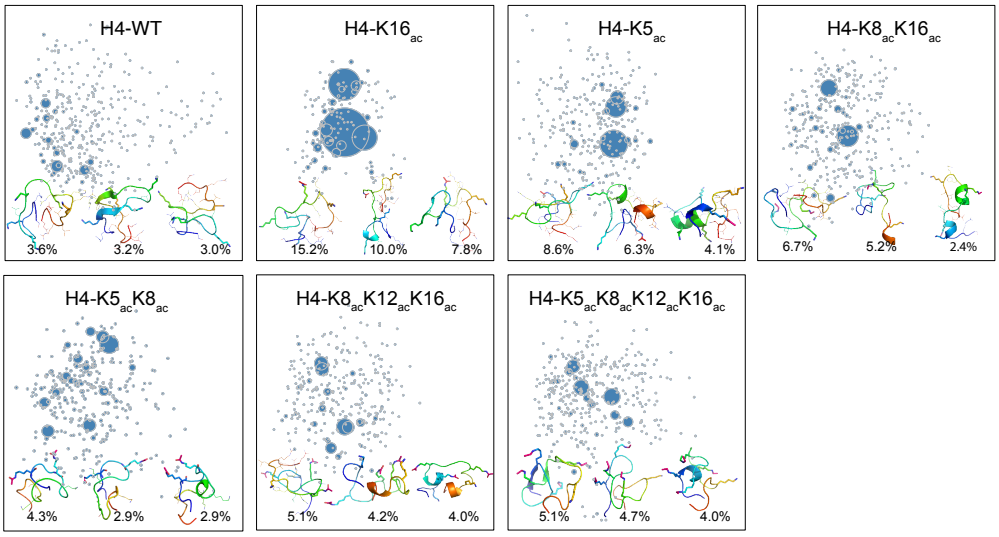


Figure 4.2: **Conformational clustering analysis of each of the H4 models.** In each panel we present the results of the clustering analysis. Each cluster is represented by a point, with its size proportional to the number of trajectory frames in it. The distances between points are approximately equal to their dissimilarities. The three most representative structures (i.e. largest clusters) are shown in each panel along with the percentage of the sampled ensemble that that cluster represents. Structures are shown colored in rainbow mode from blue (N-terminal) to red (C-terminal), and with elements of secondary structure shown in cartoon representation. Lysines are shown in sticks representation.

other, indicating high dissimilarity. In contrast, the sole acetylation of K16 has a dramatic effect on the conformational preferences of the peptide, significantly reducing the number of clusters. For H4-K16$_{ac}$ we identified three prominent structures which represent 15%, 10% and 8% of all the sampled structures.

To explore how acetylation affects the heterogeneity of the conformational ensemble, we determined the cumulative number of clusters necessary to account for a fraction of the sampled structures (Figure 4.3.A). For example, the largest cluster identified for the WT represents 3.5% of all the sampled structures. Anal-

ogously, for the acetylated H4-K16$_{ac}$, H4-K5$_{ac}$, H4-K8$_{ac}$K16$_{ac}$, H4-K5$_{ac}$K8$_{ac}$, H4-K8$_{ac}$K12$_{ac}$K16$_{ac}$ and H4-K5$_{ac}$K8$_{ac}$K12$_{ac}$K16$_{ac}$ models, the largest clusters represent 15%, 8.6%, 6.7%, 4.3%, 5% and 5% of all the sampled structures, respectively. In general, the mono-acetylation of K16 has the largest effect in changing the conformational preferences of H4, as evidenced by the presence of large clusters. The effects of di-, tri- and tetra-acetylations appear to, in part, to counteract this effect. However, for all levels of acetylation, we observe that some conformations are favored. We will discuss below how these conformational preferences may play a role in the context of binding and recognition of histone tails.

We further characterized the heterogeneity of the conformational ensemble by measuring the average distance from all clusters to the center of mass (see Methods section). Using this metric, we determined that the WT exhibits the most heterogeneous ensemble while H4-K16$_{ac}$ has the least heterogeneous ensemble. For all other models, we observed the general trend of decreasing structural heterogeneity with increasing the extent of acetylation (Figure 4.3.B). These results again indicate that the mono-acetylation of K16 stands out from the overall trend, crucially altering the conformational landscape of the H4 histone tail. An alternative approach to quantify the conformational heterogeneity is to determine the distribution of the pairwise RMSD or Q between all sampled structure, as discussed in the Supporting Information (Supplementary Figure C.2) [41, 146]. Below, we further elaborate on the way the various levels of acetylation structurally affect the the conformational preferences of the tails.

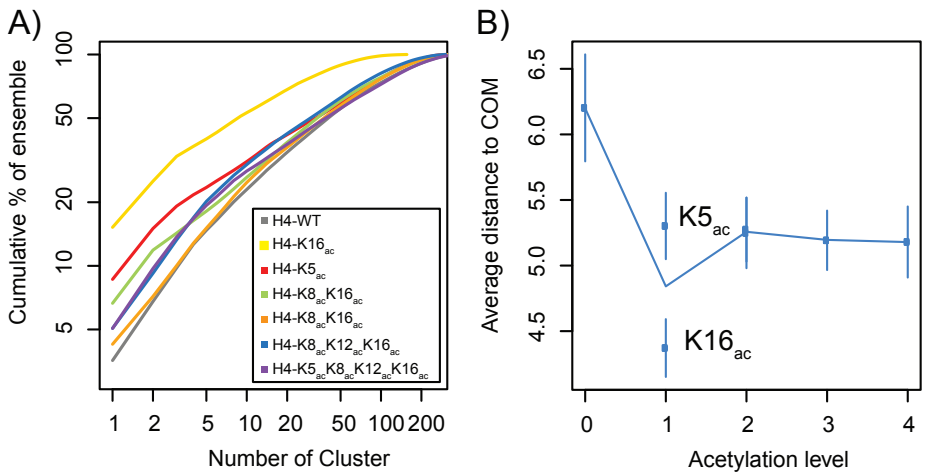


Figure 4.3: **Heterogeneity of the conformational space of WT and acetylated H4 tails.** (A) Analysis of the number of clusters necessary to account for the sampled conformations, after ordering clusters by size. (B) Average distance of all clusters to the center-of-mass.

4.3.2 Acetylation of K16, but not other lysines, leads to more extended conformations.

We used the radius of gyration (R_g) as a coarse metric to assess the type of conformations sampled by the different models. For all levels of acetylation, the average radii of gyration of the H4 histone tail are between the predicted values for globular (7.6 Å) and thermally denatured random coil (14.3 Å) proteins of the same length, indicating that the H4 tail adopts molten globule type conformations (Figure 4.4.A). Furthermore, the average R_g values for all the systems studied (Supplementary Table C.1) are closer to the predicted value for globular proteins than to the prediction for a thermally denatured random coil, suggesting that the H4 tail could include elements of secondary structure. However, the H4-K16$_{ac}$ tail ex-

hibits a slightly greater average R_g, and its R_g probability distribution features two distinct peaks, as opposed to one. We observe that, to a lesser extent, the H4-K8$_{ac}$K16$_{ac}$ model also has a bimodal R_g distribution. These results indicate that, even though the H4 tails are highly charged peptides, the global dimensions are not strictly dictated by the repulsive electrostatic interactions [147].

4.3.3 Acetylation of H4 tails induces increased helical propensities.

To better understand the preferences of the H4 tail to form secondary structure, we investigated the helical propensities, per residue, at different levels of acetylation. Our results demonstrate that the WT H4 histone tail has a small amount of α and 3_{10} helical propensity spanning the entire sequence, and, furthermore, that the helical propensity spanning the whole sequence increases with acetylation, leveling off beyond di-acetylation (Figure 4.5 and Supplementary Table C.1). The profile of helical propensity for the H4 tail mono-acetylated at lysine 16 (H4-K16$_{ac}$), however, is unique; upon K16 mono-acetylation, the H4 histone tail undergoes a significant structural rearrangement whereby the helical propensity becomes highly localized, featuring a specific 3_{10} helix from residues 7-9, which is formed $\sim$30% of the time. The formation of this helix is intermittent throughout the simulation (Supplementary Figure C.3). This individual helix is the single most prominent element of secondary structure among all levels of acetylation (Figure 4.5). In contrast to the mono-acetylation of K16, the H4 tail mono-acetylated at lysine 5 (H4-K5$_{ac}$) induces a slight increase in the 3_{10} and α helical propensities relative to the WT tail that

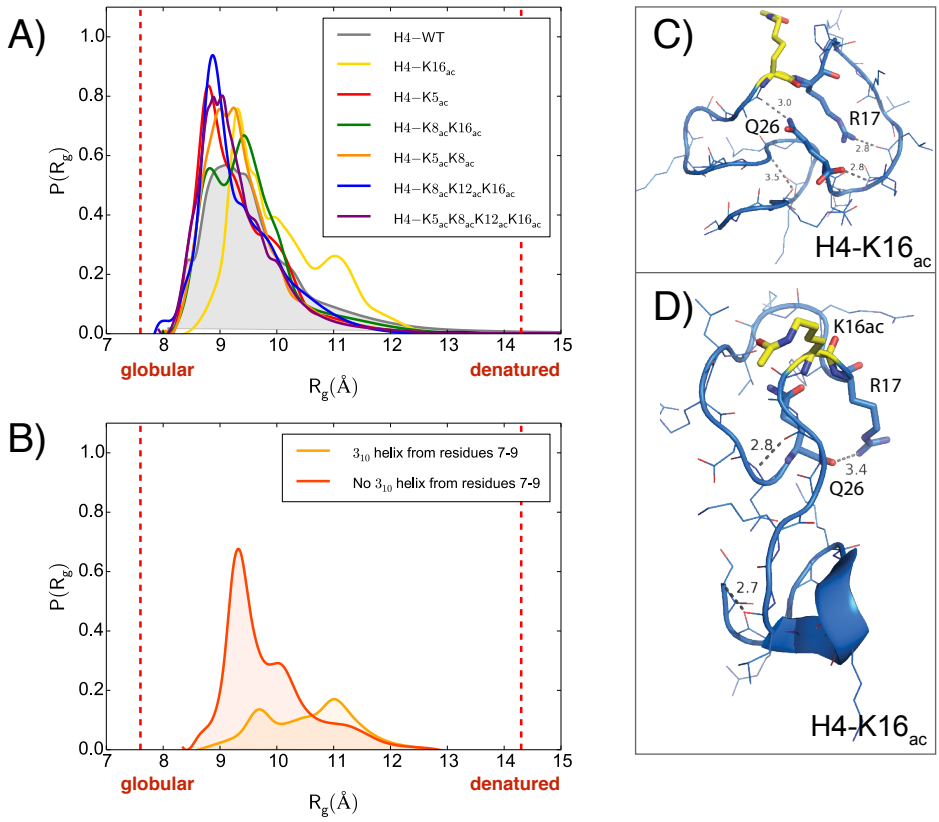


Figure 4.4: **Probability distributions of radius of gyration R_g.** (A) R_g at different levels of acetylation. The vertical red lines represent the predicted R_g values for globular (7.6 Å) and thermally denatured random coil (14.3 Å) peptides of the same length ($N = 26$ residues). (B) R_g of the H4 tail mono-acetylated at K16 divided into two groups: (1) simulation frames with a 3_{10} helix from residues 7-9, and (2) simulation frames without this specific helix. (C) Characteristic structure obtained from the most populated cluster of the H4-K16$_{ac}$ showing the stabilizing interactions between R17 and Q26 with the backbone. The acetylated K16 is shown in yellow. (D) Sample structure from the second most populated cluster exhibiting the H4-K16$_{ac}$ characteristic 3_{10} helix. This structure is stabilized mainly by backbone-backbone hydrogen bonds. The acetylated K16 is shown in yellow.

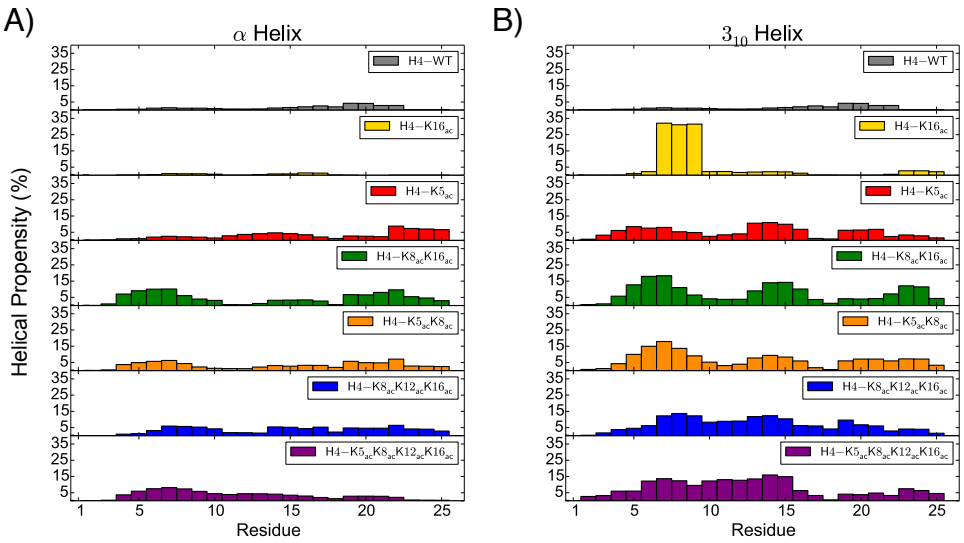


Figure 4.5: **Helical propensity per residue.** (A) α-helical propensity per residue as a fraction of time, across all levels of acetylation. (B) 3_{10} helical propensity per residue.

continues to span the entire sequence, the average helical propensity per residue increasing from ∼5% for the H4-WT tail to ∼7% for the H4-K5$_{ac}$ tail (Supplementary Table C.1). The significant difference between the structural effects of K16 mono-acetylation and K5 mono-acetylation could be explained, in part, by the location of these specific residues in the sequence of the H4 histone tail; K16 is found within a region of high charge density close to the C-terminus of the tail, and, in contrast, K5 is located in a region of relatively low charge density close to the N-terminus (Figure 4.1). Di-, tri-, and tetra-acetylated H4 tails exhibit a greater increase in the 3_{10} and α helical propensities than K5, spanning the entire sequence, further increasing the average helical propensities per residue to ∼10% (Supplementary Table C.1, Supplementary Figure C.4). Furthermore, the two patterns of di-acetylation

have similar effects on the helical propensity of the H4 tail (Figure 4.5, Supplementary Table C.1, Supplementary Figure C.4). For all levels of acetylation, the helical propensity is primarily 3_{10} helix, and the majority of helices formed are short, either three or four residues long. The overarching trend of helical propensity increasing with acetylation (Supplementary Figure C.4) is consistent with a previous circular dichroism (CD) experiment, which reported the α helical content of the H4 tail increases monotonically with progressive acetylation [136], and several recent computational MD studies performed with explicit solvent [39–42], but not with a computational study with implicit solvent [130]. However, quantitative comparison with experiments is currently not feasible, given that helicities were experimentally measured in the context of the nucleosomal core particles. Histone cores and nucleosomal DNA can alter the structural preferences of the H4 histone tail, and may contribute to the higher helicities observed by Wang *et al.* [136] (Supplementary Figure C.4).

The unique effect of the sole acetylation of K16 on helical propensity is consistent with our clustering analysis, where an H4-K16$_{ac}$ structure with a 3_{10} helix from residues 7-9 is representative of the second most populated cluster (Figure 4.2 and 4.4.C). The presence of the 3_{10} helix from residues 7-9 also explains why the H4-K16$_{ac}$ model exhibits a higher R_g. For example, by dividing the H4-K16$_{ac}$ sampled structures into two groups, one with the 3_{10} helix from residues 7-9 and the other without this helix, we discovered that structures characterized by the presence of this specific helix correspond to the more extended conformations (Figure 4.4.C) and contribute to the second peak in the probability distribution (Figure 4.4.B).

By performing a similar division of the H4-K8$_{ac}$K16$_{ac}$ sampled structures based on helical structure, we found specific 3_{10} helices contribute to slightly more extended conformations as well (Supplementary Figure C.6).

4.3.4 Acetylation of H4 tails increases the long-range contact occupancies.

The general features of conformational ensembles depend, in part, on the formation of secondary structure and other stabilizing interactions, such as hydrogen bonds, hydrophobic contacts and salt-bridges. As already described, acetylation is associated with an increase in helix propensities (Figure 4.5). To explore the role of non-local hydrogen bonds, we characterized the hydrogen bond occupancies of the peptide side chains. We found that acetylated lysines have a higher probability of participating in hydrogen bonds (Figure 4.6.A). These hydrogen bonds include mainly lysine-backbone contacts (Figure 4.6.B). Furthermore, by analyzing the hydrogen occupancies of other residues in the peptide we determined that the increased hydrogen-bond occupancies is specific to acetylated-lysines, and that hydrogen bonds occupancies of other side-chain do not increase with acetylation level (Supplementary Figure C.7). We also determined that salt bridges play a prime role in stabilizing the H4 tail conformations for all models, except H4-K16$_{ac}$ (Supplementary Figure C.8).

To further understand the role of acetylation in the conformational ensemble we analyzed the inter-residue contact preferences. By dividing the peptide in five

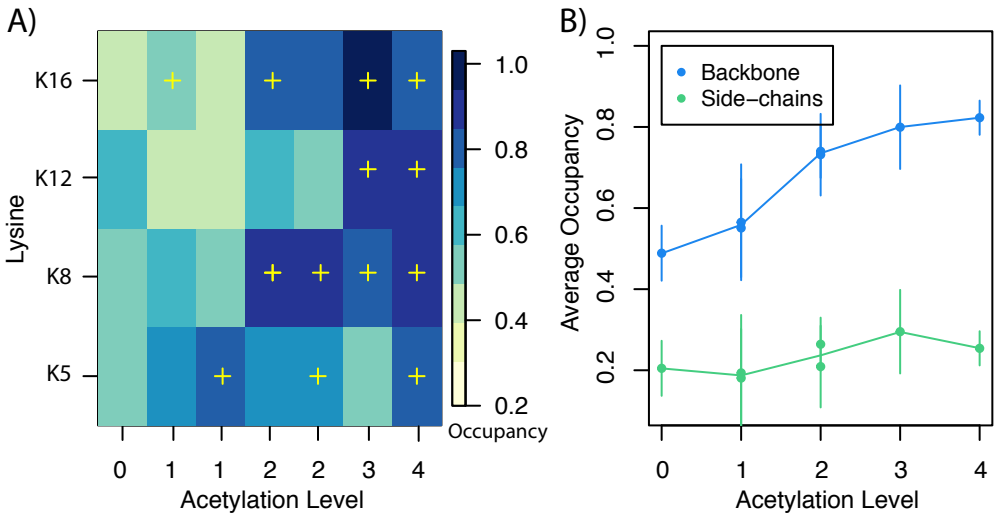


Figure 4.6: **Hydrogen-bond occupancies of lysines.** (A) Hydrogen bond occupancies measured as the percentage of time that every lysine is forming a hydrogen bond. Yellow crosses indicate acetylated lysines. (B) Average hydrogen bond occupancies at different levels of acetylation. Hydrogen bonds have been divided between lysine-backbone (blue) and lysine-side-chain (green).

segments, one of which is the basic patch, we computed the relative segment contact occupancies with respect to the WT. Figure 4.7 shows that the sole acetylation of K16 disrupts the contacts (i.e. lower occupancies) between the basic patch and the first half of the peptide, which is in agreement with the results showing that the H4-K16$_{ac}$ model samples more elongated conformations. For all other studied levels of acetylation, we observe an increase of contact occupancies of the basic patch with other segments of the peptide. Consequently, in general, acetylation has a cumulative effect of making the peptide more cohesive, favoring the formation of contacts between different parts of the peptide, as shown by the hydrogen bonds (Figure 4.6) and the contact maps (Figure 4.7) analyses.

At a molecular level, the unique effects of K16 acetylation can be traced to the local conformational preferences. Our previous work demonstrated that, in the WT H4 tail, K16 and other residues in the region of high charge density are conformationally constrained, sampling only a fraction of the sterically allowed conformations described by the Ramachandran plot [39]. Figures 4.8 and S9 reveal that the acetylation of K16 further reduces the backbone's conformational flexibility, especially for K16 and R17. These changes in the Ramachandran plots reflect the fact that residues K16 and R17 adopt mostly a trans conformation, while in all other models, these residues sample an equilibrium between the cis and trans conformations. The corresponding entropy reduction can be explained, in part, by the steric constraints of adding an acetyl group to the basic patch. However, this effect is reversed by further acetylating the H4 tail, where additional acetylated lysines promote intra-chain contact formation, overcoming the steric constraints in the segment flanking

the $K16_{ac}$ residue. Consequently, the formation or breaking of the contacts between the basic patch and the rest of the peptide sensitively depend on the interplay between the electrostatic and steric interactions and the entropic effects.

Finally, to rationalize the unique features of the H4-$K16_{ac}$ we propose that the sole acetylation of K16 results in a cooperative transition, where, the acetylation of K16 fixes the stereochemistry of residues in the basic patch and favors the trans conformation of residues K16 and R17 (Figure 4.8), resulting, subsequently, in the formation of the stable 3_{10} helix in the chain segment from residues 7 to 9. Furthermore, this 3_{10} helix structurally divides the tail into two *domains*, where the interactions between the residues before the secondary structural element and the residues after are highly disrupted (Figure 4.7 and 4.4.C). This molecular level explanation highlights both local and global effects of K16 acetylation, whose uniqueness stems from K16's location in the region of high positive charge density and sterically constrained side-chains.

4.3.5 Proposed model for the recognition of acetylated H4 tails.

We determined that even though the average radii of gyration of H4 tails remain mostly constant, progressive lysine acetylation significantly changes the conformational preferences and long-range contacts of H4 tails [148], thus modifying the sampled ensemble. To understand how these changes might affect the recognition of H4 tails by other nucleosomes or proteins involved in chromatin structure regulation, we analyzed how the lysine residues are positioned at different levels of acetylation.

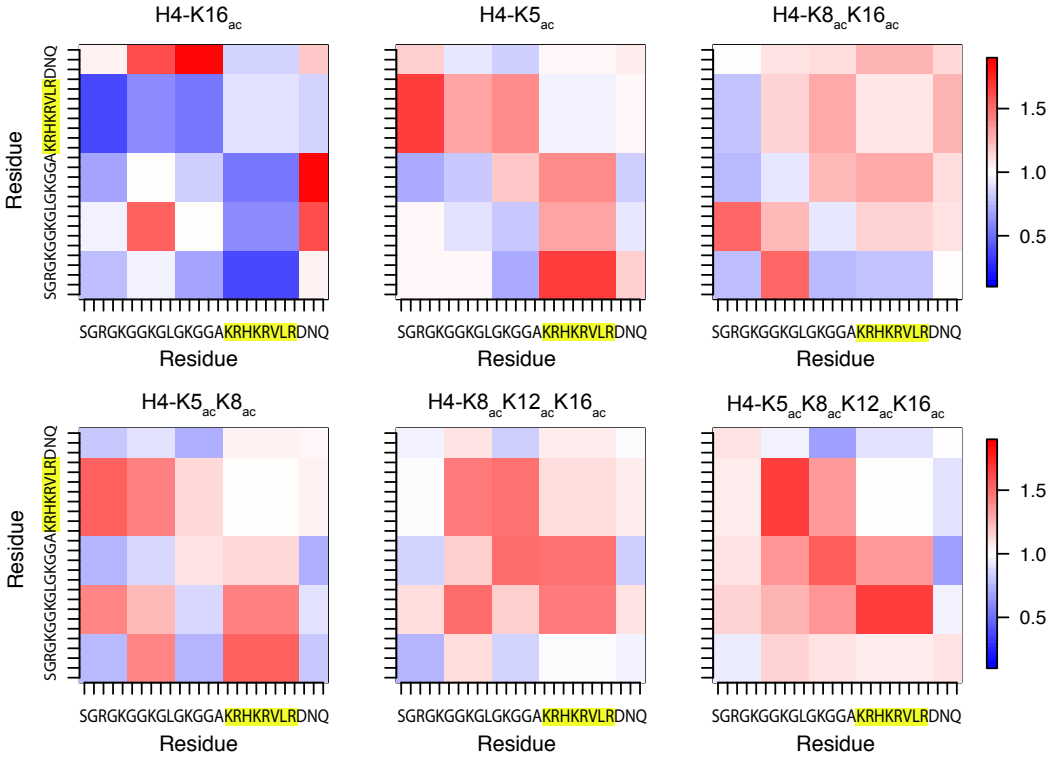


Figure 4.7: **Contact maps between different peptide segments.** Contact maps were computed by, first, determining the occupancy of the segment contacts and, then considering the ratio between the acetylated model and the WT. The ratios presented range from blue (a contact occupancy decrease with respect to the WT) to red (a contact occupancy increase with respect to the WT). The region of high positive charge density, residues 16-23, is highlighted in yellow.

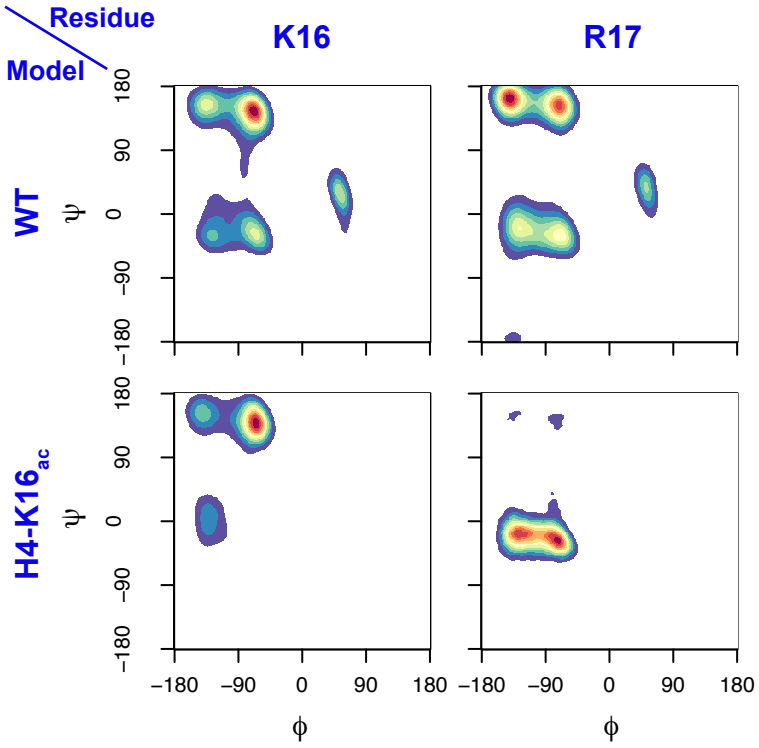


Figure 4.8: **Ramachandran plots of residues K16 and R17.**

By measuring the average distance between the NZ atoms of all lysines, we found that increasing the level of acetylation correlates with a smaller average distance between lysine side-chains (Figure 4.9), which can be explained, in part, by the overall charge reduction upon acetylation. By favoring conformations that bring acetylated lysines close to each other, H4 tails can create spatial clusters that act as recognition patches or docking sites for acetylation-dependent histone tail-binding proteins, via conformational selection, induced fit or a mixed mechanism [134, 149, 150]. This is in agreement with structural studies showing that, for proper binding, acetylation sites have to be closely spaced, such that a single acetyl-lysine binding protein can recognize more that one acetylation mark [134, 149].

As already discussed, the acetylation of K16 has effects that are unique in many of the considered metrics, exhibiting, for example, the lowest conformational heterogeneity and favoring structures with a specific 3_{10} helix spanning residues 7-9. This helix effectively promotes elongated structures with higher R_g, that position the acetylated K16 residue opposite to the N-terminal of the peptide, hence, exposing the lysine 16 residue, making it available for a specific recognition by various binding partners, leading, in turn, to unique biological consequences for this particular modification, in agreement with various experimental observations.

4.4 Conclusion

Our analysis of all-atom REMD simulations of the H4 N-terminal histone tail highlighted the effects of different levels of acetylation on the conformational

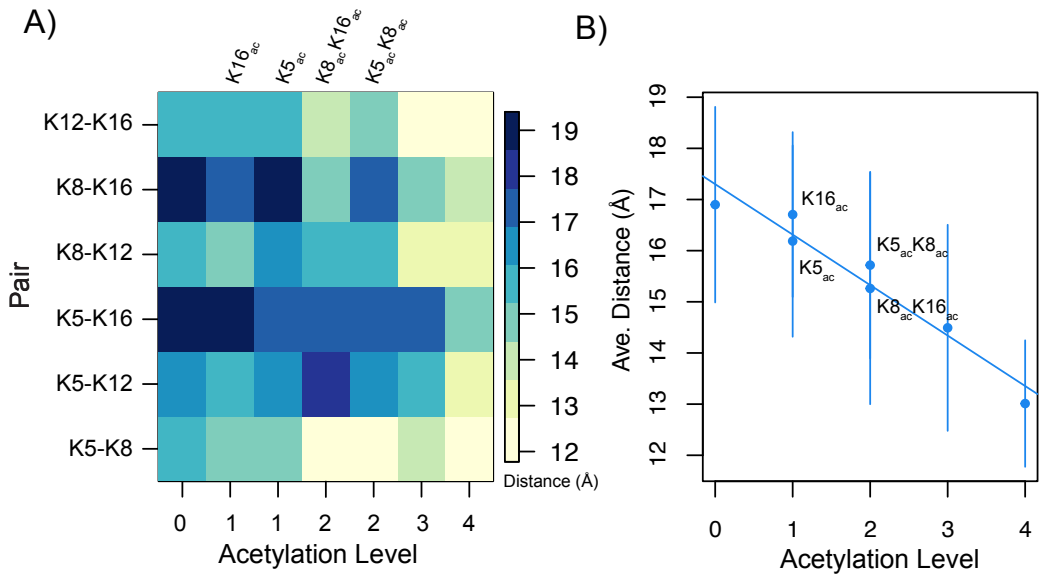


Figure 4.9: **Average distance between lysine side-chains.** (A) Average distance between all lysine pairs. (B) Average distance between all lysines at different levels of acetylation.

preferences. We demonstrated that, with the exception of the mono-acetylation of K16, progressive acetylation has a cumulative effect on both global and specific features of the conformational ensemble of the H4 tail. For example, our clustering analysis revealed that conformational heterogeneity decreases with acetylation. We also find that progressive acetylation results in higher helical propensities, both 3_{10} and α-helices.

Acetylation influences specific interactions between amino acids by increasing the hydrogen bond occupancy of acetylated lysines. The described structural changes occur mostly without significantly changing the average radius of gyration, suggesting that acetylation results in local perturbations that modify the structural preferences and heterogeneity of the H4 tail conformational ensemble. Over-

all, our investigation suggests that the electrostatic charge reduction and increased hydrophobicity upon acetylation are responsible for the cumulative effects of this post-translational modification on the H4 histone tail. Furthermore, by using a variety of metrics to characterize the disordered state of the H4 tail we demonstrate that local changes, such a formation of secondary structure, do not necessarily change some of the global or average properties of the polypeptide (i.e. radius of gyration, lysine hydrogen bonding and inter-lysine distances).

While we identified many of the effects of acetylation to be cumulative in nature, our analysis illustrates that the effect of K16 mono-acetylation is unique. K16 is found within a region of high positive charge density of the H4 tail, which plays an important role in binding and recognition interactions [38]. We illustrate how the mono-acetylation of K16 has unique global effects in the corresponding conformational ensemble, which is the least heterogeneous and exhibits a larger radius of gyration. We determined that the acetylation of lysine 16 effectively introduces a soft entropic penalty, rigidifying the chain in the vicinity of lysine 16. These local effects on one segment of the H4 tail induce the formation of highly localized, specific helix in the H4-K16$_{ac}$ system, leading to more elongated chain conformations. Furthermore, these elongated conformations may play a key role in exposing the acetylated K16 residue for participating in *direct* and *effector mediated* interactions with various chromatin regulatory proteins and DNA, which have a special role in chromatin structure [34] and transcription regulations [131].

We propose that spatial clustering of the acetyl-lysines will create recognition patches that could facilitate the recruitment of effector proteins via conformational

selection, induced fit or a mixed mechanism. This result is consistent with structural studies that show that the binding of H4 tails to bromodomains, or other acetylation-dependent histone tail-binding proteins, often require patterns of acetylation marks which, in turn, are highly sensitive to modifications flanking the acetylation site [134, 149]. These studies also show that these proteins bind more strongly to H4 tails with higher levels of acetylation [149]. Overall, our study suggests that the acetylation code for an isolated H4 histone tail has largely cumulative effects in the conformational preferences of the peptide. However, highly specific effects were seen for one of the acetylation patterns. Future research should address how the presence of histone cores and/or nucleosomal DNA further modifies these conformational preferences.

4.5 Summary and Biological Implications

In summary, progressive acetylation has a cumulative effect on the structural preferences of the H4 histone tail by reducing the overall charge and increasing the hydrophobicity of the peptide. However, the mono-acetylation of K16 uniquely makes the tail more rigid and leads to more extended conformations, which could play an important role in higher order chromatin structure and transcription regulation. Furthermore, increasing the level of acetylation of the H4 tail leads to spatially clustered acetyl-lysines, which could serve as recognition patches or docking sites for acetylation-dependent histone tail-binding proteins. Overall, acetylation has largely cumulative effects on the conformational preferences of the H4 histone tail, however,

the effects of the mono-acetylation of K16 are highly specific.

Appendix A: Chapter 2 Supplementary Information

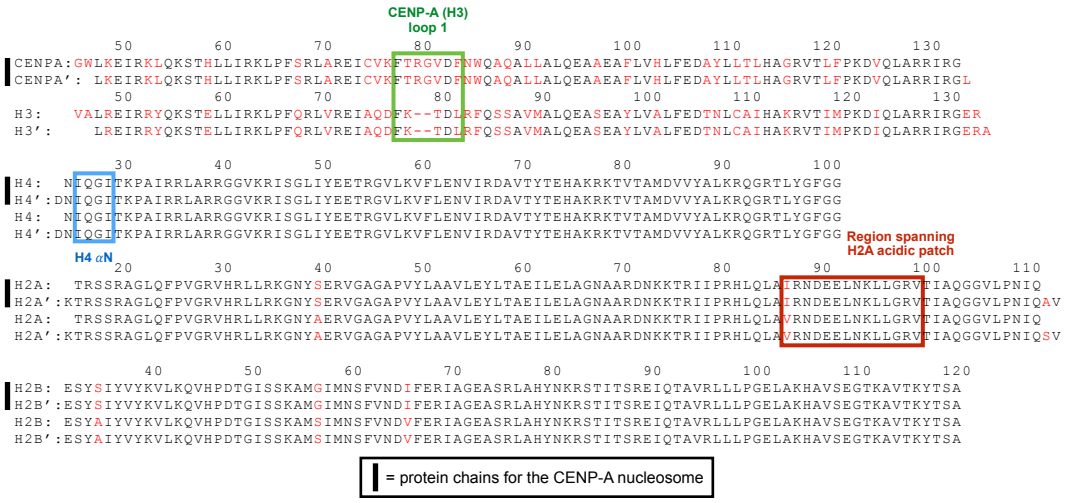


Figure A.1: **Amino acid sequences of the H3 and CENP-A systems.** Amino acid sequence alignment for the H3 and CENP-A systems. Boxes highlight important regions.

Table A.1: Distances between dimer centers-of-mass (Å)

System	**(H3/H4) to (H3'/H4')**	**(H3/H4) to (H2A/H2B)**	**(H3'/H4') to (H2A'/H2B')**
H3 nucleosome	34.3 ± 0.18	33.4 ± 0.27	32.4 ± 0.21
CENP-A nucleosome	34.7 ± 0.22	33.7 ± 0.34	32.8 ± 0.30
H3 octamer	33.9 ± 0.21	33.7 ± 0.25	32.8 ± 0.27
CENP-A octamer	34.5 ± 0.29	33.6 ± 0.39	33.0 ± 0.35

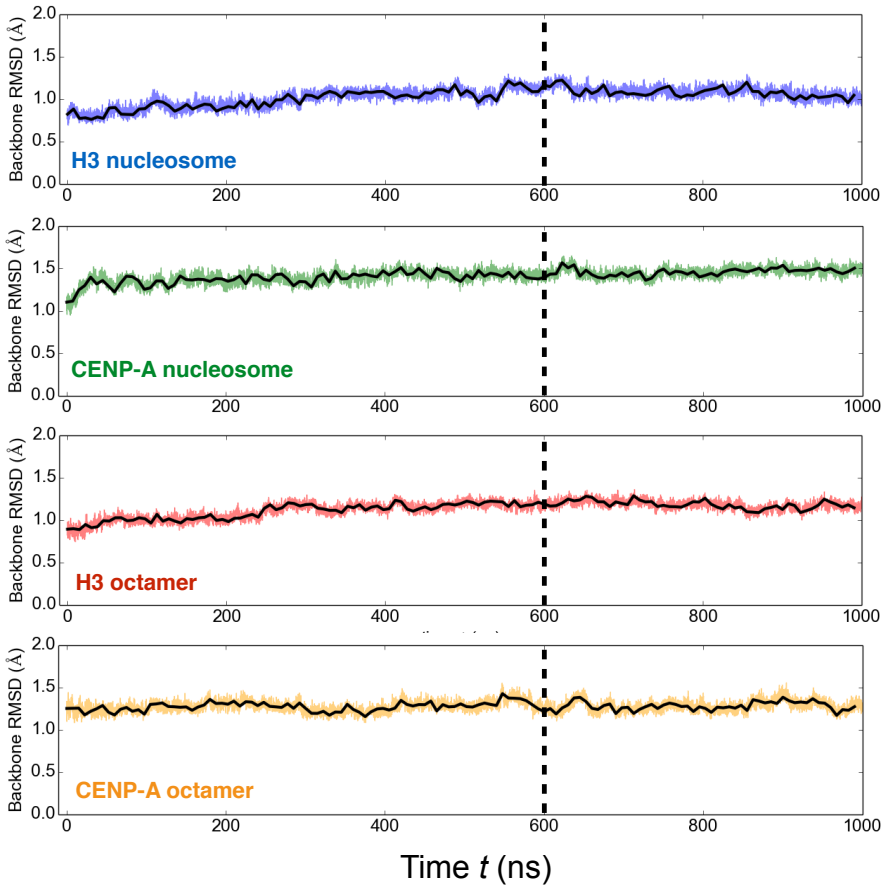


Figure A.2: **H3 and CENP-A systems reach stable equilibrium after 600 ns.** Protein backbone RMSD to the energy-minimized structures as a function of time for the H3 and CENP-A systems. The solid black lines represent running averages. The vertical dashed lines indicate the amount of time removed to ensure a stable equilibrium for each system.

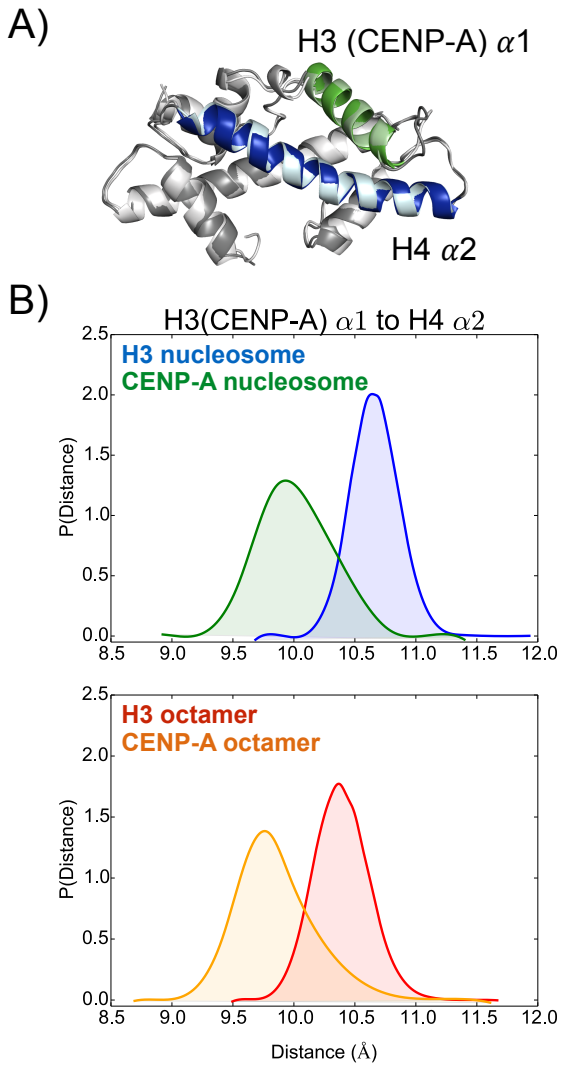


Figure A.3: **CENP-A/H4 is more compact, on average, than H3/H4 in the context of the nucleosome and octamer structures.** (A) Crystal structure distances between H3 α1 and H4 α2 (9.9 Å), and between CENP-A α1 and H4 α2 (10.1 Å). (B) Probability density functions for the distances between these two specific helices for the H3 and CENP-A systems. Structure figures drawn in Pymol.

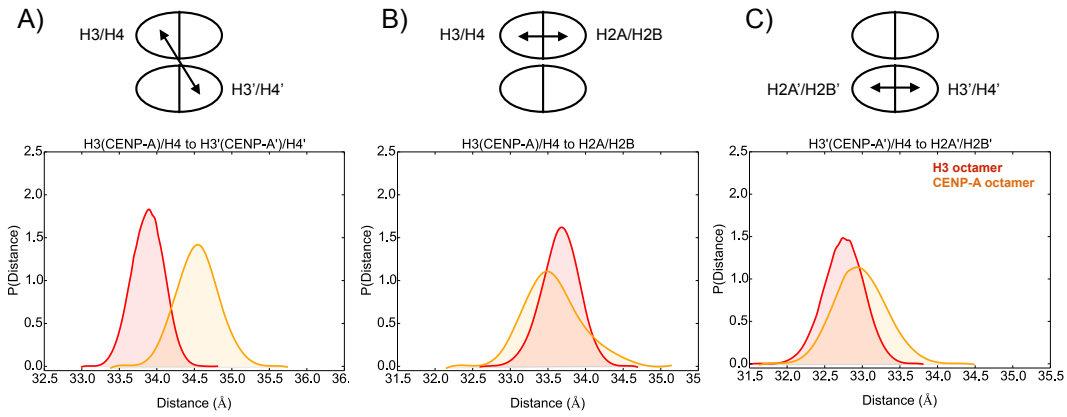


Figure A.4: **The CENP-A octamer exhibits greater global flexibility than the H3 octamer** Probability density functions for the distances within tetramers for the H3 and CENP-A octamers.

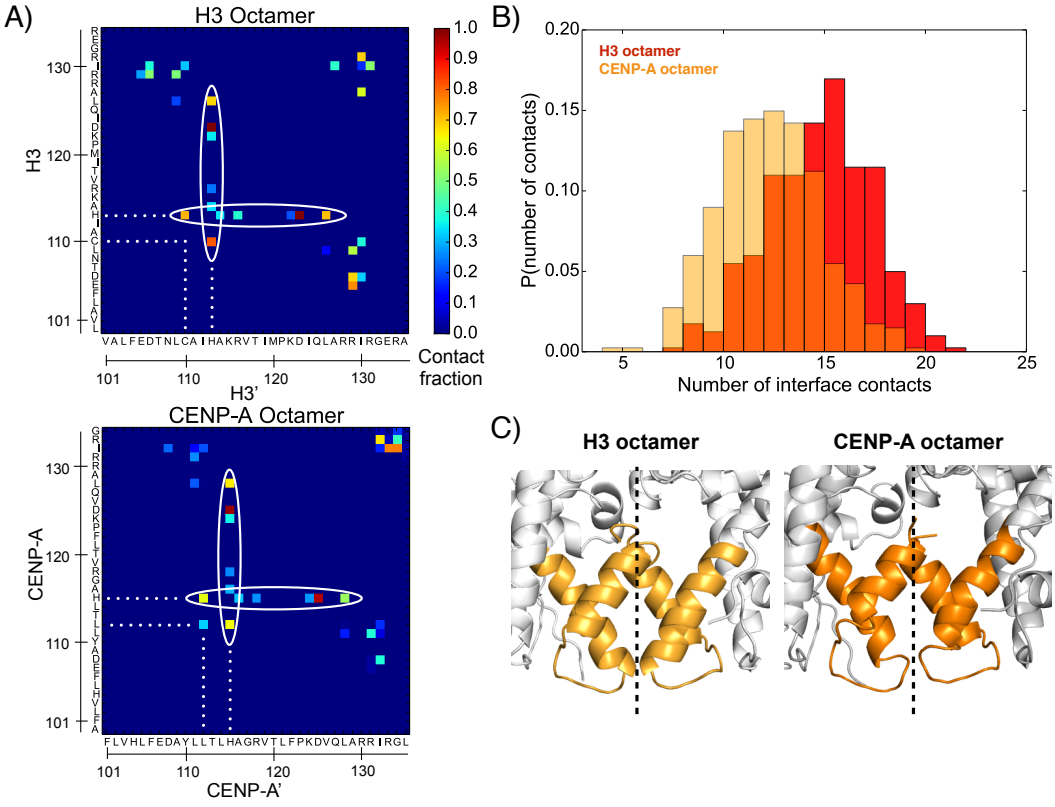


Figure A.5: **The CENP-A octamer dimerization interface is formed by fewer, and weaker, contacts than the corresponding interface in the H3 octamer.** (A) Contact maps for the dimerization interfaces of the H3 and CENP-A octamers. (B) Histograms of the number of H3–H3' and CENP-A–CENP-A' contacts. (C) The dimerization interfaces for H3–H3' (light orange) and CENP-A–CENP-A' (dark orange). The dashed line represents the pseudo-dyad. Structure figures drawn in Pymol.

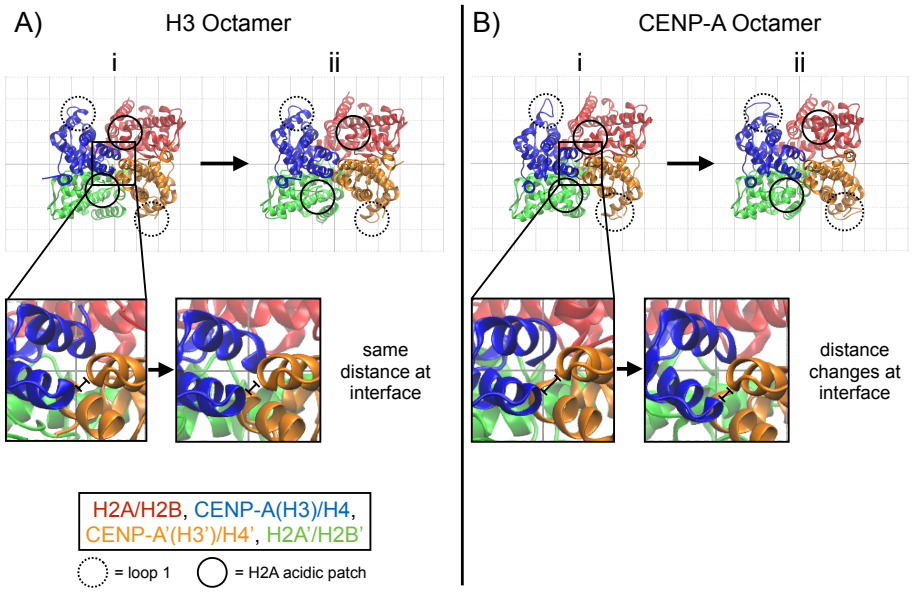


Figure A.6: **Greater distortion at the CENP-A dimerization interface, relative to the H3 octamer, is a major mode of motion of the CENP-A octamer.** (A) Two-dimensional plots displaying the top principal component for the H3 octamer viewed from the side of the DNA supercoil axis. (B) Plots for the top principal component of the CENP-A octamer. The insets display the motion of the H3–H3' and CENP-A–CENP-A' dimerization interfaces. Structure figures drawn in VMD.

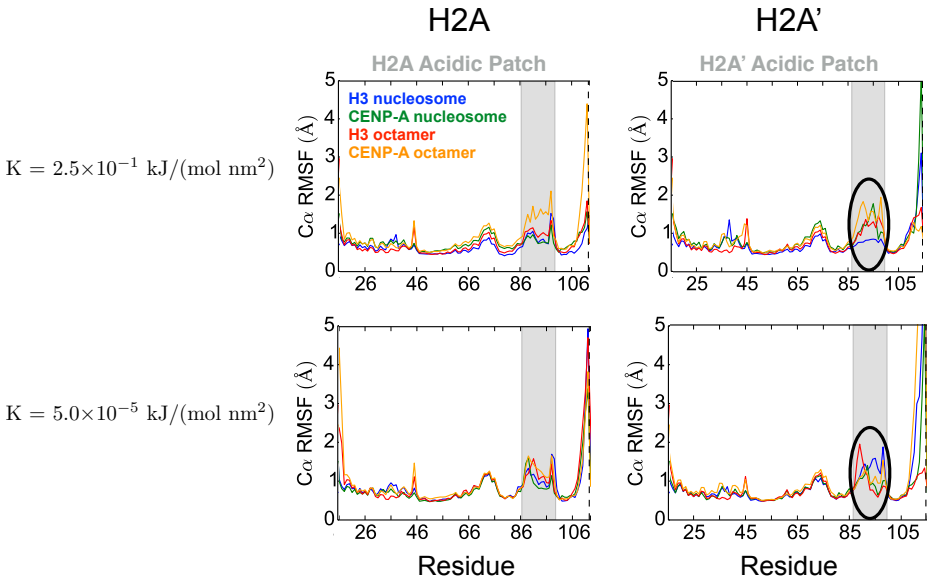


Figure A.7: **Asymmetric variability in H2A acidic patch mobility.** Cα RMSF of proteins H2A and H2A' in all four systems considered. Gray boxes define the regions spanning the H2A acidic patch (residues 87-100) in both of the heterotypic halves of the protein cores. Labels along the top identify the heterotypic half, and labels along the left-hand side indicate the strength of harmonic position restraints applied to heavy atoms. Black ovals highlight differences in H2A' acidic patch mobility, discussed further in the main text.

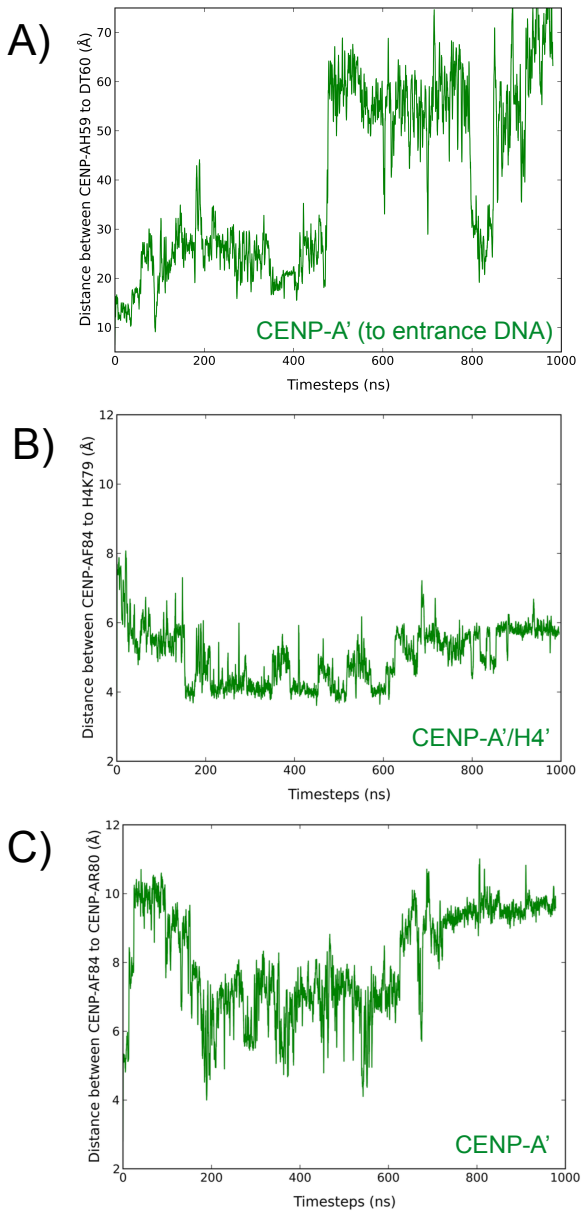


Figure A.8: **Key residue substitutions in CENP-A contribute to DNA instability near the pseudo-dyad.** As a function of time, we traced the distances between (A) CENP-A' His 59 and bp -60, (B) CENP-A' F84 and H4' K79, and between (C) CENP-A' F84 and CENP-A' R80. We specifically measured (A) NE2 of His 59 to P atom of DA-60; (B) CE1 of F84 to CB atom of K79; and (C) CE1 of F84 to CB atom of R80. CENP-A' His 59 interacts with entry DNA until detachment occurs, and interactions between CENP-A' R80, F84 and H4' K79 develop a hydrophobic pocket, weakening the CENP-A protein core affinity for DNA.

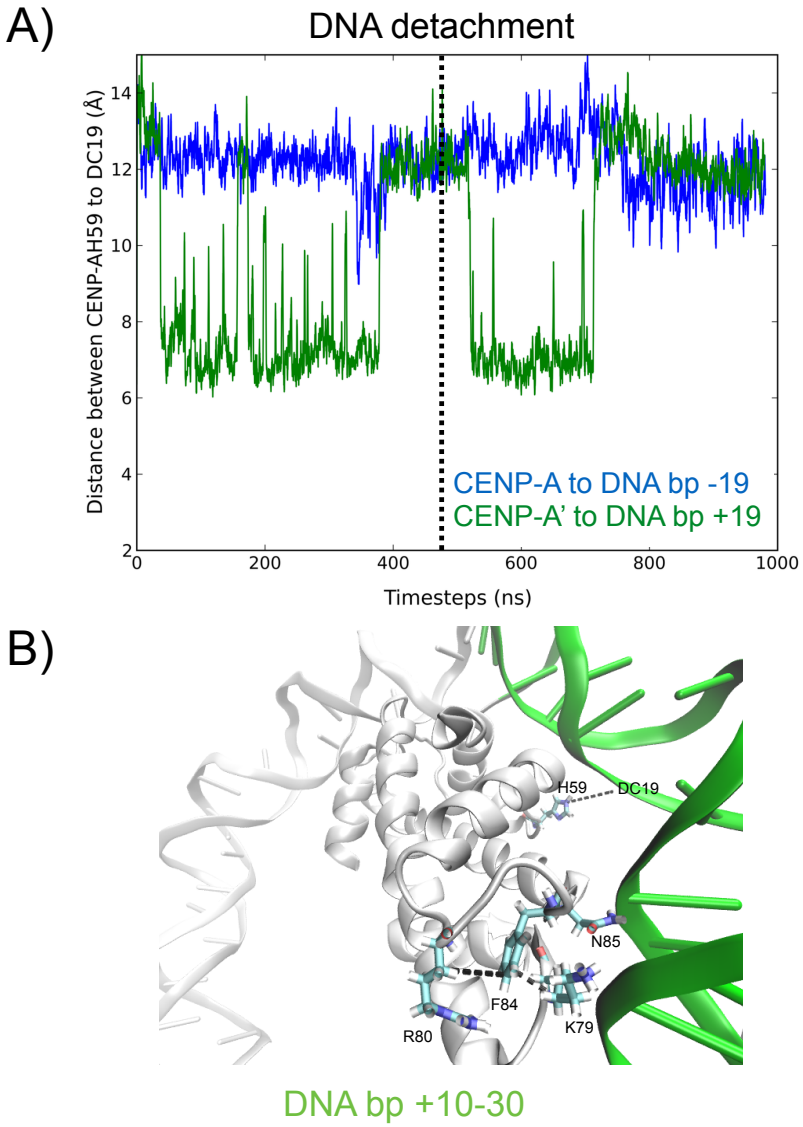


Figure A.9: **The asymmetric role of CENP-A H59 in DNA flexibility.** (A) Time traces of the distances between CENP-A H59 and DNA bp -19, and between CENP-A' H'59 and DNA bp +19. This illustrates that CENP-A' H'59 and CENP-A H59 play different roles in pseudo-dyad proximal DNA structural flexibility. (B) Residues from the CENP-A nucleosome, shown as sticks, that contribute to the relative instability of CENP-A nucleosomal DNA from basepairs +10 to +30. Structure figure drawn in VMD.

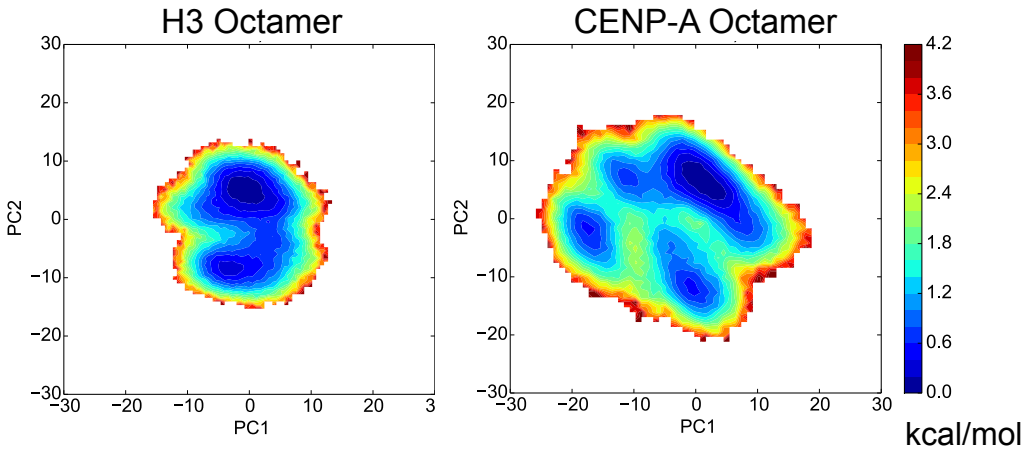
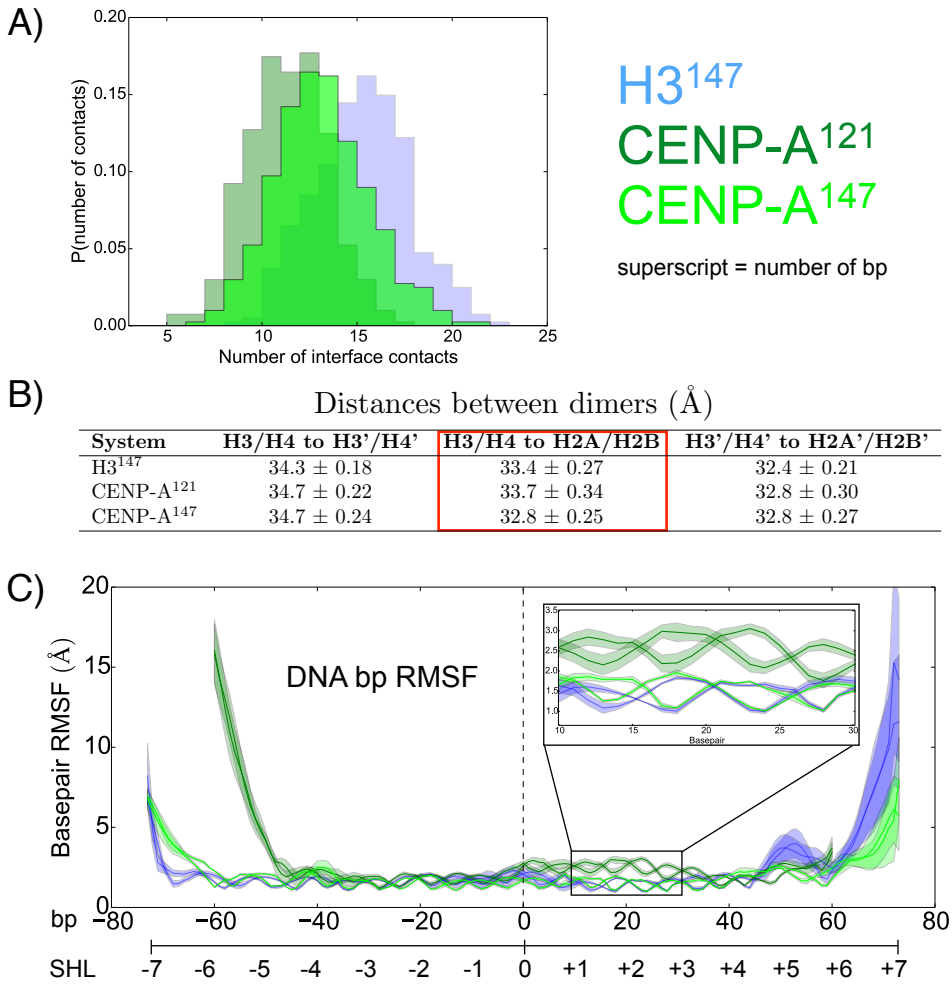


Figure A.10: **The characteristic free energy landscape is more rugged for the CENP-A octamer than for the H3 octamer.** Free energy projection of the H3 and CENP-A octamers onto their respective first two principal components.



B)

System	H3/H4 to H3'/H4'	H3/H4 to H2A/H2B	H3'/H4' to H2A'/H2B'
$H3^{147}$	34.3 ± 0.18	33.4 ± 0.27	32.4 ± 0.21
CENP-A^{121}	34.7 ± 0.22	33.7 ± 0.34	32.8 ± 0.30
CENP-A^{147}	34.7 ± 0.24	32.8 ± 0.25	32.8 ± 0.27

Figure A.11: **The effects of DNA length and sequence on CENP-A nucleosome dynamics.** A comparison between the CENP-A nucleosome with the 121 bp from 3AN2, denoted CENP-A^{121}, the CENP-A nucleosome with the 147 bp from 1KX5, denoted CENP-A^{147}, and the canonical H3 nucleosome from 1KX5, denoted $H3^{147}$. (A) Histograms of the number of CENP-A:CENP-A' dimerization interface contacts. (B) Distances between histone dimers, the red box highlighting the most significance difference. (C) Root mean squared fluctuations (RMSF) per base pair. The shaded areas represent ± one standard deviation. Two lines for each system correspond to the two strands of DNA. With the longer sequence from the crystal structure for the canonical H3 nucleosome (PDB ID 1KX5), the CENP-A nucleosomal DNA base pairs +10 to +30 remain more stably associated with the histone core.

Appendix B: Chapter 3 Supplementary Information

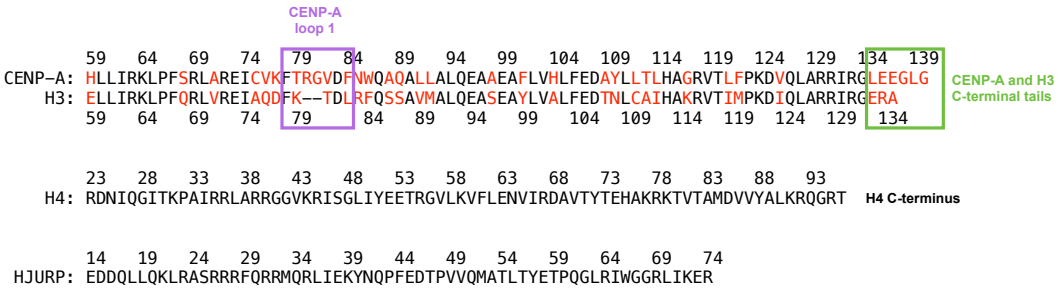


Figure B.1: **Amino acid sequences of the H3 and CENP-A within dimers.** Amino acid sequence alignment for the H3 and CENP-A dimers. Red letters indicate sequence differences, and boxes highlight important regions.

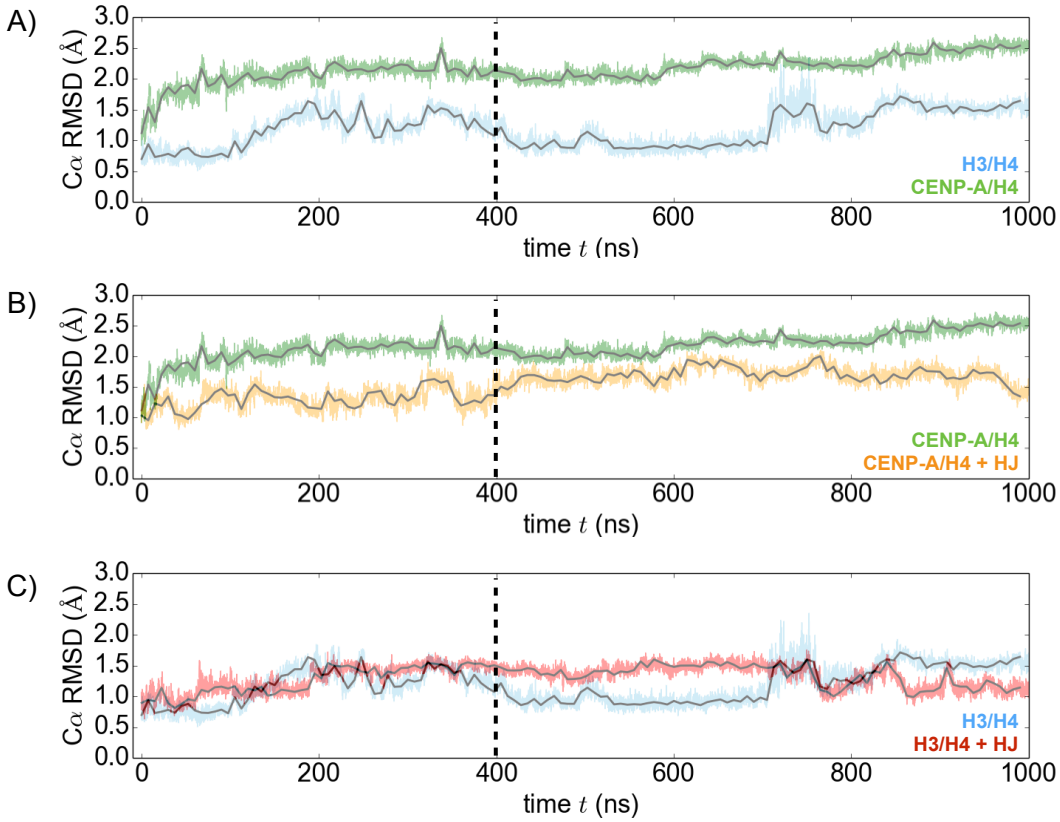


Figure B.2: **H3 and CENP-A dimers reach stable equilibrium after 400 ns.** Protein backbone RMSD to the experimentally resolved crystal structures as a function of time for the H3 and CENP-A dimers. The solid black lines represent running averages. The vertical dashed lines indicate the amount of time removed to ensure a stable equilibrium for each system.

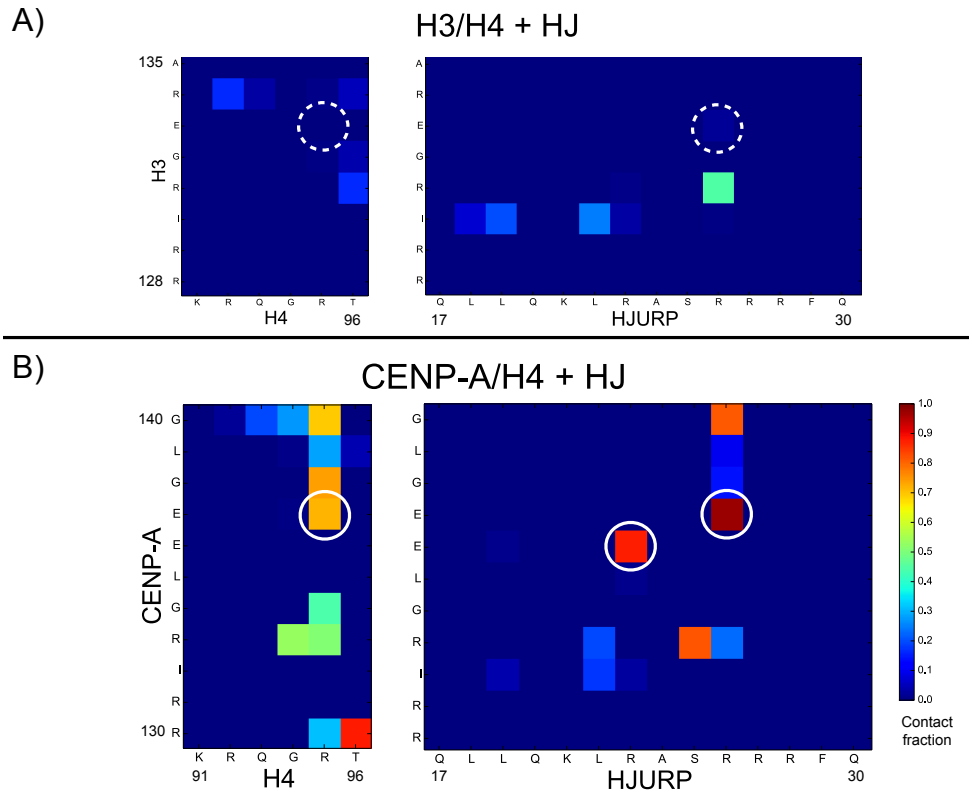


Figure B.3: **HJURP forms electrostatic interactions with the CENP-A C-terminus.** Contact maps of (A) the H3/H4 structure in a complex with HJURP, and (B) CENP-A/H4 in a complex with HJURP reveal that the α domain of HJURP weakly interacts with the C-termini of H3 and H4, and forms strong electrostatic contacts with the C-termini of CENP-A and H4.

Appendix C: Chapter 4 Supplementary Information

Pairwise RMSD and Q analysis: To characterize the heterogeneity of the unfolded ensemble we considered the pairwise RMSD and Q between all structures in simulated trajectories. By this approach we built a histogram of the RMSD or Q for all the pairs of conformations sampled. The mean and standard deviation of those distributions provide a fingerprint of the conformational ensemble heterogeneity. For details on how to calculate the pairwise Q to characterize the heterogeneity of the conformational ensemble see. [41, 146] Results are shown in figure S2.

Salt bridges analysis: We identified salt bridges between the one negatively charged aspartic acid residue and the positively charged arginines. An arginine and the aspartic acid were considered to form a salt bridge if the distance between the arginine center of positive charge, CZ, and one of the negatively charged oxygen atoms of aspartic acid, OD1 or OD2, was less than 4.0 Å. Results are shown in figure S9.

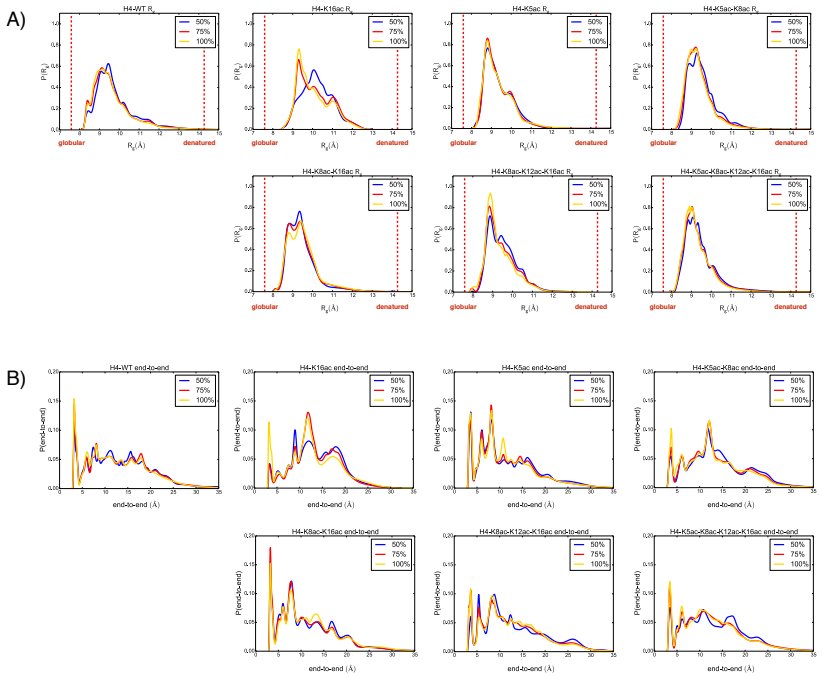


Figure C.1: **Convergence of the replica exchange simulations.** (A) Radius of gyration probability distributions obtained considering the full REMD at 300K trajectories (100%) and data sets with a fraction of the trajectories (75% and 50%). In all cases we considered the first X% of the trajectories. (B) End-to-end distance probability distributions obtained considering the full REMD at 300K trajectories (100%) and data sets with a fraction of the trajectories (75% and 50%). We observe that for data sets where the trajectories longer than 75% of the trajectories, the differences in the sampled conformations are negligible.

Table C.1: Average R_g and helical propensity for all levels of acetylation.

System	Ac level	$\langle R_g \rangle$ (Å)	Helical $(3_{10} + \alpha)\%$
H4-WT	0	9.7 ± 1.3	5.2 ± 0.4
H4-K16$_{ac}$	1	10.0 ± 0.8	5.0 ± 0.4
H4-K5$_{ac}$	1	9.4 ± 0.8	7.7 ± 1.5
H4-K8$_{ac}$K16$_{ac}$	2	9.5 ± 0.8	11.4 ± 2.2
H4-K5$_{ac}$K8$_{ac}$	2	9.4 ± 0.8	9.2 ± 0.4
H4-K8$_{ac}$K12$_{ac}$K16$_{ac}$	3	9.4 ± 0.8	9.7 ± 0.6
H4-K5$_{ac}$K8$_{ac}$K12$_{ac}$K16$_{ac}$	4	9.3 ± 0.8	10.5 ± 0.7

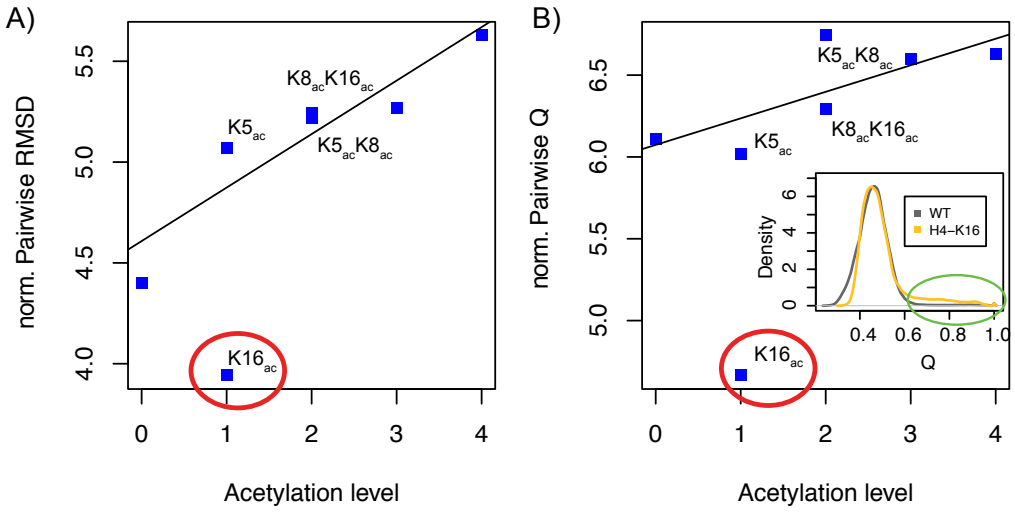


Figure C.2: **Heterogeneity of the conformational ensemble** (A) Mean pairwise RMSD among all sampled structures normalized by the standard deviation. (B) Mean pairwise Q among all sampled structures normalized by the standard deviation. Linear fits where done for all cases, except the H4-K16$_{ac}$ model. Using both metric we determined that the H4-K16$_{ac}$ system exhibits the least conformational heterogeneity.

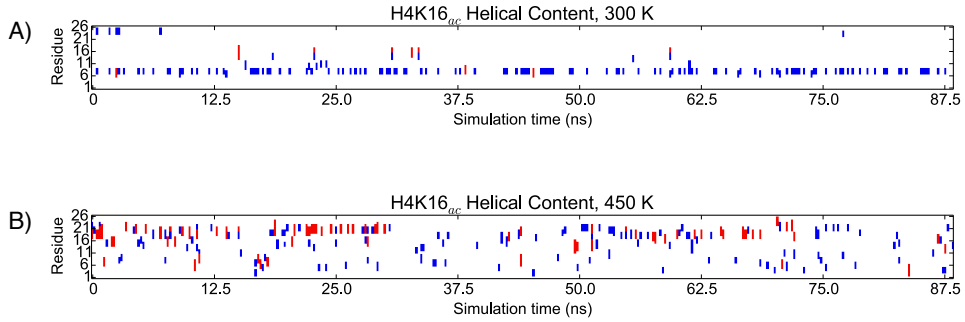


Figure C.3: **H4-K16$_{ac}$ helical content by simulation time** 3_{10} (blue) and α (red) helical content of the H4-K16$_{ac}$ tail by residue, as a function of simulation time for temperature trajectories at (A) 300 K, and (B) 450 K.

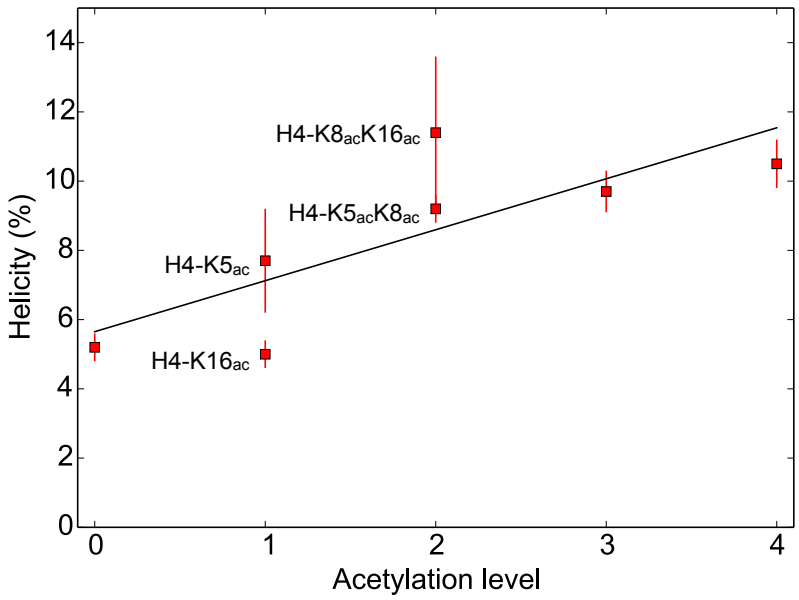


Figure C.4: **Percentage of helical residues.** Helical content of the histone H4 tail as a function of the number of acetylated lysines.

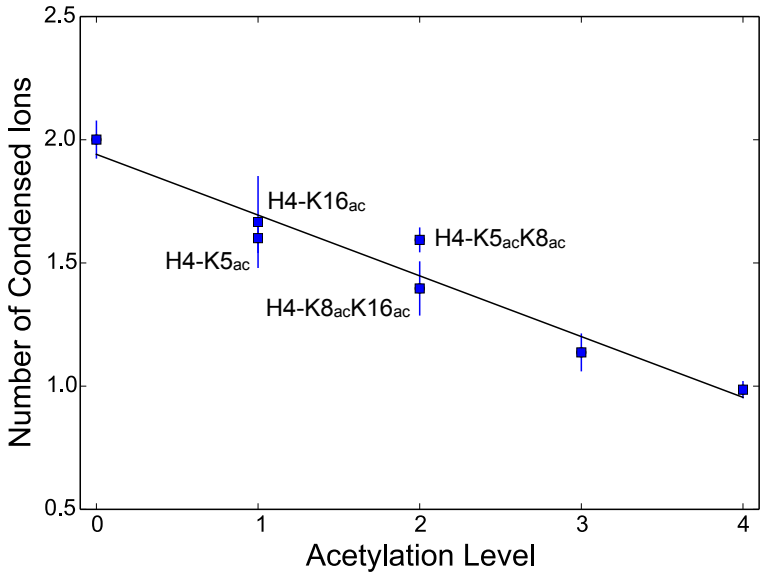


Figure C.5: **Ion Association.** Average number of condensed anions (Cl^-) around the positively charged H4 tail as a function of acetylation. A chloride ion is considered to be associated with the H4 tail if it is within the Bjerrum length (7.5 Å) of the head group nitrogen of an arginine or lysine.

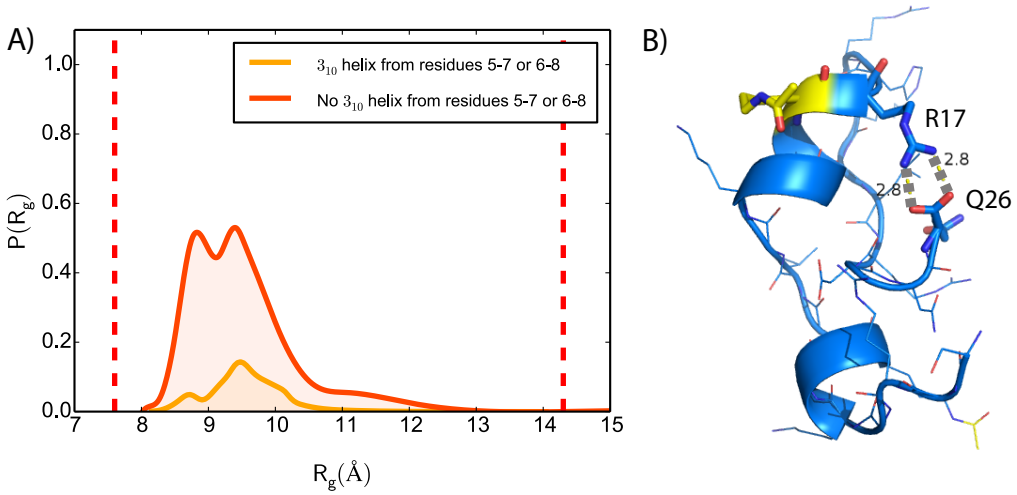


Figure C.6: **The H4-K8$_{ac}$K16$_{ac}$ model.** (A) The R_g of the H4-K8$_{ac}$K16$_{ac}$ di-acetylated model where the simulations frames were divided into two groups: (1) simulation frames with a 3_{10} helix formed from residues 5 to 7 or 6 to 8, and (2) simulation frames without these specific helices. (B) Characteristic structure obtained from the third most populated cluster exhibiting an elongated conformation. The acetylated K16 is shown in yellow.

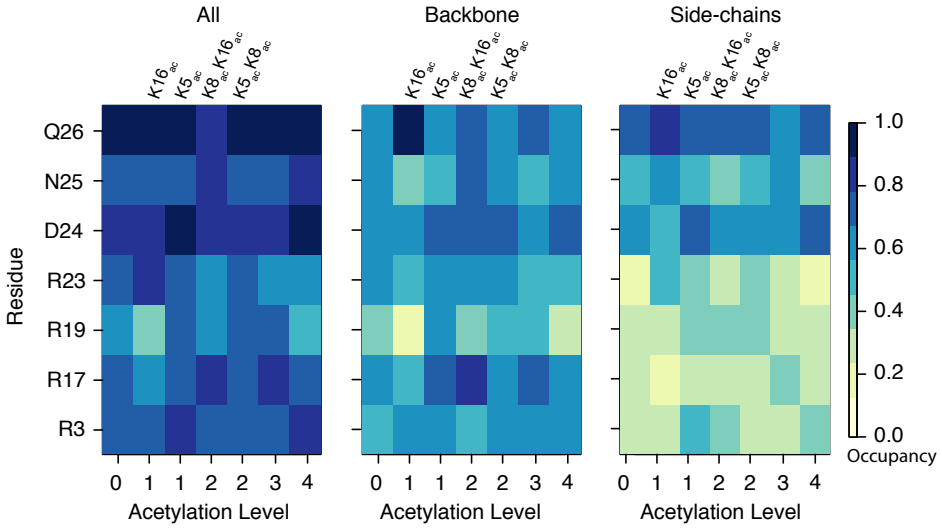


Figure C.7: **Hydrogen bonds occupancies.**

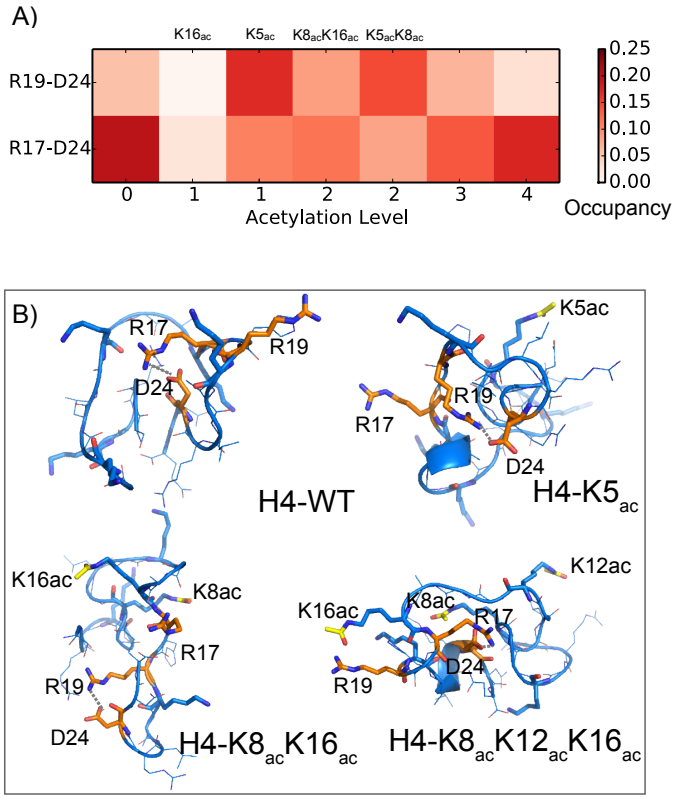


Figure C.8: **Stabilizing interactions and salt bridge occupancies.** (A) Salt bridge occupancies between the arginines (R17 or R19) and the aspartic acid (D24) at different levels of acetylation, measured as the percentage of time that the salt bridges are formed. (B) Sample structures are different levels of acetylations where salt-bridges stabilize the structures. Residues R17, R19 and D24 are shown in orange. Acetylations are shown in yellow.

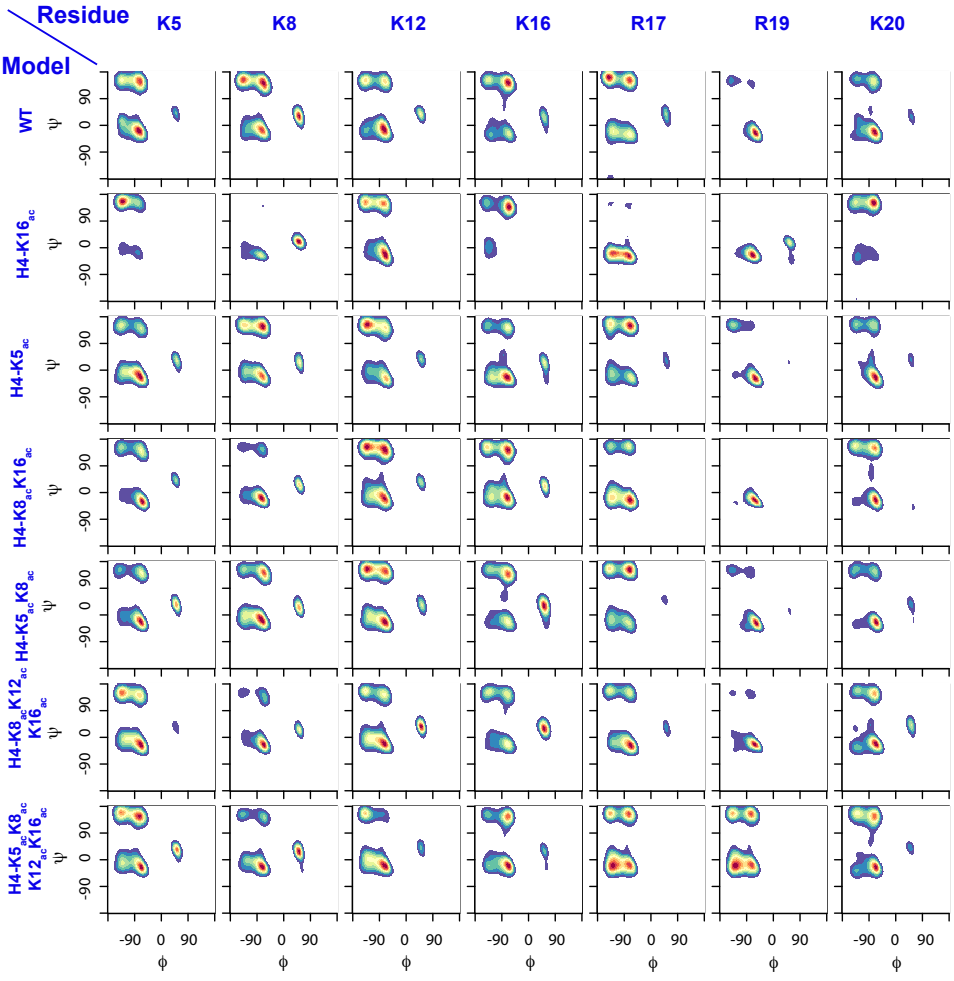


Figure C.9: **Ramachandran plots for selected residues.**

Bibliography

[1] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260, 1997.

[2] Ada L Olins and Donald E Olins. Spheroid chromatin units (ν bodies). *Science*, 183(4122):330–332, 1974.

[3] Roger D Kornberg. Chromatin structure: a repeating unit of histones and dna. *Science*, 184(4139):868–871, 1974.

[4] Curt A Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.*, 319(5):1097–1113, June 2002.

[5] Adolfo Ruiz-Carrillo, José L Jorcano, Gabi Eder, and Rudolf Lurz. In vitro core particle and nucleosome assembly at physiological ionic strength. *Proceedings of the National Academy of Sciences*, 76(7):3284–3288, 1979.

[6] Hiroaki Tachiwana, Wataru Kagawa, Tatsuya Shiga, Akihisa Osakabe, Yuta Miya, Kengo Saito, Yoko Hayashi-Takanaka, Takashi Oda, Mamoru Sato, Sam-Yong Park, Hiroshi Kimura, and Hitoshi Kurumizaka. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*, 476(7359):232–235, August 2011.

[7] Richard D Shelby, Omid Vafa, and Kevin F Sullivan. Assembly of CENP-A into Centromeric Chromatin Requires a Cooperative Array of Nucleosomal DNA Contact Sites. *J. Cell Biol.*, 136(3):501–513, 1997.

[8] Kinya Yoda, Satoshi Ando, Setsuo Morishita, Kenichi Houmura, Keiji Hashimoto, Kunio Takeyasu, and Tuneko Okazaki. Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. *Proc. Natl. Acad. Sci. U. S. A.*, 97(13):7266–71, June 2000.

[9] Yoshinori Tanaka, Hiroaki Tachiwana, Kinya Yoda, Hiroshi Masumoto, Tsuneko Okazaki, Hitoshi Kurumizaka, and Shigeyuki Yokoyama. Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. *J. Biol. Chem.*, 280(50):41609–18, December 2005.

[10] Raymond Camahort, Manjunatha Shivaraju, Mark Mattingly, Bing Li, Shima Nakanishi, Dongxiao Zhu, Ali Shilatifard, Jerry L Workman, and Jennifer L Gerton. Cse4 is part of an octameric nucleosome in budding yeast. *Mol. Cell*, 35(6):794–805, September 2009.

[11] Yamini Dalal, Hongda Wang, Stuart Lindsay, and Steven Henikoff. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS Biol.*, 5(8):e218, August 2007.

[12] Gaku Mizuguchi, Hua Xiao, Jan Wisniewski, M Mitchell Smith, and Carl Wu. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell*, 129(6):1153–64, June 2007.

[13] Jessica S Williams, Takeshi Hayashi, Mitsuhiro Yanagida, and Paul Russell. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol. Cell*, 33(3):287–98, February 2009.

[14] Takehito Furuyama and Steven Henikoff. Centromeric nucleosomes induce positive dna supercoils. *Cell*, 138(1):104–113, 2009.

[15] Nikolina Sekulic, Emily A Bassett, Danielle J Rogers, and Ben E Black. The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature*, 467(7313):347–351, September 2010.

[16] Mekonnen Lemma Dechassa, Katharina Wyns, Ming Li, Michael A Hall, Michelle D Wang, and Karolin Luger. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat. Commun.*, 2(May):313, January 2011.

[17] Weiguo Zhang, Serafin U Colmenares, and Gary H Karpen. Assembly of drosophila centromeric nucleosomes requires cid dimerization. *Molecular cell*, 45(2):263–269, 2012.

[18] Manjunatha Shivaraju, Jay R Unruh, Brian D Slaughter, Mark Mattingly, Judith Berman, and Jennifer L Gerton. Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell*, 150(2):304–316, 2012.

[19] Minh Bui, Emilios K Dimitriadis, Christian Hoischen, Eunkyung An, Delphine Quénet, Sindy Giebe, Aleksandra Nita-Lazar, Stephan Diekmann, and Yamini Dalal. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell*, 150(2):317–326, July 2012.

[20] Takehito Furuyama, Christine A Codomo, and Steven Henikoff. Reconstitution of hemisomes on budding yeast centromeric dna. *Nucleic acids research*, 41(11):5769–5783, 2013.

[21] Dan Hasson, Tanya Panchenko, Kevan J Salimian, Mishah U Salman, Nikolina Sekulic, Alicia Alonso, Peter E Warburton, and Ben E Black. The octamer is the major form of cenp-a nucleosomes at human centromeres. *Nature structural & molecular biology*, 20(6):687–695, 2013.

[22] Jan Wisniewski, Bassam Hajj, Jiji Chen, Gaku Mizuguchi, Hua Xiao, Debbie Wei, Maxime Dahan, and Carl Wu. Imaging the fate of histone cse4 reveals de novo replacement in s phase and subsequent stable residence at centromeres. *eLife*, 3:e02203, 2014.

[23] Steven Henikoff, Srinivas Ramachandran, Kristina Krassovsky, Terri D Bryson, Christine A Codomo, Kristin Brogaard, Jonathan Widom, Ji-Ping Wang, and Jorja G Henikoff. The budding yeast centromere dna element ii wraps a stable cse4 hemisome in either orientation in vivo. *Elife*, 3:e01861, 2014.

[24] Ben E Black, Daniel R Foltz, Srinivas Chakravarthy, and Karolin Luger. Structural determinants for generating centromeric chromatin. *Nature*, 430(July):578–582, 2004.

[25] Matthew D D Miell, Colin J Fuller, Annika Guse, Helena M Barysz, Andrew Downes, Tom Owen-Hughes, Juri Rappsilber, Aaron F Straight, and Robin C Allshire. CENP-A confers a reduction in height on octameric nucleosomes. *Nat. Struct. Mol. Biol.*, 20(6):763–5, June 2013.

[26] Sergei Y Ponomarev, Vakhtang Putkaradze, and Thomas C Bishop. Relaxation dynamics of nucleosomal dna. *Physical Chemistry Chemical Physics*, 11(45):10633–10643, 2009.

[27] Danilo Roccatano, Andre Barthel, and Martin Zacharias. Structural Flexibility of the Nucleosome Core Particle at Atomic Resolution Studied by Molecular Dynamics Simulation. *Biopolymers*, 85(5):407–421, 2007.

[28] Christopher K Materese, Alexey Savelyev, and Garegin A Papoian. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J. Am. Chem. Soc.*, 131(41):15005–13, October 2009.

[29] Hidetoshi Kono, Kazuyoshi Shirayama, Yasuhiro Arimura, Hiroaki Tachiwana, and Hitoshi Kurumizaka. Two arginine residues suppress the flexibility of nucleosomal DNA in the canonical nucleosome core. *PLoS One*, 10(3):e0120635, January 2015.

[30] Mithun Biswas, Karine Voltz, Jeremy C Smith, and Jörg Langowski. Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.*, 7(12):e1002279, December 2011.

[31] James Allan, Nerina Harborne, Donald C Rau, and Hannah Gould. Participation of core histone "tails" in the stabilization of the chromatin solenoid. *The Journal of cell biology*, 93(2):285–297, 1982.

[32] M Garcia-Ramirez, FENG Dong, and Juan Ausio. Role of the histone "tails" in the folding of oligonucleosomes depleted of histone h1. *Journal of Biological Chemistry*, 267(27):19587–19595, 1992.

[33] P M Schwarz, A Felthauser, T M Fletcher, and J C Hansen. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry*, 35(13):4009–15, April 1996.

[34] Michael Shogren-Knaak, Haruhiko Ishii, Jian-Min Sun, Michael J Pazin, James R Davie, and Craig L Peterson. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science Signaling*, 311(5762):844–847, February 2006.

[35] Pu-Yeh Kan, Tamara L Caterino, and Jeffrey J Hayes. The H4 tail domain participates in intra- and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays. *Molecular and cellular biology*, 29(2):538–546, January 2009.

[36] Darren Yang and Gaurav Arya. Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation . *Physical Chemistry Chemical Physics*, 13(7):2911–2921, 2011.

[37] Anna A Kalashnikova, Mary E Porter-Goff, Uma M Muthurajan, Karolin Luger, and Jeffrey C Hansen. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of the Royal Society, Interface / the Royal Society*, 10(82):20121022–20121022, May 2013.

[38] Abdollah Allahverdi, Renliang Yang, Nikolay Korolev, Yanping Fan, Curt a Davey, Chuan-Fa Liu, and Lars Nordenskiöld. The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.*, 39(5):1680–91, March 2011.

[39] Davit A Potoyan and Garegin A Papoian. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.*, 133(19):7405–15, May 2011.

[40] Roberto D Lins and Ursula Röthlisberger. Influence of long-range electrostatic treatments on the folding of the n-terminal h4 histone tail peptide. *Journal of Chemical Theory and Computation*, 2(2):246–250, 2006.

[41] Davit A Potoyan and Garegin A Papoian. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl. Acad. Sci. U. S. A.*, 109(44):17857–62, October 2012.

[42] Nikolay Korolev, Hang Yu, Alexander P Lyubartsev, and Lars Nordenskiöld. Molecular dynamics simulations demonstrate the regulation of DNA-DNA attraction by H4 histone tail acetylations and mutations. *Biopolymers*, 101(10):1051–64, October 2014.

[43] David Winogradoff, Ignacia Echeverria, Davit A Potoyan, and Garegin A Papoian. The acetylation landscape of the h4 histone tail: disentangling the interplay between the specific and cumulative effects. *Journal of the American Chemical Society*, 137(19):6245–6253, 2015.

[44] Natalia Conde e Silva, Ben E. Black, Andrei Sivolob, Jan Filipski, Don W. Cleveland, and Ariel Prunell. CENP-A-containing Nucleosomes: Easier Disassembly versus Exclusive Centromeric Localization. *J. Mol. Biol.*, 370(3):555–573, July 2007.

[45] Marcin P Walkiewicz, Emilios K Dimitriadis, and Yamini Dalal. CENP-A octamers do not confer a reduction in nucleosome height by AFM. *Nat. Struct. Mol. Biol.*, 21(1):2–3, January 2014.

[46] Christine A Codomo, Takehito Furuyama, and Steven Henikoff. CENP-A octamers do not confer a reduction in nucleosome height by AFM. *Nat. Struct. Mol. Biol.*, 21(1):4–5, January 2014.

[47] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):1–10, April 2013.

[48] Robert B. Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, 113(26):9004–9015, 2009.

[49] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, June 2010.

[50] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E Cheatham, Charles A Laughton, and Modesto Orozco. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, 92(11):3817–3829, June 2007.

[51] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A Second Generation Force Field for the Simulation of Proteins , Nucleic Acids , and Organic Molecules. *J. Am. Chem. Soc.*, 117(6):5179–5197, 1995.

[52] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[53] Berk Hess. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.*, 4(1):116–122, January 2008.

[54] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[55] M. Parrinello. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182, 1981.

[56] David J Barlow and JM Thornton. Ion-pairs in proteins. *Journal of molecular biology*, 168(4):867–885, 1983.

[57] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins Struct. Funct. Bioinforma.*, 17(4):412–425, 1993.

[58] Ben E Black, Melissa A Brock, Sabrina Bédard, Virgil L Woods, and Don W Cleveland. An epigenetic mark generated by the incorporation of cenp-a into centromeric nucleosomes. *Proceedings of the National Academy of Sciences*, 104(12):5008–5013, 2007.

[59] Samantha J Falk, Lucie Y Guo, Nikolina Sekulic, Evan M Smoak, Tomoyasu Mani, Glennis A Logsdon, Kushol Gupta, Lars ET Jansen, Gregory D Van Duyne, Sergei A Vinogradov, Michael A Lampson, and Ben E Black. Cenp-c reshapes and stabilizes cenp-a nucleosomes at the centromere. *Science*, 348(6235):699–703, 2015.

[60] Anna A Kalashnikova, Mary E Porter-Goff, Uma M Muthurajan, Karolin Luger, and Jeffrey C Hansen. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of The Royal Society Interface*, 10(82):20121022, 2013.

[61] Karim-Jean Armache, Joseph D Garlick, Daniele Canzio, Geeta J Narlikar, and Robert E Kingston. Structural basis of silencing: Sir3 bah domain in complex with a nucleosome at 3.0 å resolution. *Science*, 334(6058):977–982, 2011.

[62] Andrew J Barbera, Jayanth V Chodaparambil, Brenna Kelley-Clarke, Vladimir Joukov, Johannes C Walter, Karolin Luger, and Kenneth M Kaye. The nucleosomal surface as a docking station for kaposi's sarcoma herpesvirus lana. *Science*, 311(5762):856–861, 2006.

[63] Hidenori Kato, Jiansheng Jiang, Bing-Rui Zhou, Marieke Rozendaal, Hanqiao Feng, Rodolfo Ghirlando, T Sam Xiao, Aaron F Straight, and Yawen Bai. A

conserved mechanism for centromeric nucleosome recognition by centromere protein cenp-c. *Science*, 340(6136):1110–1113, 2013.

[64] Michael G Poirier, Eugene Oh, Hannah S Tims, and Jonathan Widom. Dynamics and function of compact nucleosome arrays. *Nature structural & molecular biology*, 16(9):938–944, 2009.

[65] Ho Sung Rhee, Alain R Bataille, Liye Zhang, and B Franklin Pugh. Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell*, 159(6):1377–1388, 2014.

[66] Thuy T M Ngo, Qiucen Zhang, Ruobo Zhou, Jaya G Yodh, and Taekjip Ha. Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility. *Cell*, 160(6):1135–44, March 2015.

[67] A Magalhaes, B Maigret, J Hoflack, JNF Gomes, and HA Scheraga. Contribution of unusual arginine-arginine short-range interactions to stabilization and recognition in proteins. *Journal of protein chemistry*, 13(2):195–215, 1994.

[68] Jiří Vondrášek, Philip E Mason, Jan Heyda, Kim D Collins, and Pavel Jungwirth. The molecular origin of like-charge arginine- arginine pairing in water. *The Journal of Physical Chemistry B*, 113(27):9041–9045, 2009.

[69] Garegin A Papoian, Johan Ulander, Michael P Eastwood, Zaida Luthey-Schulten, and Peter G Wolynes. Water in protein structure prediction. *Proc. Natl. Acad. Sci. U. S. A.*, 101(10):3352–3357, 2004.

[70] Peter G Wolynes, Jose N Onuchic, and D Thirumalai. Navigating the folding routes. *Science*, pages 1619–1619, 1995.

[71] Osamu Miyashita, Peter G Wolynes, and Jose N Onuchic. Simple energy landscape model for the kinetics of functional transitions in proteins. *The Journal of Physical Chemistry B*, 109(5):1959–1969, 2005.

[72] Timothy J Richmond and Curt A Davey. The structure of DNA in the nucleosome core. *Nature*, 423(6936):145–150, May 2003.

[73] Ravindra D Makde, Joseph R England, Hemant P Yennawar, and Song Tan. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*, 467(7315):562–6, September 2010.

[74] William C Earnshaw, Kevin E Sullivan, Carol A Cooke, and A Kaiser. Molecular Cloning of cDNA for CENP-B, the Major Human Centromere Autoantigen. *J. Cell Biol.*, 104(April):817–829, 1987.

[75] Hiroshi Masumoto, Hisao Masukata, Naohito Nozaki, and Tuneko Okazaki. A Human Centromere Antigen (CENP-B) Interacts with a Short Specific Sequence kn Alphoid DNA, a Human Centromeric Satellite. *J. Cell Biol.*, 109(November):1963–1973, 1989.

[76] Kinya Yoda, Katsumi Kitagawa, Hiroshi Masumoto, Yoshinao Muro, and Tuneko Okazaki. A Human Centromere Protein, CENP-B, Has a DNA Binding Domain Containing Four Potential Helices at the. *J. Cell Biol.*, 119(6):1413–1427, 1992.

[77] Yoshinori Tanaka, Osamu Nureki, Hitoshi Kurumizaka, Shuya Fukai, Shinichi Kawaguchi, Mari Ikuta, Junji Iwahara, and Tsuneko Okazaki. Crystal structure of the CENP-B protein ± DNA complex : the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J.*, 20(23):6612–6618, 2001.

[78] Christopher W Carroll, Mariana C C Silva, Kristina M Godek, Lars E T Jansen, and Aaron F Straight. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat. Cell Biol.*, 11(7):896–902, July 2009.

[79] Tanya Panchenko, Troy C Sorensen, Christopher L Woodcock, Zhong-Yuan Kan, and Stacey Wood. Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proc. Natl. Acad. Sci. U. S. A.*, 108(40):16588–16593, 2011.

[80] Yasuhiro Arimura, Kazuyoshi Shirayama, Naoki Horikoshi, Risa Fujita, Hiroyuki Taguchi, Wataru Kagawa, Tatsuo Fukagawa, Geneviève Almouzni, and Hitoshi Kurumizaka. Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci. Rep.*, 4(7115):1–7, January 2014.

[81] Junnan Fang, Yuting Liu, Yun Wei, Wenqiang Deng, Zhouliang Yu, Li Huang, Yan Teng, Ting Yao, Qinglong You, Haihe Ruan, Ping Chen, Rui-Ming Xu, and Guohong Li. Structural transitions of centromeric chromatin regulate the cell cycle-dependent recruitment of cenp-n. *Gene Dev.*, 29:1–16, 2015.

[82] Yamini Dalal, Takehito Furuyama, Danielle Vermaak, and Steven Henikoff. Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.*, 104(41):15974–15981, 2007.

[83] Maxim Y Sheinin, Ming Li, Mohammad Soltani, Karolin Luger, and Michelle D Wang. Torque modulates nucleosome stability and facilitates h2a/h2b dimer loss. *Nature communications*, 4, 2013.

[84] Marinela Perpelescu, Naohito Nozaki, Chikashi Obuse, Hua Yang, and Kinya Yoda. Active establishment of centromeric cenp-a chromatin by rsf complex. *J. Cell Biol.*, 185(3):397–407, 2009.

[85] Karolin Luger. Dynamic nucleosomes. *Chromosome Res.*, 14(1):5–16, 2006.

[86] Xu Lu, Matthew D Simon, Jayanth V Chodaparambil, Jeffrey C Hansen, Kevan M Shokat, and Karolin Luger. The effect of h3k79 dimethylation and

h4k20 trimethylation on nucleosome and chromatin structure. *Nat. Struct. Mol. Biol.*, 15(10):1122–1124, 2008.

[87] Marek Simon, Justin A North, John C Shimko, Robert A Forties, Michelle B Ferdinand, Mridula Manohar, Meng Zhang, Richard Fishel, Jennifer J Ottesen, and Michael G Poirier. Histone fold modifications control nucleosome unwrapping and disassembly. *Proc. Natl. Acad. Sci. U. S. A.*, 108(31):12711–12716, 2011.

[88] Andrew J Andrews and Karolin Luger. Nucleosome structure (s) and stability: variations on a theme. *Ann. Rev. Biophys.*, 40:99–117, 2011.

[89] Daniel R Foltz, Lars ET Jansen, Aaron O Bailey, John R Yates, Emily A Bassett, Stacey Wood, Ben E Black, and Don W Cleveland. Centromere-specific assembly of cenp-a nucleosomes is mediated by hjurp. *Cell*, 137(3):472–484, 2009.

[90] Elaine M Dunleavy, Danièle Roche, Hideaki Tagami, Nicolas Lacoste, Dominique Ray-Gallet, Yusuke Nakamura, Yataro Daigo, Yoshihiro Nakatani, and Geneviève Almouzni-Pettinotti. Hjurp is a cell-cycle-dependent maintenance and deposition factor of cenp-a at centromeres. *Cell*, 137(3):485–497, 2009.

[91] Sam Stoler, Kelly Rogers, Scott Weitze, Lisa Morey, Molly Fitzgerald-Hayes, and Richard E Baker. Scm3, an essential saccharomyces cerevisiae centromere protein required for g2/m progression and cse4 localization. *Proceedings of the National Academy of Sciences*, 104(25):10571–10576, 2007.

[92] Luis Sanchez-Pulido, Alison L Pidoux, Chris P Ponting, and Robin C Allshire. Common ancestry of the cenp-a chaperones scm3 and hjurp. *Cell*, 137(7):1173, 2009.

[93] Alison L Pidoux, Eun Shik Choi, Johanna KR Abbott, Xingkun Liu, Alexander Kagansky, Araceli G Castillo, Georgina L Hamilton, William Richardson, Juri Rappsilber, Xiangwei He, and Robin C Allshire. Fission yeast scm3: A cenp-a receptor required for integrity of subkinetochore chromatin. *Molecular cell*, 33(3):299–311, 2009.

[94] In Suk Joung and Thomas E Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The journal of physical chemistry B*, 112(30):9020–9041, 2008.

[95] Hao Hu, Yang Liu, Mingzhu Wang, Junnan Fang, Hongda Huang, Na Yang, Yanbo Li, Jianyu Wang, Xuebiao Yao, Yunyu Shi, Guohong Li, and Rui-Ming Xu. Structure of a cenp-a–histone h4 heterodimer in complex with chaperone hjurp. *Genes & development*, 25(9):901–906, 2011.

[96] Richard W Kriwacki, Ludger Hengst, Linda Tennant, Steven I Reed, and Peter E Wright. Structural studies of p21waf1/cip1/sdi1 in the free and cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences*, 93(21):11504–11509, 1996.

[97] Peter E Wright and H Jane Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, 293(2):321–331, 1999.

[98] A Keith Dunker, Celeste J Brown, J David Lawson, Lilia M Iakoucheva, and Zoran Obradovic. Intrinsic disorder and protein function. *Biochemistry*, 41(21):6573–6582, 2002.

[99] Peter Tompa. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25(9):847–855, 2003.

[100] Garegin A Papoian and Peter G Wolynes. The physics and bioinformatics of binding and folding—an energy landscape perspective. *Biopolymers*, 68(3):333–349, 2003.

[101] Garegin A Papoian. Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proceedings of the National Academy of Sciences*, 105(38):14237–14238, 2008.

[102] Vladimir N Uversky and A Keith Dunker. Understanding protein non-folding. *Biochim. Biophys. Acta*, 1804(6):1231–1264, June 2010.

[103] H Jane Dyson. Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys.*, 44(4):467–518, November 2011.

[104] Robin van der Lee, Marija Buljan, Benjamin Lang, Robert J Weatheritt, Gary W Daughdrill, A Keith Dunker, Monika Fuxreiter, Julian Gough, Joerg Gsponer, David T Jones, Philip M Kim, Richard W Kriwacki, Christopher J Oldfield, Rohit V Pappu, Peter Tompa, Vladimir N Uversky, Peter E Wright, and M Madan Babu. Classification of intrinsically disordered regions and proteins. *Chemical reviews*, 114(13):6589–6631, July 2014.

[105] Jennifer J Kohler, Steven J Metallo, Tanya L Schneider, and Alanna Schepartz. Dna specificity enhanced by sequential binding of protein monomers. *Proceedings of the National Academy of Sciences*, 96(21):11735–11739, 1999.

[106] Jiangang Liu, Narayanan B Perumal, Christopher J Oldfield, Eric W Su, Vladimir N Uversky, and A Keith Dunker. Intrinsic disorder in transcription factors. *Biochemistry*, 45(22):6873–6888, 2006.

[107] Monika Fuxreiter, Peter Tompa, István Simon, Vladimir N Uversky, Jeffrey C Hansen, and Francisco J Asturias. Malleable machines take shape in eukaryotic transcriptional regulation. *Nature chemical biology*, 4(12):728–737, 2008.

[108] Ruth S Spolar and M Thomas Record Jr. Coupling of local folding to site-specific binding of proteins to dna. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 263(5148):777–784, 1994.

[109] Peter H von Hippel. From "simple" dna-protein interactions to the macromolecular machines of gene expression. *Annual review of biophysics and biomolecular structure*, 36:79, 2007.

[110] Dana Vuzman and Yaakov Levy. Intrinsically disordered regions as affinity tuners in protein–dna interactions. *Molecular BioSystems*, 8(1):47–57, 2012.

[111] Paul C Whitford, Karissa Y Sanbonmatsu, and José N Onuchic. Biomolecular dynamics: order–disorder transitions and energy landscapes. *Reports on Progress in Physics*, 75(7):076601, 2012.

[112] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in biochemical sciences*, 33(1):2–8, 2008.

[113] Monika Fuxreiter. Fuzziness: linking regulation to protein dynamics. *Molecular BioSystems*, 8(1):168–177, 2012.

[114] Tanja Mittag, Lewis E Kay, and Julie D Forman-Kay. Protein dynamics and conformational disorder in molecular recognition. *Journal of molecular recognition*, 23(2):105–116, 2010.

[115] Alaji Bah, Robert M Vernon, Zeba Siddiqui, Mickael Krzeminski, Ranjith Muhandiram, Charlie Zhao, Nahum Sonenberg, Lewis E Kay, and Julie D Forman-Kay. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*, 519(7541):106–109, March 2015.

[116] Karl J Niklas, Sarah E Bondos, A Keith Dunker, and Stuart A Newman. Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Frontiers in cell and developmental biology*, 3:8, 2015.

[117] Tamaki Suganuma and Jerry L Workman. Signals and Combinatorial Functions of Histone Modifications. *Annual review of biochemistry*, 80(1):473–499, July 2011.

[118] Andrew J Bannister and Tony Kouzarides. Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–395, March 2011.

[119] Gabriel E Zentner and Steven Henikoff. Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*, 20(3):259–266, March 2013.

[120] Dinshaw J Patel and Zhanxin Wang. Readout of Epigenetic Modifications. *Annual review of biochemistry*, 82(1):81–118, June 2013.

[121] Brian D Strahl and C David Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, January 2000.

[122] Scott B Rothbart and Brian D Strahl. Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(8):627–643, August 2014.

[123] Sharon Pepenella, Kevin J Murphy, and Jeffrey J Hayes. A Distinct Switch in Interactions of the Histone H4 Tail Domain upon Salt-dependent Folding of Nucleosome Arrays. *The Journal of biological chemistry*, 289(39):27342–27351, September 2014.

[124] Geeta J Narlikar, Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell*, 154(3):490–503, August 2013.

[125] William L Hwang, Sebastian Deindl, Bryan T Harada, and Xiaowei Zhuang. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature*, 512(7513):213–217, August 2014.

[126] Sean D Taverna, Haitao Li, Alexander J Ruthenburg, C David Allis, and Dinshaw J Patel. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology*, 14(11):1025–1040, November 2007.

[127] Marta Garcia-Ramirez, Corinne Rocchini, and Juan Ausio. Modulation of chromatin folding by histone acetylation. *Journal of Biological Chemistry*, 270(30):17923–17928, 1995.

[128] Christin Tse, Takashi Sera, Alan P Wolffe, and Jeffrey C Hansen. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by rna polymerase iii. *Molecular and cellular biology*, 18(8):4629–4638, 1998.

[129] Philip J J Robinson, Woojin An, Andrew Routh, Fabrizio Martino, Lynda Chapman, Robert G Roeder, and Daniela Rhodes. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.*, 381(4):816–25, September 2008.

[130] Darren Yang and Gaurav Arya. Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation. *Phys. Chem. Chem. Phys.*, 13(7):2911–21, February 2011.

[131] Michael F Dion, Steven J Altschuler, Lani F Wu, and Oliver J Rando. Genomic characterization reveals a simple histone H4 acetylation code. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15):5501–5506, April 2005.

[132] Junbiao Dai, Edel M Hyland, Daniel S Yuan, Hailiang Huang, Joel S Bader, and Jef D Boeke. Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. *Cell*, 134(6):1066–1078, September 2008.

[133] Benedetta Dorigo, Thomas Schalch, Kerstin Bystricky, and Timothy J Richmond. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *Journal of molecular biology*, 327(1):85–96, March 2003.

[134] Jeanne Morinière, Sophie Rousseaux, Ulrich Steuerwald, Montserrat Soler-López, Sandrine Curtet, Anne-Laure Vitte, Jérôme Govin, Jonathan Gaucher, Karin Sadoul, Darren J Hart, Jeroen Krijgsveld, Saadi Khochbin, Christoph W Müller, and Carlo Petosa. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461(7264):664–668, October 2009.

[135] Jean-Louis Banères, Aimée Martin, and Joseph Parello. The n tails of histones h3 and h4 adopt a highly structured conformation in the nucleosome. *Journal of molecular biology*, 273(3):503–508, 1997.

[136] Xiaoying Wang, Susan C Moore, Mario Laszckzak, and Juan Ausió. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J. Biol. Chem.*, 275(45):35013–20, November 2000.

[137] DA Case, TA Darden, TE Cheatham III, CL Simmerling, J Wang, RE Duke, R Luo, RC Walker, W Zhang, and KM *et al.* Merz. Amber 12. *University of California, San Francisco*, 1(2):3, 2012.

[138] KE Van Holde. *Chromatin*. NY: Springer-Verlag, 1988.

[139] Doug Phanstiel, Justin Brumbaugh, W Travis Berggren, Kevin Conard, Xuezhu Feng, Mark E Levenstein, Graeme C McAlister, James a Thomson, and Joshua J Coon. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.*, 105(11):4093–4098, 2008.

[140] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1):141–151, 1999.

[141] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, 10(15):2073–7, April 2008.

[142] Jeffrey Skolnick, Andrzej Kolinski, and Angel R Ortiz. MONSSTER: a method for folding globular proteins with a small number of distance restraints. *J. Mol. Biol.*, 265(2):217–241, 1997.

[143] Feng Ding, Ramesh K. Jha, and Nikolay V. Dokholyan. Scaling behavior and structure of denatured proteins. *Structure*, 13(7):1047–1054, July 2005.

[144] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[145] Jianyin Shao, Stephen W Tanner, Nephi Thompson, and Thomas E Cheatham. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation*, 3(6):2312–2334, November 2007.

[146] Ignacia Echeverria and Garegin A Papoian. Structural Heterogeneity and Dynamics of the Unfolded Ensemble. *Israel Journal of Chemistry*, 54(8-9):1293–1301, August 2014.

[147] Albert H Mao, Scott L Crick, Andreas Vitalis, Caitlin L Chicoine, and Rohit V Pappu. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*, 107(18):8183–8188, 2010.

[148] Bowu Luan, Nicholas Lyle, Rohit V Pappu, and Daniel P Raleigh. Denatured state ensembles with the same radii of gyration can form significantly different long-range contacts. *Biochemistry*, 53(1):39–47, January 2014.

[149] Panagis Filippakopoulos, Sarah Picaud, Maria Mangos, Tracy Keates, Jean-Philippe Lambert, Dalia Barsyte-Lovejoy, Ildiko Felletar, Rudolf Volkmer, Susanne Müller, Tony Pawson, Anne-Claude Gingras, Cheryl H Arrowsmith, and Stefan Knapp. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell*, 149(1):214–231, March 2012.

[150] Pavel I Zhuravlev and Garegin A Papoian. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Quarterly reviews of biophysics*, 43(03):295–332, 2010.