ABSTRACT

Title of dissertation: UNCOVERING THE BIOPHYSICAL MECHANISMS OF HISTONE COMPLEX ASSEMBLY

Haiqing Zhao, Doctor of Philosophy, 2018
Biophysis Program

Dissertation directed by: Professor Garegin Papoian
Institute for Physical Science and Technology
Department of Chemistry and Biochemistry

At the most basic level, inheritance in living beings occurs by passing the genomic information such as the DNA sequences from the parent generation to the offspring generation. Hence, it is a fundamental goal for every generation to efficiently express the genomic information and safely pass it on to the next generation. In human and other eukaryotic species, this mission is mediated via chromatin, a macromolecule with intricate hierarchical structure. The fundamental unit of chromatin is called a nucleosome, a complex of histone proteins wrapped around with DNA. To carry out diverse biological functions such as transcription and DNA replication, the DNA-protein complex must dynamically transition between more compact, closed states and more accessible, open ones. To fully understand the chromatin structure and dynamics, it is essential to comprehend the basic structural unit of chromatin, nucleosome.

In this dissertation, I present my doctoral research in the exploration of the

nucleosome dynamics problem, focusing on the assembly process of histone proteins. From histone monomer to dimer, then to tetramer, octamer, and nucleosome, I used different computational modeling theories and techniques, together with different experimental collaborations, to investigate the overall thermodynamics and specific mechanistic details of nucleosome dynamics at different levels. My work has shed light on the fundamental principles governing the histone protein folding and histone complex assembly, in particular, highlighting similarities and differences between the canonical and variant CENP-A histones.

UNCOVERING THE BIOPHYSICAL MECHANISMS OF HISTONE COMPLEX ASSEMBLY

by

Haiqing Zhao

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
2018

Advisory Committee:
Professor Garegin A. Papoian, Chair/Advisor
Professor Yamini Dalal
Professor David Fushman
Professor Maria K. Cameron
Professor Jeffery B. Klauda
Professor Sergei Sukharev, Dean's Representative

© Copyright by
Haiqing Zhao
2018

To my grandma
for constant love, diligence, patience, and kindness.
May peace be around her forever.

Acknowledgments

More than eighty percent of my life has been spent in school so far. Facing towards the end, it is time to express my sincere thankfulness to all the people who have helped me along this long way.

First, I offer my sincerest gratitude to my PhD adviser, Prof. Garegin Papoian, whom I have ever worked with for the longest time. He accepted me as his Ph.D student and supported me throughout my whole doctoral training. I was deeply impressed by his curiosity and persistent pursue of natural science no matter what detailed subject or category the question may relate to. His broad knowledge across physics, chemistry, mathematics, and computer science always inspired me consciously and unconsciously. I often got ignited right after I discussed my questions with him. Plus, Garyk's mild-mannered character always made our conversation enjoyable. He taught me humbleness and respectfulness towards people but independence and critical thinking towards their sciences.

I give the same amount of thanks to Dr. Yamini Dalal as well, my co-adviser for the three-year-long UMD-NCI partnership program. I was fortunate to be selected and have the opportunity to work at NIH, the world-class research center for health sciences. From Yamini and other lab members I started to realize the importance of my theoretical and computational researches. The annual ethics case study at NIH told me how to be a serious and responsible scientist, beyond pursuing one's own curiosity. From here, I learned how to be a professional, mature, responsible researcher and scholar.

I am grateful to many other wonderful professors at Maryland, Profs. Maria Cameron and David Fushman for collaborating with me on two projects where I learned about applied mathematics and structural biology, Profs. Chris Jarzynski and Dave Thirumalai for teaching me fantastic statistic and ploymer physics, Profs. Sergei Sukharev, Dorothy Beckett and Jeff Klauda for providing me personal guidance about both science and career development. I thank all the administrative stuffs who helped me with their abundant patience for my administrative paperworks.

I am fortunate to have fantastic group members and friends during these years. I thank David Winogradoff, Longhua Hu, Ignacia Echeverria, Hao Wu, Yang Shen, Guang Shi, and Hongcheng Xu for many useful discussions and for all the fun we have had, Delphine Quenet, Minh Bui, Jon Nye, Daniel Melters and Tatini Rakshit for teaching me centromere biology. My time at college park was enjoyable also due to my fabulous roommates Ruiliang Bai, Xunnong Xu, Yuwei Cui, Brooke Okada, Derek Sober, Beth Domingo and the groups I attend which have become an important part of my life here, the Grace Fellowship (Peter Pan, Zheng Duan, Bobby Zhou, Yue Qin, Jenny Fang, Yongmu Huang, Yixiao Zhang), the Wallace Church, the Wallace music team (Stan, Eileen, Andy, Joyce, Wade), our music band *WHO IS FRANK*, and our *UMD-CSSA* soccer team. I cannot forget the scheduled list who volunteered to pick me up and take care of me during my ACL surgery. I want to express my heartfelt thanks to all them. They are treasures of my life. Thanks to my friend, now my wife, Sizhu(Alex) Li for her consistent encouragements, mental comforts and supports. It is my best luck to meet and marry her towards the end

of my student life.

I also thank Prof. Yanting Wang from the Chinese Academy of Science, Prof. Taiyu Zheng from the Northeast Normal University, and Prof. Uli Hansmann for introducing and encouraging me to pursue the direction of biophysics.

Lastly and most importantly, I would like to give the most thanks to my family, mom and dad, my dearest grandma, also aunt and uncle, for their unconditional love and consistent support throughout all the years. Studying abroad, I could not visit them often. However, without them, I cannot have the high-quality education, complete this degree and have the honor of being called Doctor.

Table of Contents

List of Tables

List of Figures

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
AWSEM	The Associative Memory, Water Mediated, Structure and Energy Model
bp	base pair
CENP-A	CENtromere Protein-A
CG	Coarse Grained
DNA	DeoxyriboNucleic Acid
FE	Free Energy
FM	Fragment Memory
FRET	Frster Resonance Energy Transfer
HA	Homologue Allowed
HE	Homologue Excluded
HJURP	Holliday Junction Recognition Protein
IDP, IDR	Intrinsically Disordered Protein, Intrinsically Disordered Region
PTM	Post-Translational Modification
WT	Wild Type
PDB	Protein Data Bank
RMSD	Root-Mean-Square Deviation
RMSIP	Root-Mean Square Inner Product
RMSF	Root-Mean-Square Fluctuations
R_g	Radius of gyration
COM	Center Of Mass
PCA	Principal Component Analysis
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
PMF	Potential of Mean Force
REMD	Replica-Exchange Molecular Dynamics
Å	Ångstrom (10^{-10} m)
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
TIP3P	Three-point explicit water model
WHAM	Weighted Histogram Analysis Method
IPST	Institute for Physical Science and Technology
UMD	University of Maryland at College Park
NIH	National Institutes of Health
NCI	National Cancer Institute

Chapter 1: Introduction of the Dissertation

The DNA of higher organisms associate with ions and proteins, forming chromatin to fit inside the nucleus. The chromatin dynamically transits between inactive state and more accessible state in order to orchestrate various DNA-templated processes such as transcription, translation, replication, and repair. The nucleosome, known as the fundamental structural unit of chromatin, consists of a segment of DNA wrapped around eight histone proteins. This dissertation investigates the assembly mechanisms of the underlying protein components, and the nucleosome itself, from a biophysics perspective by using computational modeling methodologies. Different simulation techniques are applied aiming to achieve the length and time-scales that are physiologically relevant to chromatin biology.

In this chapter, I first introduce the biology background followed by the structural aspects and related studies of histone complexes. In the third part of the introduction I will overview the protein folding problem, including the folding theory, after which molecular dynamics simulations and a protein coarse-grained model, called AWSEM, are introduced. Lastly, an overview of the subsequent chapters is outlined.

1.1 Chromatin and Centromere: the biology background of this study

The 21st century is claimed as the Century for Biology [8], marked by the outstanding breakthrough of the completion of the Human Genome Project in 2003. Sequencing of 99% of the euchromatic human genome with 99.99% accuracy has been reached [9]. How to better understand and use these big data pose a crucial long-term research challenge both in computational and experimental Life science. In the human cell, about two meters of DNA need to be tightly packaged into 23 pairs of chromosomes, in order to be less voluminous and fit inside the micron-diameter nucleus. A special region within the chromosome, or chromatin is centromere, the chromosomal loci where kinetochores assemble. Most eukaryotic centromeres are composed of repetitive arrays of non-coding DNA. During cell division, the spindle fibers attach to the centromere via the kinetochore and then pull out the sister chromatids to finish the chromosome segregation. To better understand the mechanisms of chromatin functions, we need to investigate its structure and dynamics.

Studying the chromatin structure is a complicated problem due to the size, its complex biological environment, such as the specific phase in the cell cycle. For example, in interphase, where cells spend most of its time, the cell is undergoing a period of growth which involves a high amount of DNA replication, gene transcription, and protein synthesis. During this time, the chromatin is relatively loosely packaged. However, during cell division, chromatin undergoes further condensation to form the chromosome ensuring that all the replicated genetic information will correctly pass into the new daughter cells. In general, chromatin organization has

three different levels: I. DNA wrapping around the central domain proteins, forming the nucleosome, the building block and the smallest protein-DNA complex in chromatin; II. Nucleosome arrays condensing into chromatin fiber; III. Higher level packaging of chromatin fiber into chromosome during cell division.

Despite being extensively investigated within many fields such as cell biology, molecular biology, biochemistry, and polymer physics during the last four decades, the structure of chromatin remains poorly understood. As mentioned above, chromatin conformations are dynamic, showing cell-cycle dependent structural organization. Most scientific controversies are focused on the level II, the chromatin fiber. In most of the cell life time (interphase), chromatin is a loosely-packed zig-zag array of nucleosomes ("beads-on-a-string"), — a single, long, 10-nm chromatin fiber —- where DNA duplication and transcriptions frequently occur [10–12]. At the molecular level, the fibers interact or undergo packaging through nucleosome-nucleosome and nucleosome-linker DNA contacts mediated by histone tails. This dissertation investigates the dynamics of nucleosomes, from the internal histone monomer, to different oligomers, then to the surrounding DNA stability, using a bottom-up approach aiming to provide the basic understanding of structural dynamics towards understanding the fundamental unit of chromatin organization.

1.2 Nucleosome & Histone Complexes

The research object of this dissertation, nucleosome and histone protein complex are introduced in this section. The nucleosome is the basic structural unit of

chromatin. A segment of DNA wrapped around eight histone protein cores constitute a typical nucleosome. However, histone proteins have diverse variants. These variations in histone primary sequence lead to changes in chromatin structure and dynamics, serving as the foundation of genomic regulation *in vivo*.

1.2.1 Canonical Histone

The core canonical histones making up the nucleosome are H3, H4, H2A, and H2B. Among the most evolutionarily conserved proteins, histone monomers share the same structural motif called "histone-fold". Each monomer has three long α-helices connected by two loops (Figures 1.1.A and 1.2), together forming a shallow U shape. Respectively, these helices and two loops are named as α1, α2, α3, L(oop)1, and L(oop)2. Two histone monomers form a dimer, which is the most fundamental histone structural unit *in vivo* and *in vitro*. All histone dimers follow the same "handshake" structural motif: the long α2 helix of each lays across that of its partner leading the three α-helices of one slotting into three α-helices of the other (Figure 1.1.A), via extensive hydrogen bonding and hydrophobic interactions between the two monomers. L1 and L2 help to hold the supporting helices together and provide sites for interaction with the DNA backbone. The rest of the binding interactions comes from the end of α1, either histone tails or αN helix in H3. The investigation of this general structural motif for histone proteins is discussed in chapter 2, where we found that most histone dimers have similar folding or binding mechanisms.

In the nucleosomal context, the two H3/H4 dimers form a histone tetramer,

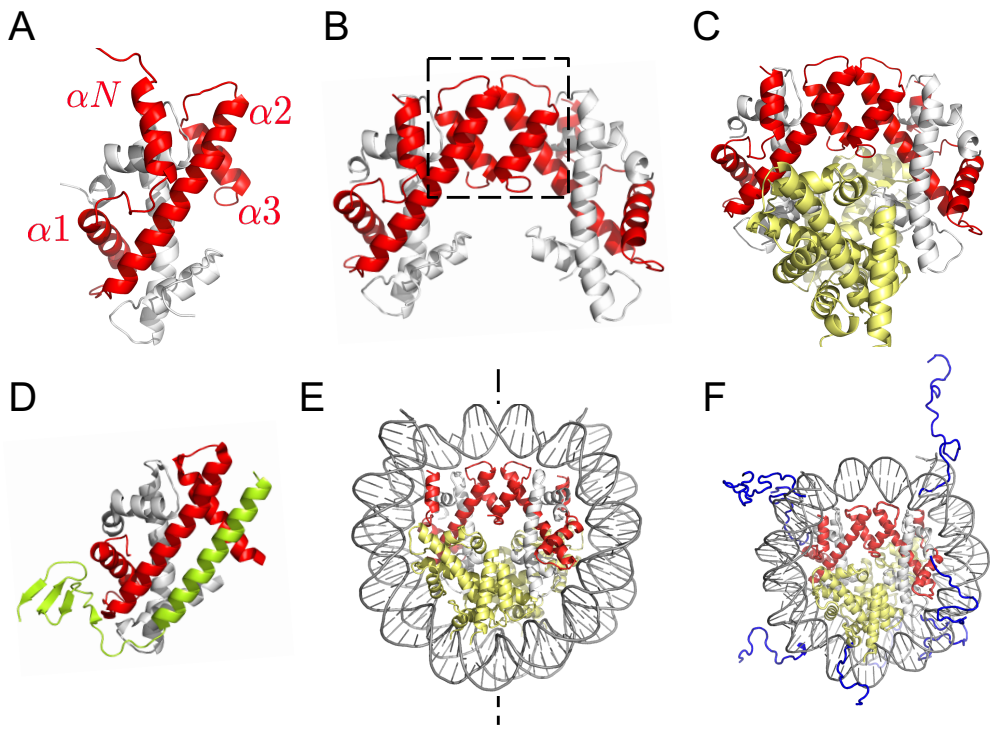

Figure 1.1: **Structure representations of histone dimer, tetramer, octamer, chaperone, as well as nucleosome and histone tails.** (A) A typical structure of histone dimer marks a handshake structural motif, consisting of three helices (α1, α2, and α3) from each histone. H3/H4 (red/white) dimer is shown here as an example. (B) Two H3/H4 dimers associate into a histone tetramer, interacting through a four-helix bundle (black box). (C) The $(H3/H4)_2$ tetramer is flanked on both sides by H2A/H2B dimers (yellow) to make an histone octamer. (D) Histone chaperons are usually invovled in histone assembly. Shown here is the structure complex of histone H3 variant CENP-A (red), H4 (white) and chaperone HJURP (green). (E) Structure of nucleosome is made of DNA (silver) wrapping around an histone octamer. The nucleosome is rotational symmetric about the pseudo-dyad axis, marked as a dashed line. (F) Histone tails (darkblue) are displayed in the context of nucleosome.

primarily interacting through the H3:H3 interface, a four-helix bundle with two helices from each H3 histone, close to the central base pair of DNA at the pseudo-dyad (Figure 1.1.E). The two dimers H2A/H2B interact with the tetramer at each side,

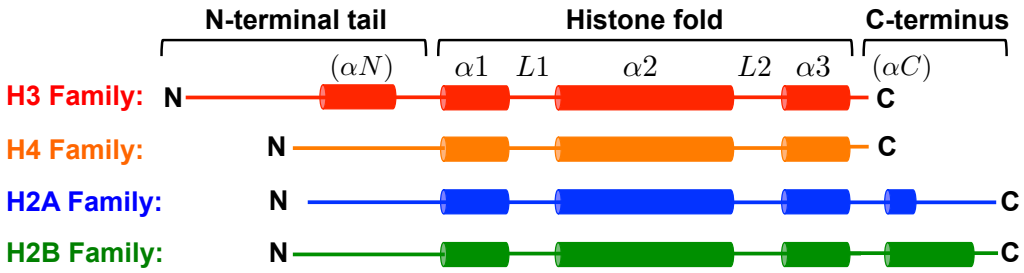

Figure 1.2: **A schematic representation of the histone structural motif and terminal tails.** Core histone proteins have three central α helices connected by two loops, known as the "histone-fold". Additionally, there is an αN helix in H3 and an αC helix in H2A and H2B respectively. Each histone features an N-terminal tail, and histone H2A includes a longer C-terminal region than other histones.

through the interface between H2B and H4, a similar four-helix bundle region as in the tetramer [2]. Histone tetramer, as the largest protein domain in the protein core, initiates nucleosome assembly and recognizes nucleosome positioning signals [13], serving as the structural basis for nucleosomal dynamics. A deeper understanding of histone tetramer dynamics could provide novel insights into nucleosome assembly and unveil possible pathways that have evolved to deal with the mechanical stress associated with chromosome segregation.

The 147 bp of nucleosomal DNA is approximately one persistence length of free DNA under physiological conditions (l_p ~50 nm), yet it wraps around the histone core in almost two superhelical turns, thus adopting a highly bent conformation. There are 14 regularly spaced major points of contact between protein and DNA, located once every 10.5 bp where the DNA minor groove faces towards the histone core, first observed in the original high-resolution crystal structure of the nucleosome [2]. Assembly of the nucleosome occurs in discrete steps *in vitro* depending on the

ion concentration [14], suggesting the possibility of intermediate dimer, tetramer, or hexamer structures *in vivo*.

Together with the intermediate subsets of histone complexes or nucleosome assembly, the variabilities in the histone and DNA sequences the structure and dynamics of chromatin, playing important roles in the regulation of gene expression *in vivo*. In this dissertation, I investigate the fundamental folding and binding mechanisms of the assembly process of nucleosome, from histone monomer, to histone dimer, tetramer, and to octamer and nucleosome.

1.2.2 Histone Variant and CENP-A

Histones are encoded by multiple copies of genes, producing variants with different amino acid sequences. Across all core histone families, the sequence identity of variants ranges from 50% to 99% with respect to the canonical histone sequence (Table 1). The replacement of canonical histones can affect nucleosome stability, creating functionally distinct chromatin domains [15]. Indeed, variation in histone primary sequence serves as the foundation of genomic regulation *in vivo* by leading to functional changes in chromatin structure and dynamics. During a typical cell cycle, canonical histones are mostly expressed during the interphase whereas histone variants can be expressed throughout the entire cell cycle [16]. This replication-independent nature of histone variants implies their unique functions other than being the DNA carriers.

Histone variants widely exist *in vivo* for most species. Among all types of his-

tones, each has a different number of sequence variants, with important implications for histone evolution [17] and nucleosome assembly dynamics. Interestingly, H4 is the only histone that does not have any reported sequence variant [18]. Whether or not the absence of histone variants for H4 reflects greater structural integrity remains unclear. To address this question, we studied the dynamics of histone H4 and compared it to that of its folding partner; this is the focus of chapter 3. Our work revealed that histone H4 adopts configurations closer to the native state than its partner (Figure B.8.A) [19], demonstrating the structural resilience that is predicted from its high sequence conservation and the absence of variants. Thus, H4 could provide a consistent reinforcing structural framework for histone dimers and higher order structures, while the H3 family, including canonical H3 and its variants, provides variability to the structure and function.

One important histone variant within the histone H3 family is the variant CENP-A (CenH3). Biologically, CENP-A specifies the unique location of the centromere, required for proper chromosome segregation during cell division [5, 20–23]. Recent studies reveal that CENP-A is overexpressed and mislocalized in more than 20 types of cancers and tumors [24]. However, structural dynamics regarding to the CENP-A related histone complexes remain obscure. Although the functions between CENP-A and canonical H3 differ distinctly, their crystal structures are almost identical. Researching the dynamics differences between CENP-A and H3 would be scientifically interesting and would help understand the relation between dynamics and function.

Among all the histone varaint types, CENP-A is of high interest in research

because the unique location and function of CENP-A during cell division make it a practicable candidate to study the structure-function relationship *in vivo*. Plus, the dominant structure of CENP-A-involved histone complex in different biology contexts remains disputable. The CENP-A octameric nucleosome are shown to be more flexible than the caonical histone complexes *in silico* and *in vitro* [6, 7]; the $(\text{CENP-A/H4})_2$ tetramer is more "rigid" than the $(\text{H3/H4})_2$ tetramer [21, 25] through the small-angle X-ray scattering experiments; however, the crystal structures of CENP-A- and H3-containing nucleosomes are virtually identical except three minor differences in the loop 1 region, C terminus, and the two ends of DNA [26–28]. Having these controversies, in chapter 3, we used a dual-resolution molecular dynamics methodology to investigate the dimer of H3/H4 and CENP-A/H4, to see whether, on the dimer level, the variant CENP-A demonstrates natural distinct dynamics from H3/H4, which might contribute to its unique biology *in vivo* [19]. By performing both all-atom and coarse-grained (AWSEM model [29]) simulations, the CENP-A/H4 dimer is more structurally variable than the canonical H3 dimer and that the centromere-specific chaperone HJURP stabilizes CENP-A and prevent the promiscuous mis-assembly of the CENP-A/H4 dimer. A similar conclusion is also suggested from our earlier nucleosome studies [7], where we discovered that CENP-A encodes enhanced distortability to the octameric nucleosome through the shearing contacts within the dimerization interface, which may allow for enhanced flexing of the histone core under mitotic tension.

The motivation for us to further study the histone tetramer (both canonical and variant) is not only based on the structural fact that the histone tetramer

is the central protein component composing the whole nucleosome, but also, for *in vivo* human cells, CENP-A nucleosomes are found to oscillate between stable tetramer and stable octamers, and the oscillation correlates with opening and closing of the CENP-A chromatin fiber [5]. Interestingly, *in vitro* deuterium exchange experiments, which measure the accessibility of protein backbones, indicate that the CENP-A histone tetramer is more stable, or rigid, than its canonical counterpart H3 tetramer [21, 25]. As discussed in chapter 4, we applied a coupling replica-exchange and umbrella sampling method to calculate the binding free energy between two histone dimers. The computed free energy uncovered distinctive free energy profiles for H3 and CENP-A tetramers. From the perspective of thermodynamics, CENP-A is more likely than canonical H3 to remain as a histone tetramer. Dynamically, CENP-A forms a more rigid tetramer with a fixed interface while canonical H3 tetramer displays a swelling motion around the binding interface. Furthermore, the addition of H2A/H2B onto the two sides of the tetramer has stabilization effect on the canonical histone but leads the CENP-A octamer to retain a memory of two states — one intrinsic compact state as in solo tetramer, the other similar to the familiar nucleosomal context. This finding provides a physical explanation why the variant CENP-A appears to be more globally dynamic than canonical H3 in octameric nucleosomes.

1.3 Energy Landscape Theory and AWSEM Model: the theory, model and methodology

In this section, I will introduce the problem of protein folding in general, the energy landscape theory with the folding funnel hypothesis, the molecular dynamics (MD) simulation and a unique model used in this dissertation, the Associative memory, Water mediated, Structure and Energy Model, AWSEM.

1.3.1 Protein Folding Problem and Folding Funnel Theory

Protein molecules consist of heterogeneous unbranched chains of amino acids. Through folding into a particular three-dimensional globular structure, proteins are able to fulfill their specific biological function. Given the numerous degrees of freedom from all the amino acid atoms, proteins can nevertheless spontaneously fold from random coils into an ordered, characteristic structure on short time scales, for example, seconds. However, a naive search model would predict astronomically long folding times, which is called the Levinthals paradox [30, 31]. This paradox is well resolved by the folding funnel energy landscape theory [32–34]. According to this theory, the conformation phase space of a protein with specific sequence, when projected on the energy and entropy variable shows a funnel-like energy landscape. Though the landscape may have some roughness with local minima, the protein is able to escape these local traps and transition downhill towards its native state, at the bottom of the funnel. A similar approach and theory can be used to describe

the protein-protein interactions. Usually, they are rough with shallow binding funnels [35–37]. Inspired by this theory, structure-based models have arisen that enable studying of the folding and binding dynamics of biomolecules. [38–40].

1.3.2 Molecular Dynamics Simulation and AWSEM Model

Molecular dynamics (MD) simulation was originally developed in the field of theoretical physics [41, 42]. Ever since the first molecular dynamics simulation of a macromolecule of biological interest was reported in late 1970s [43], this method has developed into a strong computational technique that can be effectively applied to study the dynamics of biomacromolecular systems. In the most common version, MD simulation trajectories describe the temporal evolution of a huge number of individual particles, typically atoms. The trajectories are determined by numerically solving the Newton′s equations of motion for a system of interacting particles, where their interactions are defined using a specific potential energy function called a force field. Hence, with detailed physical dynamics and chemical accuracy, MD simulations can provide deep insights into the fundamental mechanisms for complex, highly dynamical biological systems, whose details are thus far not accessible to either *in vivo* or *in vitro* experiments.

In this thesis, we undertook Molecule Dynamics simulations, with the associative memory, water mediated, structure and energy model (AWSEM) [44] to study the histone protein assembly process. AWSEM is a coarse-grained protein model where three beads are chosen to represent the positions of every amino acid.

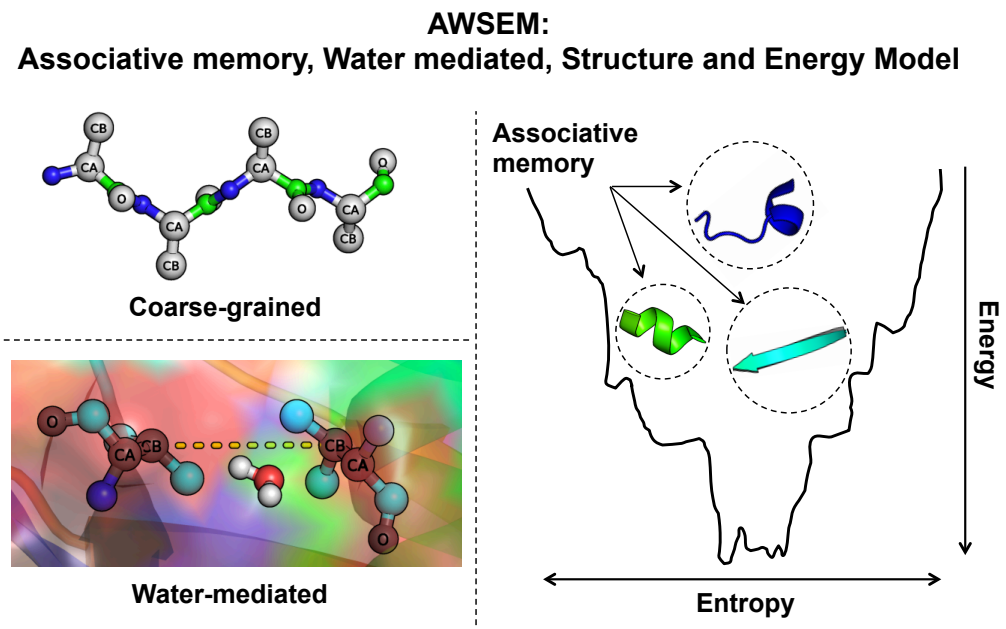

Figure 1.3: **Associative memory, Water mediated, Structure and Energy Model (AWSEM)** AWSEM force field uses three beads (C_α, C_β, O) to represent one amino acid, and features a water-mediated potential to describe the role of water in mediating protein-protein recognition. Associative memory (or fragment memory) is a local structure based term, bioinformatically aligning the target protein sequence towards short peptide fragments with known conformations.

Together with the implicit-solvent and fragment-memory features, AWSEM is computationally very efficient and has been used to perform de novo protein structure prediction [44] and protein-protein binding interface prediction [45].

$$V_{AWSEM} = V_{backbone} + V_{contact} + V_{burial} + V_{h-bond} + V_{AM} \tag{1.1}$$

In AWSEM, both physically motivated potential terms such as hydrogen bonding and bioinformatically based local structure biasing term are included. $V_{backbone}$ describes the detailed chemical connections to maintain the simulated chain protein-like (Figure 1.3). The physically motivated potentials $V_{contact}$, V_{burial}, and V_{h-bond} reflects different aspects of protein physics. $V_{contact}$ defines a tertiary interaction that acts between a pair of residues that are ten-residue or further apart in sequence. Dependent on the amino acid type, $V_{contact}$ also takes the local amino acid density into account. In the case of low and high local amino acid density, water-mediated [46] and protein-mediated interactions (in $V_{contact}$ term) are applied respectively. No other explicit or implicit water model is used in AWSEM. The burial term V_{burial} is another amino acid dependent term that reflects the hydrophobicity or hydrophilicity. V_{burial} specifies the preference of an amino acid to be buried inside or to be on the surface of the protein. V_{h-bond} term defines hydrogen bonding networks that are responsible for the formation of α helices or β hairpins. All the physical potentials are self-consistently optimized in a way that maximizes the ratio of the folding temperature to the glass transition temperature for the model, T_f/T_g which ensures the funneled energy landscape nature of protein folding. The bioinformatical term,

called fragment memory or associate memory potential, V_{AM} is a Gō-like [47] potential, but uses fragments of the target sequence. The fragment memory library is generated by aligning the target sequence to the online experimentally determined PDB information. In this research, the associate memory database is based on a single monomeric protein structure.

To overcome the local minima on the rough energy landscapes and obtain a sufficient sampling phase space which is ergodic, different enhanced sampling methods are used. Those methods include the replica-exchange method and the umbrella method. For the replica-exchange method, or called parallel tempering [48], different independent MD simulations are run in parallel carried out at different temperatures. Two replicas are randomly set to exchange if the probability of the exchanging replica satisfies the Metropolis Monte Carlo weight to the accepting replica which implies the accessibility between the two spaces. Through this, the energy barriers between different minima can be overcome so as to get a sufficiently sampled conformational space. The umbrella sampling method works through adding a biased energy term of a selected reaction coordinate [49] ξ, typically in a harmonic form $U(\xi(r)) = k(\xi(r) - \xi_o)^2$, along different positions of this coordinate axis, in order to cancel the influences of the potential energy barriers. The free energy profile is then computed after the removal of the bias, through the Weighted Histogram Analysis Method (WHAM) method [4].

1.4 Overview: the summary and the outline of the thesis

This dissertation investigates the assembly dynamics of histone complexes and the entire nucleosome thereof using different molecular dynamics techniques. We order the results chapters by length scale, focusing first on an individual dimer, then tetramer, octamer, and nucleosome. Each chapter consists of the introduction, methods, results, a discussion section where we describe the potential biological importance of our findings and the conclusion.

In Chapter 2, we use AWSEM to examine the protein folding and binding mechanism for histone dimer, the elementary protein structural unit of the nucleosome. Our simulations and NMR experimental data show that all the examined histone dimers, H3/H4, CENP-A/H4 and H2A/H2B, show a similar folding/binding mechanism.

In Chapter 3, we employ a dual-resolution MD methodology to investigate the conformational dynamics difference of the H3/H4 and CENP-A/H4 underlying their distinguishable biology functions. We found that CENP-A/H4 dimer is significantly more dynamic than H3/H4, and chaperone HJURP prevents this promiscuous misassembly of CENP-A, serving both as a folding and binding chaperone.

In Chapter 4, we use the coupled replica-exchange and umbrellas sampling, and constant temperature simulations, to characterize the thermodynamical features and the dynamics details for the canonical and variant CENP-A histone tetramers. The calculated binding free energy of histone tetramer and the analyses of the long time-scale MD simulation together suggest that CENP-A forms a more stable

histone tetramer which explains the controversies between previous computational studies and experimental observations. Finally, we discussed the biological meaning of this research by proposing two different histone assembly ways.

Chapter 2: The Dynamics and Thermodynamics of Histone Folding

The chapter is based on the unpublished work of the author: H. Zhao, H. Wu, D. Abeykoon, A. Guseman, D. Fushman and G. Papoian; (2018)

2.1 Introduction

In eukaryotic cells, histone proteins organize the genomic DNA into chromatin with the basic packaging subunit nucleosome. The histone proteins of the core nucleosome are organized in a form of octamer, comprised of four pairs of dimers. Two pair of dimers form a tetramer [2]. During decades of studies, no functional histone monomer or homodimer were ever reported or observed. Hence, the histone dimer, specifically the heterodimer, is known to be the smallest protein unit in eukaryotic chromatin. Besides these core canonical histones, variant histones have evolved for diverse biological functions. Interestingly, despite their function and sequence diversities, all histones are characterized with the same structural motif, known as the histone-fold, where two monomers fold into a handshake motif to form a dimer [2]. Two monomer components, each consisting of a helix-loop-helix frame, dimerize in an intertwined, head-to-tail manner.

Extensive studies in biochemistry, biophysics and cell biology have centered on

the structure and function of the canonical and variant histone nucleosomes, interrogating the relationships between the sequence, structure, and function of histones in various biological backgrounds [5,23,50,51]. Limited work has been done towards understanding of histone protein folding dynamics. Previous experiment showed that unfolding of histone H2A/H2B is a two-state transition [52]. However, detailed folding mechanisms of the histone dimer or the histone-fold remain unknown. Investigating the histone dimer from the perspective of protein folding is essential for better understanding of the higher level organization of histone proteins, such as histone tetramer, octamer and nucleosome. Plus, because histones are among the most conserved eukaryotic proteins, a clear understanding of their dynamics may shed light on the evolutionary origin of the histone fold.

AWSEM is a coarse-grained protein force field used in Molecular Dynamics (MD) simulations [29]. Built on the funneled free energy landscape theory [33,34], AWSEM includes both physical interaction terms and bioinformatics-inspired terms. AWSEM has been successfully used to predict single protein structures [29] as well as binding interfaces between proteins [19,45]. Recent works in AWSEM have expanded to study multiple protein assembly including protein aggregation [53,54], membrane protein folding [55,56], and protein-DNA association [57–59].

In this work, we used AWSEM MD simulations together with NMR experiments to address some of those longstanding biophysical puzzles regarding histone dimer and monomer. Taking the fragment memories from the wild protein databases in AWSEM, we tried to predict the structure of histone monomer and found that all of them tend to fold into a collapsed state that is far from the native confor-

mation. However, in the presence of the other partner histone, two histones fold into a native-like dimer. This observation is supported by NMR data. Altogether, this observed process reveals a coupled folding and binding mechanism of histone dimer formation and also asymmetrical dynamics roles of two composing histones. Through enhanced sampling simulations, we estimated the corresponding folding free energy. Finally, we discuss the histone folding mechanism from the perspectives of polymer physics and evolutionary biology. As a concluding remark, we propose that histone-fold based proteins may inherit similar structural motif and folding mechanics, however, their components could contribute differently towards to the structural dynamics thereby to support diverse functions in vivo.

2.2 Methods

2.2.1 AWSEM-MD simulation

To explore the folding mechanism of histone monomers, we first performed molecular dynamics simulations with the AWSEM model on all the four histone core proteins, namely H3, H4, H2A and H2B. In this study, the parameters in AWSEM model were tuned such that the simulated melting temperature of histone dimers is around 350 K, as observed in experiments. In addition, we employed an AWSEM-featured bioinformatic term called "fragment memory", using available protein segments as the local structural bias. In histone monomer folding simulations, the biasing segments were selected from proteins which share similar local amino acid sequences to the histone monomers. According to the criteria that

whether homologues of the simulated histone (including histone itself) are included or not, two versions of "memory" library were built respectively, "homologues allowed (HA)" and "homologues excluded (HE)". In histone dimer simulations, the local memory fragments were selected from the crystal structure of the nucleosome (PDB: 1OAI), which still only bias local structure. The length of a fragment is typically from 3 to 9 residues. Hence, only local structural bias are taken.

To better characterize the thermodynamics feature of histone protein folding, we carried out coupled replica-exchange and umbrella sampling simulations to collect sufficient conformational statistics for estimating folding free energies. We first set 10 umbrella windows linearly distributed along the chosen collective variable Q_{dimer}. At every umbrella window of Q_{dimer}, 10 different temperature replicas were run in parallel. The replica that is close to the folding temperature was then collected from every umbrella window, following which WHAM was compute the unbiased free energies.

We run AWSEM simulations using the open-source molecular dynamics software, LAMMPS (Oct 2012 version), with non-periodic shrink-wrapped boundary condition and the Nose-Hoover thermostat. The simulation timestep was set as 5 femtoseconds. The native conformations were taken from the crystal structure of nucleosome (PDB ID: 1AOI), excluding the disordered N-terminal and C-terminal tails. All annealing simulations started from the completely unfolded state, and then were slowly cooled down from 600 K to 200 running 1×10^7 steps. Ten separate simulations were carried for each system.

2.2.2 Experimental details

Protein expression and purification: H2A and H2B, unlabeled or ^{15}N labeled, were expressed in e.coli and purified from inclusion bodies using cation exchange chromatography. Their correct mass was confirmed by mass spectrometry. The plasmids were a generous gift from Dr. Tingting Yao.

NMR measurements: All NMR experiments were performed at 23°C on Bruker Avance-III NMR spectrometer equipped with TCI cryoprobe. Proteins were dissolved at 100-200 uM in 20 mM sodium phosphate buffer (pH 6.8) containing 7% D2O and 0.02% NaN_3. NMR data were processed using TopSpin (Bruker Inc.)

2.3 Results

2.3.1 Histone Monomers Cannot Fold by Themselves

To quantitatively describe the propensity for folding of histone monomers, we used the Q measure, which indicates the degree of similarity between simulated and native structures:

$$Q = \frac{1}{N_{pairs}} \sum_{i<j-2} exp[-\frac{(r_{ij} - r_{ij}^N)^2}{2\sigma_{ij}}] \tag{2.1}$$

where N_{pairs} is the number of pairs in the summation, r_{ij} is the instantaneous distance between C_α atoms of residues i and j, r_{ij} is the same distance in the native structure, and $\sigma_{ij} = (1 + |i - j|)^{0.15}$ represents the resolution of the distance difference. The range of Q is from 0 to 1, the higher value indicating that a simulated

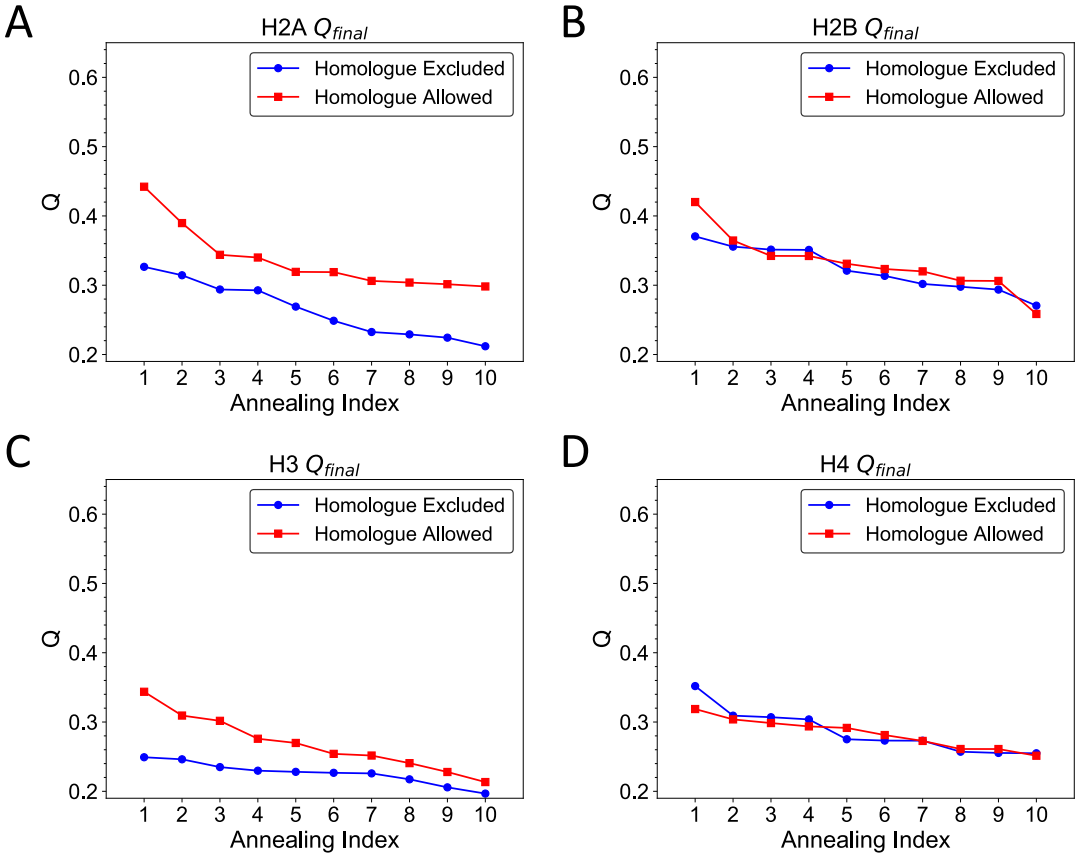

Figure 2.1: **Low Q values indicate histone monomers cannot fold to native structures independently.** Q values of (A): H3; (B): H4; (C): H2A; (D): H4 are shown in ten individual annealing simulation runs in a descending order. Results with "homologue excluded" and "homologue allowed" fragment memory databases are represented as blue circles and red squares.

conformation is more similar to the native structure.

We calculated the Q values of the last snapshot in each simulation trajectory and ranked them in a descending order (Figure 2.1). In the simulations with "homologues excluded" structural bias, the Q values for the four histone monomers are all around or below 0.35 (blue circles). When the homologues are allowed, although the Q values increase in general, they are still below 0.45 (red squares). By contrast, with almost identical setup, more than ten globular protein structures can be precisely predicted by AWSEM simulations, with Q values up to 0.65. This comparison indicates that histone monomers are not well folded as native-like in our simulations. A further exploration of the simulated structures shows that they do not form the stereotypical hand-shake motif as in dimer or tetramer states in the crystal structures, despite several similar α-helices. Figure 2.2 shows that for histone H3 and H2A, the distances between the $\alpha 1$ and $\alpha 3$ (black circled) are less than those in the native conformation, forming a closed state, which may block the binding interface with its binding partners.

2.3.2 Dynamics of the Histone Dimer Folding

Given the above monomer results, we next performed AWSEM annealing simulations to predict the binding interface of histone dimers. We would like to note that the only structural information used in these simulations was the local backbone conformations of histone monomers from the protein data bank. Hence, no structural bias was provided for the tertiary contacts within each monomer and

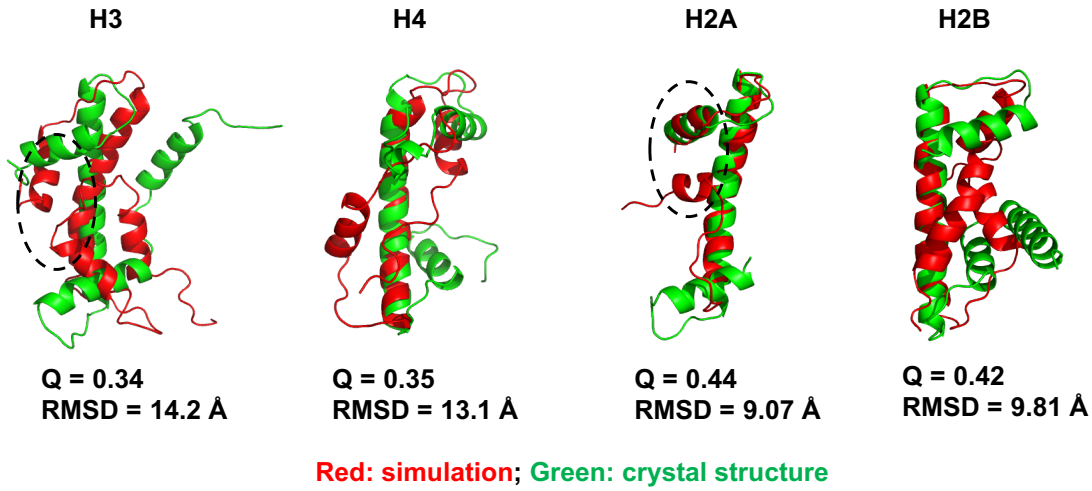

Figure 2.2: **Comparison between the predicted and native structures for H3, H4, H2A, and H2B.** Structures that have the best Q value were chosen as the final predicted structure (red). They are aligned with the native conformations (green) from the crystal structure (PDB: 1AOI).

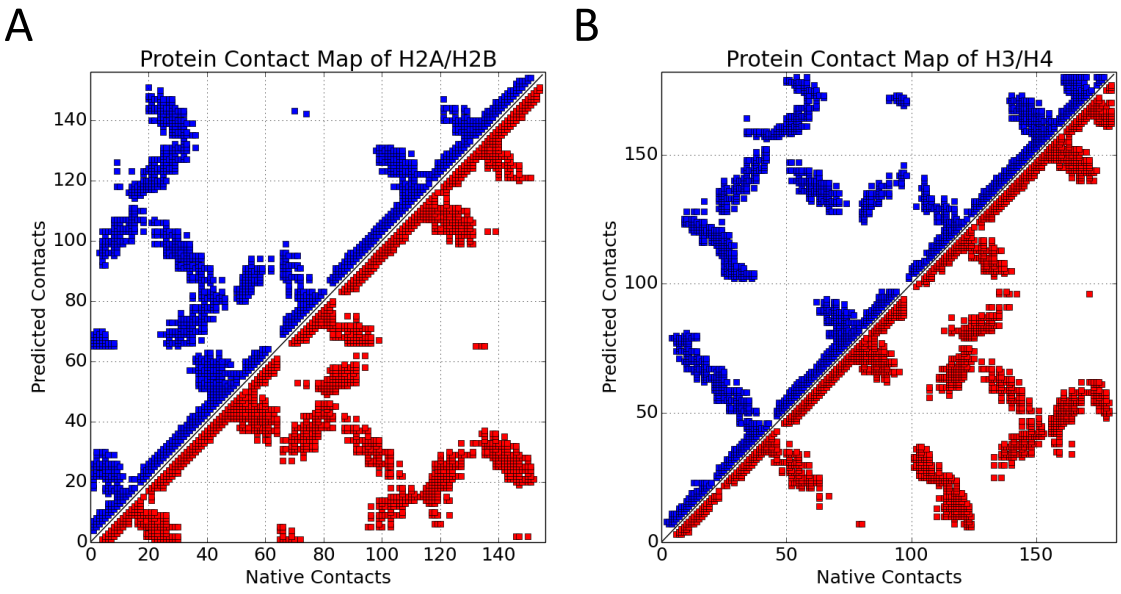

Figure 2.3: **Contacts comparison between the contacts in the simulation predicted conformation (blue) and that of the native state structure (red).** Contacts map are plotted along the native contacts and predicted contacts for H2A/H2B (A), and H3/H4 (B).

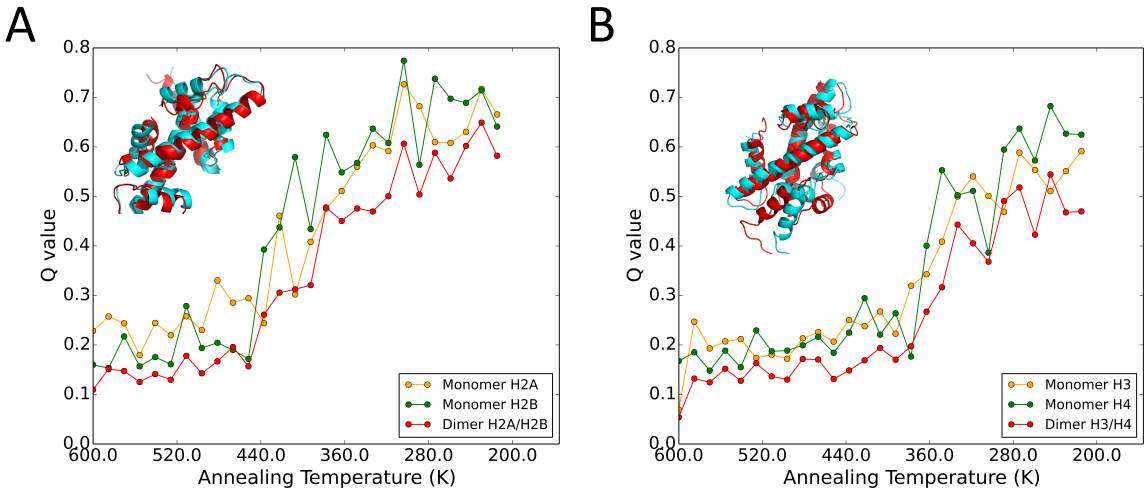

Figure 2.4: **Histone monomers help each other to fold** (A) Q values analysis for H2A/H2B annealing simulations shows that the monomer H2A (orange), H2B (green), and histone dimers (red) of fold simultaneously as the annealing temperature is cooled. (B) H3/H4 annealing simulations also shows a simultaneous folding and binding process between H3 (orange), and H4 (green) monomer, resulting in the dimer H3/H4 (red). The final folded dimeric conformations of H2A/H2B, H3/H4 (red) are aligned to the corresponding crystal structures (cyan).

between the two monomers. Following this strategy, we performed simulations on H2A/H2B, and H3/H4 dimers, respectively, starting from random coils states for each monomer. Through simulated annealing runs, the temperature was lowered from 600 K to 200 K.

First, the contacts of the predicted dimer conformation are analyzed compared to those of the native structure. Both the contacts within each monomer and that between two monomer partners have been successfully predicted for H2A/H2B (Figure 2.3.A) and H3/H4 (Figure 2.3.B). A few contacts are missing but in all, 98% of H2A/H2B are found in the predicted conformation of H2A/H2B, and 96% of the native contacts of H3/H4 were correctly predicted.

To further investigate the folding mechanisms, the structural overlaps of monomers

H2A, H2B (Figure 2.4.A), H3, H4 (Figure 2.4.B), and dimers (Figure 2.4) with the corresponding native states of monomers and dimers were calculated and plotted as a function of annealing temperature. These plots show that around 360 K, there is an obvious transition, wherein Q_{dimer} value goes from 0.3 to 0.5 for H2A/H2B, from 0.2 to 0.45 for H3/H4. The transitions in interfacial coordinates dimer occur with the similar transition of composing monomers, indicating that the two monomers fold and bind spontaneously for both H2A/H2B dimer (Figure 2.4.A) and the H3/H4 dimer (Figure 2.4.B). Interestingly, the roles of two composing monomers are not symmetrical in terms of these contributions to dimer's folding. For instance, H2B (green dots in Figure 2.4.A) is better folded than H2A during the folding trajectories. This analysis is consistent with the above monomer prediction simulations, where, on average, the final Q of H2B is better than the final Q of H2A (Figure 2.1.C and 2.1.D) Similarly, H4 maintains relatively more native-like conformation than H3 (Figure 2.4).B). This latter observation is not surprising based on our previous finding that histone H4 preferentially maintains native-like stability and in the presence of various binding partners [19].

2.3.3 Thermodynamics of Histone Dimer Folding

To further characterize the thermodynamical features of histone protein folding, we carried out coupled replica-exchange and umbrella sampling simulations. The calculated free energies are projected onto various 1D and 2D order parameters. As shown in Figure 2.5, the free energies of H2A/H2B are plotted as a function

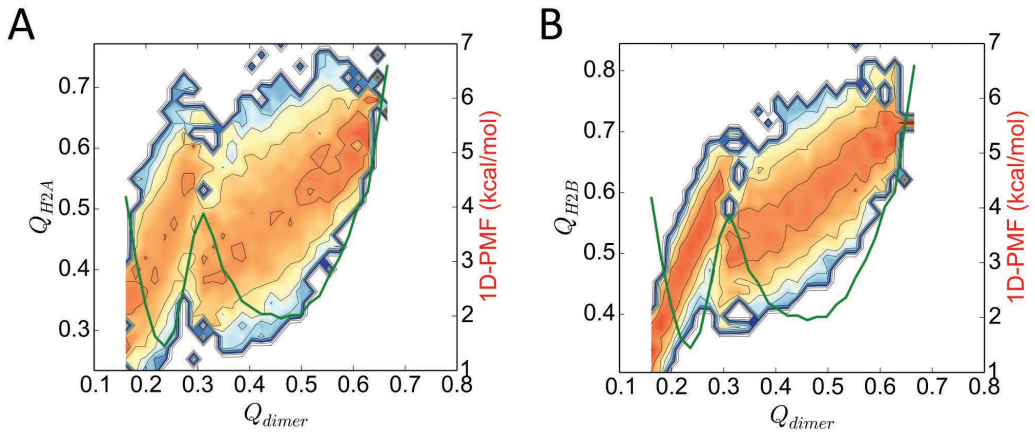

Figure 2.5: **1D and 2D Free energy profiles of histone dimer folding along the reaction coordinates of Q values of dimer and monomers of H2A (A) and H2B (B).**

of Q_{dimer} (green curve) and as a function of Q_{H2A} and Q_{dimer} (Figure 2.5.A), and Q_{H2B} and Q_{dimer} (Figure 2.5.B). From the 1D free energy curve, it is clear that the simulated H2A/H2B dimer has two states: the unfolded state referred at Q_{dimer} of 0.2 and the folded state at Q_{dimer} of 0.5. The energy barrier between these two states is about 4 kcal/mol, located at Q_{dimer}=0.3. This result is consistent with our simulated annealing simulations, where the folding transition also occurs around Q_{dimer}=0.3 for H2A/H2B (Figure 2.4).

Furthermore, on the 2D surface, the saddle point between the two minima occurs at $Q_{monomer}$ of 0.5. Indeed, if histone monomers are well folded as $Q_{monomer}$ equal to 0.6 $\sim$ 0.7, the saddle region would not apply. Instead, there is an energy barrier which a well-folded monomer need to overcome to get to the folded state. Again, the two monomers contribute differently to the thermodynamics of histone dimer. As seen in the 2D FE surface, both of the two energy minimums of histone H2B are deeper than those of H2A. Instead, the free energy landscape of H2A is rel-

atively more frustrated. This finding suggests that H2B has a more clear transition between the unfolded and folded state, compared with H2A. In other words, our thermodynamics analysis indicates that during binding/folding of H2A/H2B, H2A dynamics may be more glass like compared with smoother folding of H2B.

2.3.4 Experimental confirmation by NMR

We then carried out NMR measurements to support our simulation data. ^{1}H-^{15}N NMR spectra of H2A and H2B alone (Figure 2.6.A, E) show a narrow spread of NMR signals resulting in signal crowding in the region typical for amide signals of unstructured/unfolded proteins. The negative or close to zero signal intensities observed in the heteronuclear NOE spectrum of ^{15}N-labeled H2A recorded upon pre-saturation of amide protons (e.g., Figure 2.6.C for H2A) are a clear indication that the protein is unstructured and highly flexible. Upon addition of unlabeled H2B we observed a dramatic change in the ^{1}H-^{15}N NMR spectra of ^{15}N-labeled H2A, wherein new signals (corresponding to the bound state) appear and increase in intensity until they saturate at ca. 1:1 H2B:H2A molar ratio (Figure 2.6.B). Concomitantly, the unbound signals reduce in intensity and practically disappear at the saturation point. This behavior of the NMR signals which exhibit essentially no shifts indicates that the binding is in slow exchange regime on the NMR chemical shift time scale. In contrast to the unbound state, the signals of ^{15}N-labeled H2A in complex with H2B show a significant spread, indicating that the bound state of H2A is well structured. Also many H2A signals in the heteronuclear steady-state

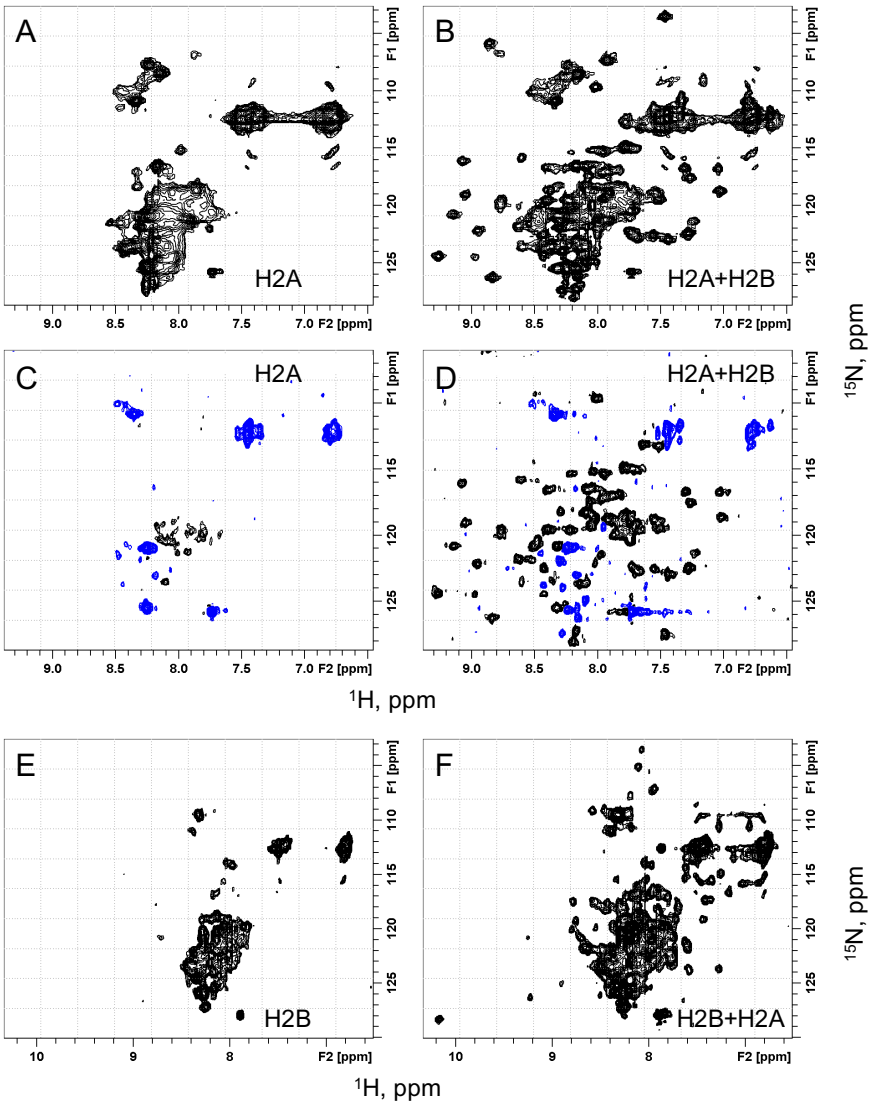

Figure 2.6: **NMR studies of H2A and H2B upon complex formation. (A-B) ^{1}H-^{15}N SOFAST-HMQC spectra of 15N-labeled H2A alone (A) and in the presence of unlabeled H2B at a 1:1 molar ratio (B). (C-D) Heteronuclear steady-state ^{15}N{^{1}H} NOE spectra recorded with amide proton presaturation (ref.) for ^{15}N-labeled H2A alone (C) and in the presence of unlabeled H2B at a 1:1 molar ratio (D). In these spectra contours with positive intensities are colored black while negative intensities are blue. (E-F) ^{1}H-^{15}N SOFAST-HMQC spectra of ^{15}N-labeled H2B alone (E) and in the presence of unlabeled H2A at a 1:1 molar ratio (F).**

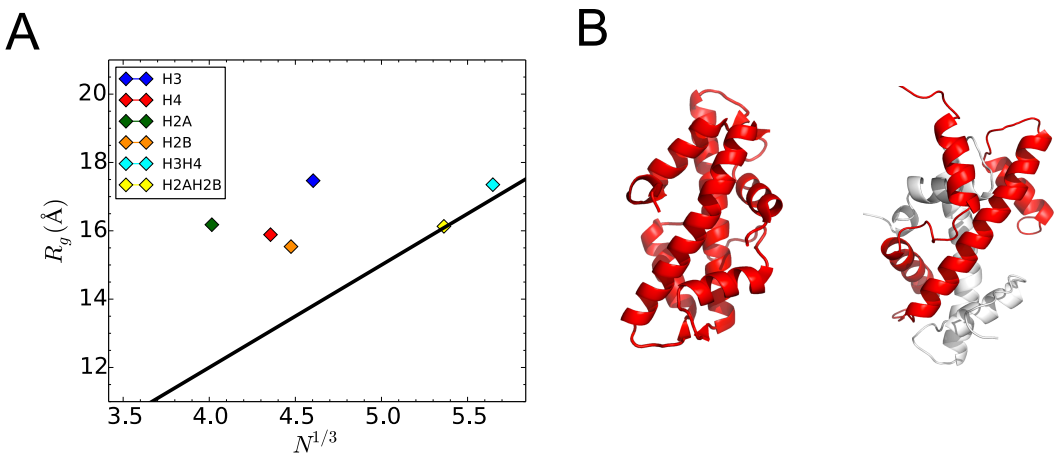

Figure 2.7: R_g **of histone monomers do not obey the Flory scaling law whereas the histone dimer does; an ancestral histone in archaea shows a monomeric chain.** (A) The ratio of R_g over the 1/3 scaling power of residue number N is plotted for histone monomer H3, H4, H2A, H2B and H3/H4, H2A/H2B with different colors respectively. The black line is the best linear fit for 403 monomeric protein datasets with a correlation of 0.9 [1]. (B) The archaea *Methanopyrus kandleri* histone (left) folds as a monomeric chain while eukaryotic histone displays a dimeric structure (right, red marks H3 and white marks H4).

NOE spectra recorded at these conditions have positive intensities, characteristic of a well-folded state of the protein (see e.g. [60]). A similar behavior was observed for ^{15}N-labeled H2B, which is unstructured in the unbound state and folds upon complex formation with H2A (Figure 2.6.E, F).

2.4 Discussions

2.4.1 Polymer scaling law

Our results indicate that histone dimers H2A/H2B and H3/H4 have similar folding and binding mechanisms, namely, binding coupled to folding. which is a coupled folding and binding process, meaning that monomers cannot be well folded

on its own but can be folded together with the other histone partner. Using the simulation methodology, we are able to provide new dynamics and thermodynamics details characterizing the different roles of the two composing histone monomer towards folding a dimer. In the following, we discuss our results from polymer biophysics perspective and the potential biological meaning of our findings.

For many classes of polymer, the radius of gyration for a polymer chain approximately follows the scaling relation: $R_g \sim N^{\nu}$, where R_g is the radius of gyration of the polymer. N is the number of bond segments (equal to the degree of polymerization) of the chain and ν is the scaling exponent [61]. Dima and Thrumalai studied the radii of gyration of 403 monomeric proteins [1]. The plot of R_g, for their dataset of proteins, as a function of protein's chain length. N follows $R_g \sim \alpha N^{1/3}$ with $\alpha \simeq 3$ with a correlation coefficient of 0.90. With this in mind, we calculated the R_g of the crystal structures of histone proteins, both monomers and dimers, computed the corresponding scaling factors, and compared them with the empirical values from [1]. As shown in Figure 2.7.A, all histone monomers have a higher scaling factor α than the average of all folded proteins (the black line). However, the scaling factors of the histone dimers give a close fit to the empirical function of monomeric proteins. Together with the geometry of histone-fold structural motif, where the three helices of one histone cross and clot into another three helices from the other, this analysis supports the point that structurally, histone dimer represents a single folding unit.

2.4.2 Histone evolution

Indeed, from the view of histone evolution, it is found that an ancient archaea, *Methanopyrus kandleri*, produces a novel, 154-residue histone (HMk) (Figure 2.7.B) which is homologous to the eukaryotic histone heterodimers, sharing the similar histone-fold structural motif [62, 63]. It is possible that the eukaryotic histones inherited the main structural motif and folding mechanism from their ancestor protein but diversified into two different partners for each protein unit, allowing for more possibilities to execute diverse biological functions needed for higher organisms. These possibilities may include but not limited to distinctive post-translational modifications on each monomer and different structural and functional roles of the two composing partners as revealed here and in our previous work [19]. Furthermore, there are a group of proteins containing the histone-fold motif and most of them are involved in a wide variety of functions related mostly to DNA metabolism [64]. Based on the analyses presented in this work, we propose that while these histone-fold based proteins may have a relatively small degree of sequence similarity, their histone-fold structure implies that they may have a similar folding mechanism as the histone dimer.

2.5 Conclusion

To conclude, in this work we have applies MD simulations and NMR experiments to study the folding and binding mechanisms of histone dimers. We quantified the dynamics and thermodynamics of histone dimer folding, demonstrating coupling

of binding and folding. The histone monomers need to be partly folded in order to form a dimer. In addition, two composing histone monomers asymmetrically contribute to the binding process, with smoother and morris rugged associate energy landscapes. Lastly, based on our analyses, we propose that the folding mechanism uncovered in this work may be operational for other proteins with the histone fold structure.

Chapter 3: Promiscuous Histone Mis-assembly is Actively Prevented by Chaperones

The chapter is based on the published work of the author: H. Zhao, D. Winogradoff, M. Bui, Y. Dalal, and G. Papoian; J. Am. Chem. Soc. 138 (40) (2016) [19]

3.1 Introduction

In eukaryotes, genomic DNA associates with histone proteins, assembling into arrays of nucleosomes. The canonical nucleosome contains 147 base pairs of DNA, wrapped around the histone octamer core with two copies each of the histones H2A, H2B, H3, and H4. [2] These core histones are among the most conserved proteins in eukaryotes, and all feature the same structural motif, known as the "histone-fold." [65] However, recent studies revealed that variant histones have evolved for diverse and specific functions. [15, 50, 66–68] Extensive studies in cell biology, biochemistry and biophysics have interrogated the relationships between the sequence, structure and function of histone variants in various biological contexts. [15, 50, 51, 66–69] Indeed, variation in histone primary sequence serves as the foundation of genomic regulation *in vivo* by leading to functional changes in chromatin structure and dynamics. [70, 71] In contrast to all the other core histones,

there are no reported variants of H4 [18]. Whether the absence of histone variants for H4 reflects greater structural integrity remains unknown, and addressing this question may shed light on the structural foundation of genetic inheritance.

Within the H3 family, the variant CENP-A (CenH3), specifies the unique location of the centromere, required for proper chromosome segregation during cell division. In particular, CENP-A is reported to be overexpressed and mislocalized into noncentromeric chromosome regions in aggressive cancer cells. [72, 73] Interestingly, the crystal structures of CENP-A and canonical H3 are nearly identical, except for minor differences in CENP-A's αN helix, and loop 1 regions [3, 26]. However, *in vivo* CENP-A-containing nucleosomes have been shown to occupy a multitude of structures. [5, 20, 25, 74–88] Our recent all-atom molecular dynamics study revealed that the octameric CENP-A nucleosome displays more structural heterogeneity on a local and global scale than its H3 counterpart [7], a result that has since been experimentally validated by FRET assays demonstrating CENP-A octameric nucleosomes *in vitro* are highly flexible, [6] in contrast to previous reports that the CENP-A nucleosome is rigidified [21, 25] *in vitro*. Since the CENP-A dimer is the key component distinguishing the CENP-A nucleosome from the canonical H3 nucleosome, we were curious whether, in isolation, or coupled to its chaperone HJURP (Holliday Junction Recognition Protein), the CENP-A/H4 dimer displays dynamics distinct from that of H3/H4, which might, in turn contribute to its unique biology *in vivo*.

Investigating the dynamics of histone variant deposition into and eviction from nucleosomes is fundamentally important, with chaperones like HJURP playing a key

role in facilitating and regulating histone delivery, exchange and removal. [89, 90] The chaperone HJURP has been demonstrated to be required for the deposition of CENP-A into the kinetochore, [91–93] but precisely how HJURP dynamically interacts with CENP-A/H4, and how HJURP mediates CENP-A's deposition through these interactions remain unclear.

To address the questions above, one could rely on molecular simulations of the CENP-A/H4 and H3/H4 dimers, and also the ternary complexes with HJURP. Usually, either atomistic or coarse-grained simulations are chosen for such studies, where the former provides finer resolution but samples less conformational space, raising issues of convergence for systems of this size. Coarse-grained simulations, on the other hand, quickly achieve equilibration, however, detailed atom-by-atom structural interactions are averaged over. In this work, we studied the same systems employing a novel dual-resolution approach, using both coarse-grained AWSEM [44] (CG-AWSEM) and all-atom molecular dynamics (MD) simulations. These two techniques complement each other: CG-AWSEM MD (*i.e.* three beads per amino acid residue) in implicit solvent samples more conformational space and explores more global properties of the histone dimers, whereas all-atom MD in explicit solvent probes specific interactions and native-state dynamics at high resolution. One of the overarching goals of our work was to cross-validate the conclusions obtained from these two independent methods, analyzing consistent findings or discrepancies in some detail.

Both CG-AWSEM and all-atom results indicate that histone H4 adopts configurations closer to the native state than either CENP-A or H3, demonstrating

the structural resilience that is predicted from its high sequence conservation and the absence of variants. The CENP-A/H4 dimer is more structurally variable than the canonical H3/H4 dimer in CG-AWSEM simulations, wherein the dimer interface of CENP-A/H4, in particular, exhibits greater conformational heterogeneity. A key component that distinguishes the dynamics of CENP-A/H4 from H3/H4 is the longer and more acidic C-terminal residues of CENP-A, which, in our simulation results, is surprisingly regulated by its chaperone HJURP. In all-atom MD simulations, we observe that HJURP facilitates the formation of a structure-inducing electrostatic network with the C-termini of CENP-A and H4, and that the N-terminal portion of CENP-A containing S68 forms key interactions with a hydrophobic pocket of HJURP. To test the hypothesis that CENP-A S68 is required for binding with HJURP, we performed *in vivo* experiments and all-atom simulations mutating this residue. Finally, we discuss the implications of our findings on the recruitment of other centromeric proteins, such as CENP-C, and propose a model in which HJURP may play dual roles in guiding CENP-A's deposition, serving both as a folding and a binding chaperone.

3.2 Methods

3.2.1 Structure preparation for MD simulations

Starting from the crystal structures for canonical H3 nucleosome (PDB ID: 1AOI [2]) and the CENP-A/H4 heterodimer with chaperone HJURP (PDB ID: 3R45 [3]), we developed all-atom and CG-AWSEM 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. Systems 1, 2, and 4 are based directly on PDB structures, or subdomains thereof, and we aligned the H3/H4 dimer to the CENP-A/H4 dimer of CENP-A/H4/HJURP to construct a CG-AWSEM model for H3/H4 in conjunction with HJURP. Finally, for the all-atom model of H3/H4/HJURP, we rotated the final three residues of H4 (-GRT) slightly after alignment to the CENP-A dimer in order to prevent structural overlaps between H4 and the newly placed HJURP. From these four models, at two different resolutions, we performed all-atom and coarse-grained MD simulations.

The CENP-A/H4/HJURP crystal (PDB: 3R45) does not include the H4 C-terminal tail. But, in the nucleosome structure, the H4 C-terminal tail is resolved and forms a few hydrophobic interactions with H3 (CENP-A) $\alpha 2$ and H4 $\alpha 3$ (Figure S14). Additional CG simulations were performed for a mixed CENP-A/H4 where CENP-A is provided from CENP-A/H4/HJURP (PDB: 3R45) and H4 from the CENP-A nucleosome (PDB: 3AN2), and for a CENP-A/H4 dimer derived solely from the CENP-A nucleosome structure (Figure S3 and Figure S15). Both simulations demonstrate that the H4 C-terminal tail is intrinsically unstable. The results of these additional runs are addressed in the discussion section and presented in the Supplementary Information.

3.2.2 Coarse-grained MD methods

For coarse-grained MD, we used Associative-memory, Water-mediated, Structure and Energy Model (AWSEM) [44] as the force field. In AWSEM, three beads – C_α, C_β (H for Glycine) and O – represent one amino acid. Water-mediated interactions [46] are applied instead of other explicit or implicit water models. Fragment memory, which is included in the associate memory potential, is set as a single memory determined by the crystal structure of the corresponding histone monomer. Fragments are non-overlapping and twelve (or fewer) residues long to ensure that it only provides a local structural bias. The interface dynamics between two molecules is purely determined by physics, not including any bioinformatics terms. To prevent the division of one dimer into two monomers, we applied a weak harmonic spring between the centers-of-mass of the two monomers (k = 0.02 kcal/(mol Å^2)). More details about AWSEM are included in the original force field study [44].

AWSEM coarse-grained MD simulations are run through LAMMPS package. Using the Nose-Hoover thermostat, we perform 200 ns NVT MD runs at 300 K with the initial velocities randomly generated for every bead drawn from a Maxwell-Boltzmann distribution. Five independent simulations with different random seeds of velocity distributions are carried out for each system. For analysis, we combine all five independent simulations after reaching equilibrated states, by deleting the first 10 ns, which is considered as the time required to reach equilibration (Figure S5). The trajectory is saved every 1000 time-steps, which is 2 ps in the coarse-grained timescale. It is worth noting that the timescale in coarse-grained simulation

is different from the timescale in all-atom simulation. Due to the faster diffusion, the same amount of CG-AWSEM simulation time samples much more conformational phase space than all-atom simulation does. CG simulations reach the convergence at around 10 ns, as shown in the RMSD and RMSIP analysis (Figure S5 and S17). It is important to note that while the timescale of atomistic simulations is absolute, and can be directly related to experimental timescales, 10 ns of CG simulations cover several orders of magnitude longer real timescale (microsecond-to-millisecond).

3.2.3 All-atom MD methods

We performed all-atom molecular dynamics (MD) in explicit solvent using the gromacs 4.5.7 MD software, [94] the amber99SB*-ILDN [95, 96] force field for proteins, the ions08 [97] force field for ions, and the TIP3P water model. 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 nm). Periodic boundary conditions were employed throughout all the simulations, and long-range electrostatics were treated with the Particle Mesh Ewald method [98]. Non-bonded Coulomb and Lennard-Jones interactions were truncated at 10 Å, and all bonds involving hydrogen were

constrained using the LINCS [99] 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 with and without the chaperone HJURP, 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 [100] with a 1.0 ps time-constant and the Parrinello-Rahman barostat [101] with a relaxation time of 2.0 ps. For analysis, we only considered the final 600 ns of the trajectories to account for further temperature and pressure equilibration. Convergence of the all-atom simulations can be seen from the RMSD (Figure S5) and Root-mean-square-inner-product (RMSIP) [102, 103] analysis (Figure S17). A detailed explanation of the RMSIP calculation is provided in the Supplementary Information.

3.2.4 *In vivo* experiments: cloning and immuno-fluorescence

Original GFP-CENP-A and mCh-CENP-A plasmids were a gift from Stephan Diekmann. To generate the mutant serine 68, we performed fusion PCR with mutant forward primers ATAAGGAAGCTGCCCTTC[GCA]CGC or ATAAGGAAGCTGC-

CCTTC[GAA]CGC with a common reverse primer GAAGGGCAGCTTCCTTATCA for the [alanine] or [glutamic acid], respectively. The whole mutant CENP-A coding sequence after fusion PCR were cloned in-frame and downstream of the EGFP and linker peptide. The plasmids were co-transfected using Roche's X-tremeGENE HP DNA Transfection Reagent (Cat # 06-366-546-001, Lot #11062300) into HeLa cells that were grown on poly-D-lysine coated coverslips. Three days after transfection, the coverslips were cytospun at 800 rpm for 5 min to reduce the number of Z-stacks during immuno-fluorescence. Coverslips were then prefixed with 4% paraformaldehyde (PFA) for 1 min, washed 3X with PEM (80 mM K-PIPES, pH: 6.8; 5 mM EGTA, pH: 7.0; 2 mM MgCl2), soluble proteins extracted with 0.5% triton-X100 in CSK buffer (10mM PIPES, pH: 6.8; 100 mM NaCl; 200 mM sucrose; 3 mM MgCl2; 1 mM EGTA) for 5 min at 4C, washed once with PEM and fixed with 1% PFA for 20 min at 4C. The coverslip was then washed 3X with PEM, air dried in the dark and mounted with Vectashield with DAPI (softset) and sealed along the edges with nail polish. Slides were stored in the dark at 4C until imaging with a DeltaVision RT system fitted with a CoolSnap charge-coupled device camera and mounted on an Olympus IX70.

3.2.5 Analysis for the MD simulation 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, investigating overall structural variation. We analyzed inter-

residue contact preferences 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. We used the STRIDE [104] algorithm to assign secondary structure to the all-atom simulation snapshots, considering the final six residues of CENP-A assigned as either 3_{10} or α to be helical. The average helical percentage was determined for each residue, and the average helicity of the CENP-A C-terminal tail was calculated as the mean of the averages for the final six residues.

To analyze the data from a more global perspective, we calculated a specific measure of structural similarity, Q [105], of all the simulation snapshots to the experimentally determined crystal structures. A widely-used quantity in protein folding theory, Q is a normalized order parameter, with higher values indicating greater higher structural resemblance between the two structures being compared.

$$Q = \frac{1}{n} \sum_{i<j-2} \exp\left[-\frac{\left(r_{ij} - r_{ij}^{native}\right)^2}{2\sigma_{ij}^2} \right], \tag{3.1}$$

where n is the total number of contacts, r_{ij} is the instantaneous distance between the Cα atoms of residues i and j, r_{ij}^{native} 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 generated probability density functions $P(Q)$ of all the simulation snapshots, where the shape of this distribution characterizes the structural heterogeneity of the related conformational ensemble. We first applied this

order parameter to interface profiles of H3/H4 and CENP-A/H4. A pair of residues from CENP-A or H3 and H4 was considered a native contact if their Cα atoms are within 12 Å in the experimentally determined x-ray crystal structure, and only native interface contacts are considered for $Q_{\text{interface}}$ calculation. Lastly, we applied this formula of structural similarity to the native state to CENP-A or H3 and H4 histones separately, which we refer to as $Q_{monomer}$.

The angle between two α helices was determined by calculating the orientation vectors for selected helices. The assessment of convergence was mainly through RMSD and RMSIP. RMSIP was calculated using the first ten eigenvectors of a given subspace. Detailed explanations of the methods used to determine helix orientation vectors and to calculate RMSIP values are provided in the Supplementary Information.

3.3 Results

In this work, we performed microsecond-scale coarse-grained and explicit-solvent atomistic MD simulations for the following systems: (1) the H3/H4 dimer; (2) the CENP-A/H4 dimer; (3) the CENP-A/H4/HJURP complex; (4) the H3/H4 dimer with HJURP. Initial conformations are based on the crystal structures of the canonical nucleosome (PDB ID: 1AOI [2]) and of the CENP-A/H4 dimer with chaperone HJURP (PDB ID: 3R45 [3]). In the Supplementary Information, we present the same analysis of coarse-grained MD simulations based on the dimer subdomain of the octameric CENP-A nucleosome (PDB ID: 3AN2 [26]). Currently, the

CENP-A/H4/HJURP structure is the only one that includes the final six residues of CENP-A. Distinguishing its structure from canonical H3, the C-terminal region of CENP-A is noted for its rapid evolution [18, 106], and functionally required for binding to CENP-C. [107] Therefore, much of our analysis focuses on the C-terminal end of CENP-A.

Coarse-grained and all-atom results are presented separately in the following two sections. CG-AWSEM results characterize global features of CENP-A and H3 dimers, examining how the histone monomers contribute separately to dimer stability, comparing the structural variability of CENP-A/H4 and H3/H4, and investigating the effect of chaperone HJURP on the CENP-A/H4 dimer. Further, contacts analyses based on all-atom MD simulations in explicit solvent provide a detailed physical description of how HJURP interacts with the CENP-A dimer, mapping key contacts between HJURP and the C- and N-terminal portions of CENP-A [108]. Lastly, *in vivo* experiments investigate the role of CENP-A S68, testing the hypotheses derived from all-atom MD contact map analysis. We have found that both simulation methods reach the same overall consensus qualitatively when performing the same analyses. Global measures from all-atom simulations are presented in the Supplemental Information.

3.3.1 CG-AWSEM MD Results

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

All core histones share the "histone-fold" structural motif, three helices connected by two loops, yet the number of sequence variants for each differs widely. This difference has important implications for histone evolution [17] and nucleosome assembly dynamics. For instance, several variants exist for the canonical histone H3 (*i.e.* H3.1) including H3.2/H3.3/CENP-A [68], while there are no variants for histone H4 reported thus far. From CG-AWSEM simulations, we first investigated how histone monomers H4 and H3, or H4 and CENP-A, contribute separately to dimer structural dynamics by calculating Q value, a normalized measure that compares the pairwise contacts in one structure to those in another (see Methods). A higher Q value (that can very between 0 and 1) indicates greater structural similarity between the two structures. Here, we calculated the Q value between the simulation snapshots and the corresponding crystal structures for H3/H4 (PDB ID: 1AOI [2]) and CENP-A/H4 (PBD ID: 3R45 [3]).

Interestingly, for all the systems studied, the conformations of H4 remain highly native-like, with an average Q value considerably greater than Q_{H3} or Q_{CENP-A}. The probability distributions of Q value for H4 are centered at ∼0.8 (Figures B.9.A,B,C,D), corresponding to root-mean-squared deviations (RMSD) ranging from 1.7 Å to 2.1 Å, whereas Q value for H3 at 0.7 corresponds to a RMSD range from

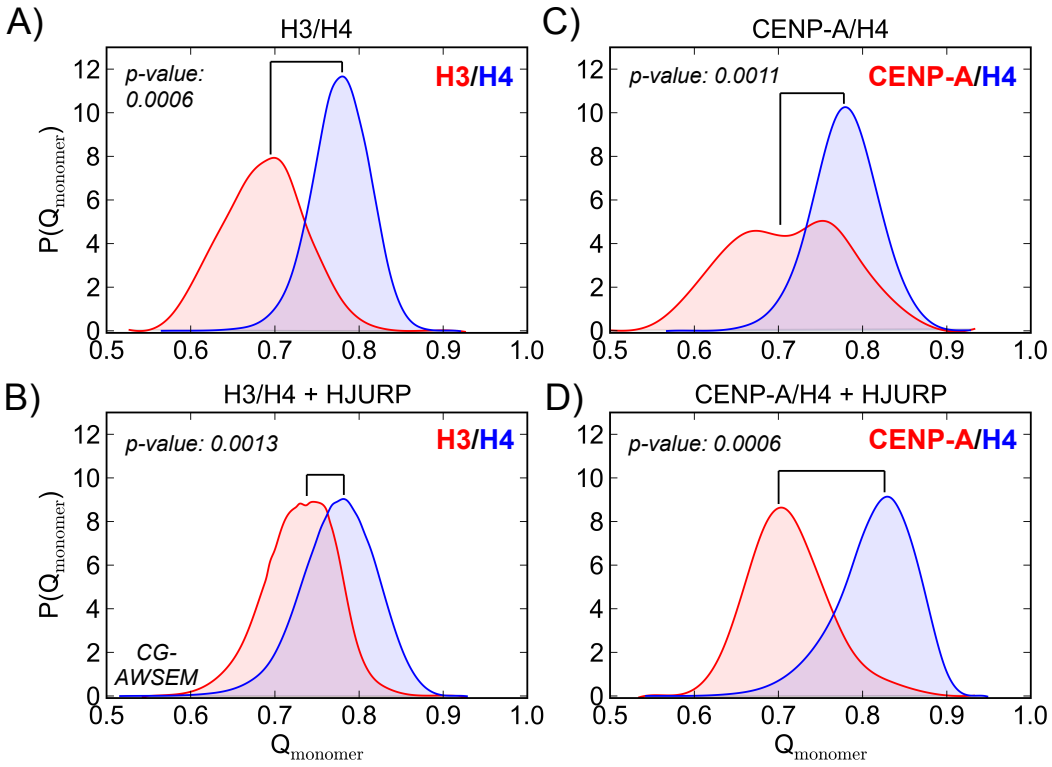

Figure 3.1: **H4 adopts conformations closer to the native state than CENP-A or H3 in CG-AWSEM simulations.** $Q_{monomer}$ characterizes a monomer's structural resemblance to its native state, defined by the corresponding monomeric conformations found in the crystal structures for H3/H4 (PDB ID: 1AOI [2]) and CENP-A/H4 (PDB ID: 3R45 [3]). Probability distributions of monomer Q are plotted for either H3 vs. H4 or CENP-A vs. H4 in (A) the H3/H4 dimer, (B) the CENP-A/H4 dimer, (C) H3/H4 in presence of HJURP and (D) the CENP-A/H4/HJURP complex. For each system, the average monomer Q value for H4 (blue) is greater than the average for CENP-A or H3 (red). Matching the CG-AWSEM results, H4 is structurally consistent in all-atom MD simulations (Figure S2).

2.0 Å to 2.6 Å, and for CENP-A Q at 0.7 corresponds to a RMSD from 2.0 Å to 2.9 Å. H4 is consistently stable in both H3/H4 and CENP-A/H4 dimers, with and without the presence of chaperone HJURP. Even though CENP-A displays large conformational variety in the CENP-A/H4 dimer, indicated by the broad distribution in $P(Q)$ (Figure B.9.C), H4 maintains native-like conformations for most of the simulation trajectories. When performing this analysis based instead on the CENP-A/H4 dimer found in the octameric CENP-A nucleosome crystal structure (PDB ID: 3AN2 [26]), we reach the same conclusion (Figure S3). Histone H4 consistently maintains native-like stability, providing a strongly reinforcing structural framework for histone dimers and higher order structures, such as the histone octamer. The intrinsic stability of H4 is independent of its dimer partner, CENP-A or H3, or the presence of chaperone HJURP.

3.3.1.2 CENP-A/H4 exhibits greater structural variability

We then examined the structural variability of the CENP-A/H4 and canonical H3/H4 dimers in CG-AWSEM simulations by calculating the root-mean-square deviations (RMSD) of Cα atoms with respect to the corresponding crystal structures. Replacing canonical H3 with CENP-A in the heterodimer leads to a greater RMSD, on average, for both CG (Figure B.10) and all-atom MD simulations (Figure S4). In the context of CG simulations, CENP-A/H4 (4.1 $\pm$ 0.5 Å) exhibits greater RMSD on average than H3/H4 (3.4 $\pm$ 0.4 Å) (Figure B.10.B). As expected, the two-residue longer loop 1 in CENP-A displays enhanced fluctuations (Figure S7).

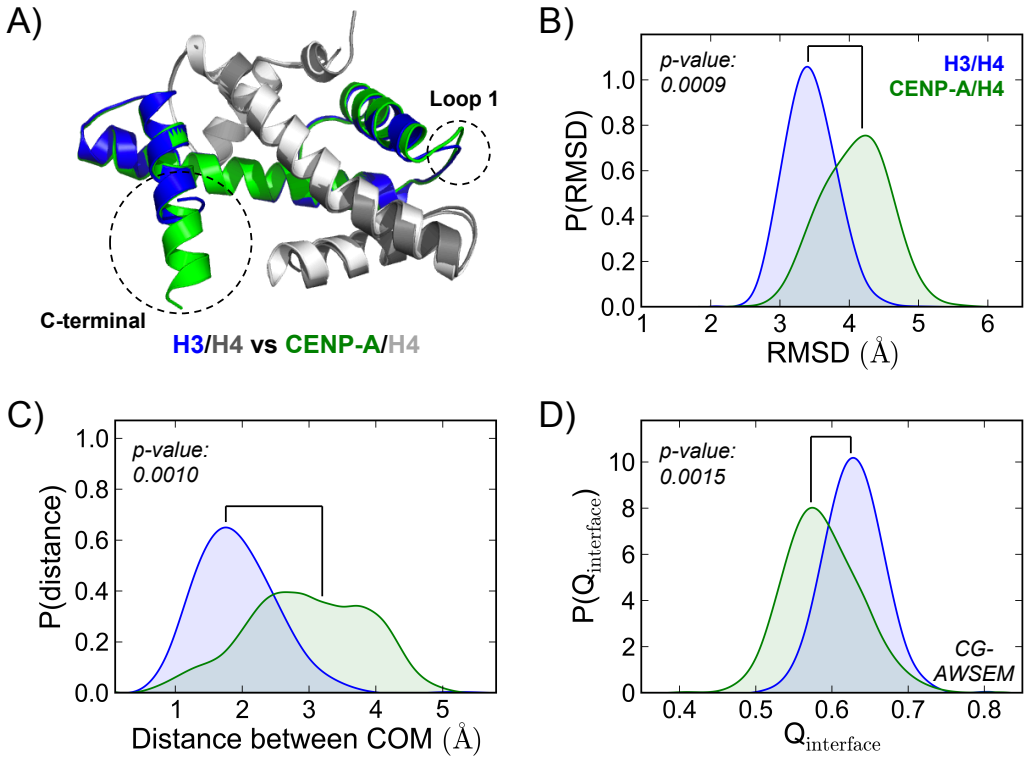

Figure 3.2: **CENP-A/H4 displays greater structural variability than H3/H4 in CG-AWSEM simulations.** (A) Structural alignment of CENP-A/H4 and H3/H4 highlights the two main structural differences between CENP-A and H3: the longer loop 1 and C-terminal regions of CENP-A (labeled by dashed circles). (B) Probability distribution functions of the Cα RMSD reveal that replacing H3 with CENP-A leads to greater structural variability in the dimer. (C) Probability distribution functions of the distance between the centers-of-mass (COM) of H3 (or CENP-A) and H4 show that CENP-A/H4 exhibits much more conformational heterogeneity. (D) Probability distribution functions of the $Q_{\text{interface}}$ with respect to the crystal structures of CENP-A/H4 (PDB ID: 3R45) and H3/H4 (PDB ID: 1AOI) for the CG-AWSEM simulation trajectories indicate that CENP-A/H4 has a more heterogeneous binding interface than H3/H4. Structure figure rendered in Pymol.

The spontaneous variability of CENP-A/H4 dimer in CG simulations is not only due to its flexible loop 1. The distance between the centers-of-mass (COM) of CENP-A and H4 occupies a much broader distribution than H3 and H4 (Figure B.10.C), indicating that the interface between CENP-A and H4 is more globally flexible. We analyzed the binding interface by calculating $Q_{interface}$, a normalized measure comparing the interface contacts in the CG simulation snapshots to those in the crystal structures (PDB IDs 1AOI for H3/H4, and 3R45 for CENP-A/H4). As shown in Figure B.10.D, the distribution of the CENP-A dimer $Q_{interface}$ is shifted considerably to the left of the same distribution for the H3 dimer, demonstrating that substituting canonical H3 with CENP-A leads to less native-like interfaces and increases the conformational heterogeneity of the dimer binding interface. Additionally, we calculated the pairwise Q value between any two conformations within one simulation trajectory. As shown in Figure S6, the pairwise Q is greater on average for H3/H4 (0.81 $\pm$ 0.04) than for CENP-A/H4 (0.73 $\pm$ 0.08) in CG simulation, implying that the higher heterogeneity of CENP-A/H4 is intrinsic and spontaneous. Overall, the isolated CENP-A/H4 dimer is more structurally variable than H3/H4 in both CG-AWSEM and all-atom simulations. These data are consistent with the greater heterogeneity seen in the CENP-A nucleosome compared to its canonical H3 counterpart *in silico*, *in vitro* and *in vivo* [5–7].

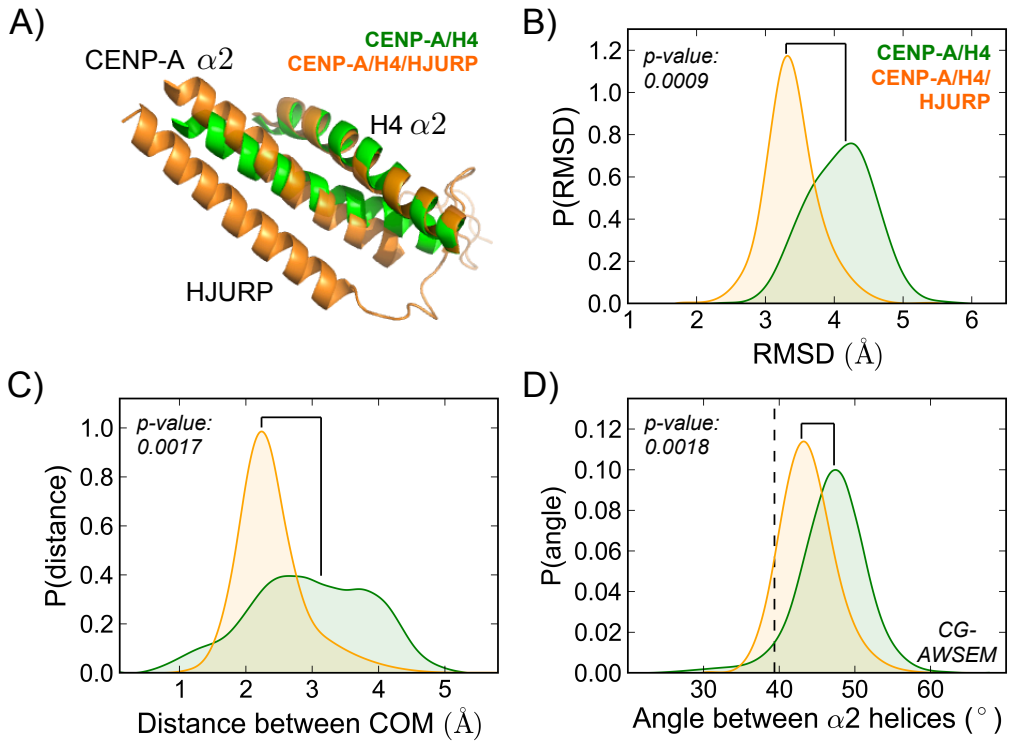

Figure 3.3: **HJURP stabilizes the overall shape of the CENP-A/H4 dimer in CG-AWSEM simulations.** (A) Representative simulation snapshots of CENP-A/H4 (green) and CENP-A/H4 in conjunction with HJURP (orange) illustrate how HJURP adjusts the overall shape of the dimer. Only the $\alpha 2$ helices of CENP-A and H4, as well as HJURP, are displayed. Introducing the CENP-A specific chaperone HJURP (B) reduces the CENP-A/H4 RMSD, on average, with respect to the crystal structure and (C) reduces the average distance between the COMs of CENP-A and H3, focusing the distribution and making the CENP-A/H4 dimer more compact and stable. (D) HJURP modifies the overall shape of the CENP-A/H4 dimer by reducing the angle between the $\alpha 2$ helices of CENP-A and H4. The reference angle from the crystal structure (40°) is illustrated by the dashed line. Structure figures rendered in Pymol. Similar analyses for the all-atom simulations can be found in Figure S8.

3.3.1.3 HJURP alters the shape of the CENP-A/H4 dimer

The data above demonstrate that, in isolation, the CENP-A/H4 dimer is structurally more variable than H3/H4 in CG simulations, which leads to the question of whether its chaperone HJURP influences the structural features of CENP-A/H4. Upon the introduction of HJURP, the RMSD distribution of the CENP-A dimer becomes tighter and shifts to the left (Figure 3.3.B), centered at 3.3 Å, which is comparable to the RMSD of H3/H4 in isolation (Figure B.10.C). Moreover, the distance between CENP-A and H4 shows much less deviation when HJURP is present (Figure 3.3.C). Therefore, in agreement with its documented role as a bonafide chaperone, HJURP stabilizes and restrains the conformational variability of the CENP-A/H4 dimer on a global scale.

$\alpha 2$ is the longest helix among the three major helices of each core histone, and provides the main supportive frame for the histone-fold structure. Thus, the shape of the CENP-A/H4 dimer can be characterized on a coarse level by the angle between the $\alpha 2$ helices of CENP-A and H4. Introducing the CENP-A-specific chaperone HJURP reduces the average angle between the $\alpha 2$ helices of CENP-A and H4 by 6° (Figure 3.3.D). The presence of HJURP tightens this distribution and brings its center closer to the reference value calculated from the crystal structure. As shown in the representative snapshot (Figure 3.3.A), HJURP modifies the orientation of CENP-A with respect to H4, bringing the CENP-A dimer's structure closer to that found in its octameric nucleosome. When performing the same analysis for all-atom MD simulations, we observe that the introduction of HJURP slightly

reduces the average RMSD (Figure S8.A). However, the distance between histone monomers and the angle between $\alpha 2$ helices remain unchanged (Figure S8.B,C). While CG-AWSEM MD simulations can explore conformational space widely, all-atom MD mainly probes dynamics near the native state, keeping global preferences relatively constant. Taken together, these results indicate that HJURP stabilizes the conformational ensemble of the CENP-A dimer and modifies the overall shape of CENP-A/H4, priming the CENP-A/H4 dimer for its deposition into the nucleosome, and, ultimately, into the centromere.

3.3.1.4 HJURP regulates the CENP-A/H4 dimer through stabilizing the C-terminal helix of CENP-A

After investigating how the introduction of HJURP influences the CENP-A dimer structure globally, we turn our focus to how HJURP specifically modifies the conformational preferences of the CENP-A monomer. The CENP-A $\alpha 3$ helix includes the final six residues at the C-terminus (i.e. LEEGLG in the human CENP-A sequence, Figure S1), which are currently thought to play an important role in CENP-A's interaction with the chaperone HJURP [3] and kinetochore protein CENP-C [107,109]. Presently, only the CENP-A/H4/HJURP complex includes an ordered CENP-A C-terminus in its crystal structure. Therefore, to better understand how HJURP dynamically affects the $\alpha 3$ helix of CENP-A, we measured the angles between the CENP-A $\alpha 1$ and $\alpha 2$ helices, and between CENP-A $\alpha 3$ and $\alpha 2$ (Figure 3.4.B).

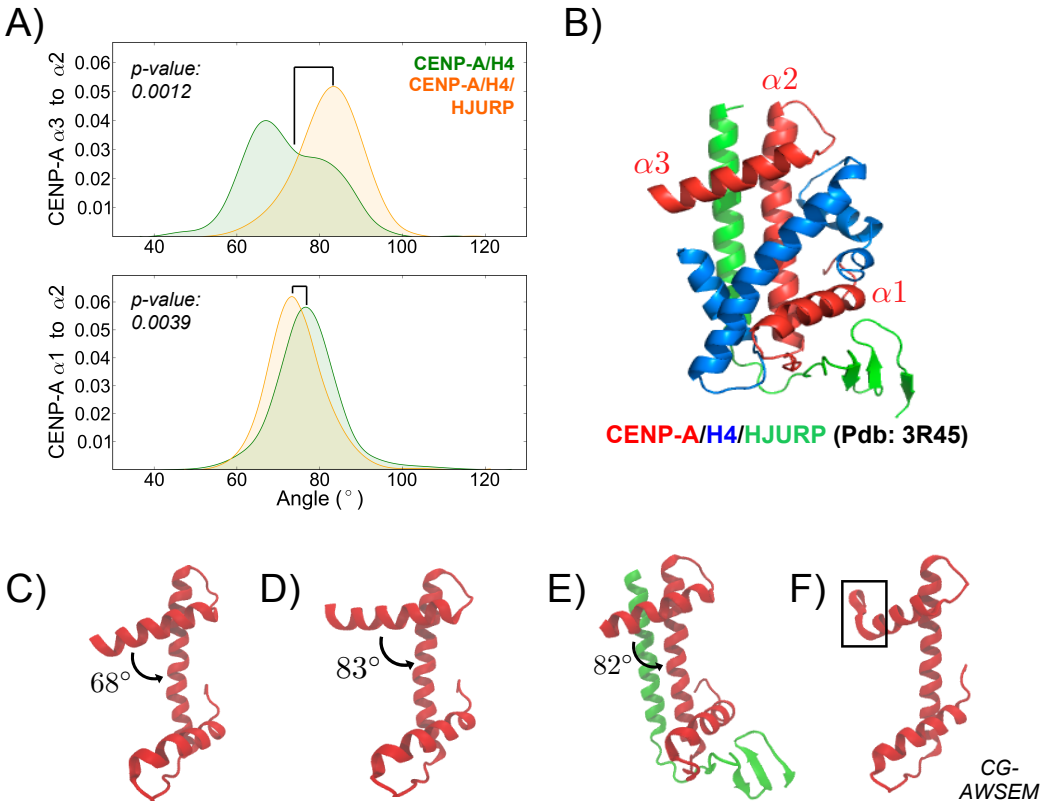

Figure 3.4: **HJURP stabilizes CENP-A α3 in CG-AWSEM simulations.** (A) Probability distributions of the angles between CENP-A α2 and α3, and between α1 and α2, demonstrate that the introduction of the chaperone HJURP stabilizes the motion of CENP-A α3 with respect to CENP-A α2. (B) The CENP-A/H4/HJURP crystal structure is shown. Helices used for the angle measurements are labeled in red. Conformations (C) and (D) correspond to the primary peak and shoulder in the distribution of the angle between α2 and α3 of CENP-A in the absence of HJURP. (E) A representative structure illustrates the most common angle between CENP-A α2 and α3 upon the introduction of HJURP. (F) In the absence of HJURP, the C-terminal end of α3 of CENP-A becomes partially unwound. Colors identify CENP-A (red) and HJURP (green). H4 is removed from the representative structures to facilitate easier observation. Structure figures rendered in VMD. Related CG trajectories can be found in the Supplemental Information (Movie S1 & S2). We observe the same overall trend when analyzing the angles between α2 and α3, and between α1 and α2, of CENP-A in the all-atom MD simulations (Figure S9).

The α3-α2 angle of CENP-A is broadly distributed, with a primary peak and a shoulder, at ∼68° and ∼82° respectively (Figure 3.4.A), corresponding to two populated states of CENP-A conformations when HJURP is absent (Figure 3.4.C,D). However, in the presence of HJURP, this angular distribution becomes tightened exclusively around the 82° peak (Figure 3.4.A,E). The preceding $Q_{monomer}$ analysis (Figure 1.C,D) also illustrates the change of Q_{CENP-A} from two populated states to one upon the introduction of HJURP. We observe the same overall trend in the all-atom MD results: the addition of HJURP stabilizes the angle between CENP-A α helices 2 and 3 (Figure S9.A) without having a significant effect on the angle between CENP-A α1 and α2 (Figure S9.B), in part because CENP-A α3 becomes partially unraveled in the absence of HJURP (Figure S9.C).

The CENP-A α3 helix is much more structurally dynamic than α1 in the CG simulations, since the CENP-A α1-α2 angle occupies only one focused peak and remains unchanged upon the introduction of HJURP (Figure 4.A). Further analysis reveals that the flexible CENP-A α3 helix could disrupt the stability of H4 α3 (Figure S11), which is consistent with all-atom contact maps (Figure 5). These results are also consistent with the experimentally determined B-factor data (Figure S10), which describes the uncertainty about the actual atom positions in X-ray crystallography. Moreover, this data provides a physical explanation of a key result from our previous CENP-A nucleosome work [7] – the shearing motion of the CENP-A nucleosome dimerization interface – wherein the interface, called the "four-helix bundle," is exactly defined by two copies of the CENP-A α3 and α2 helices. Altogether, our CG-AWSEM simulations demonstrate that HJURP regulates the

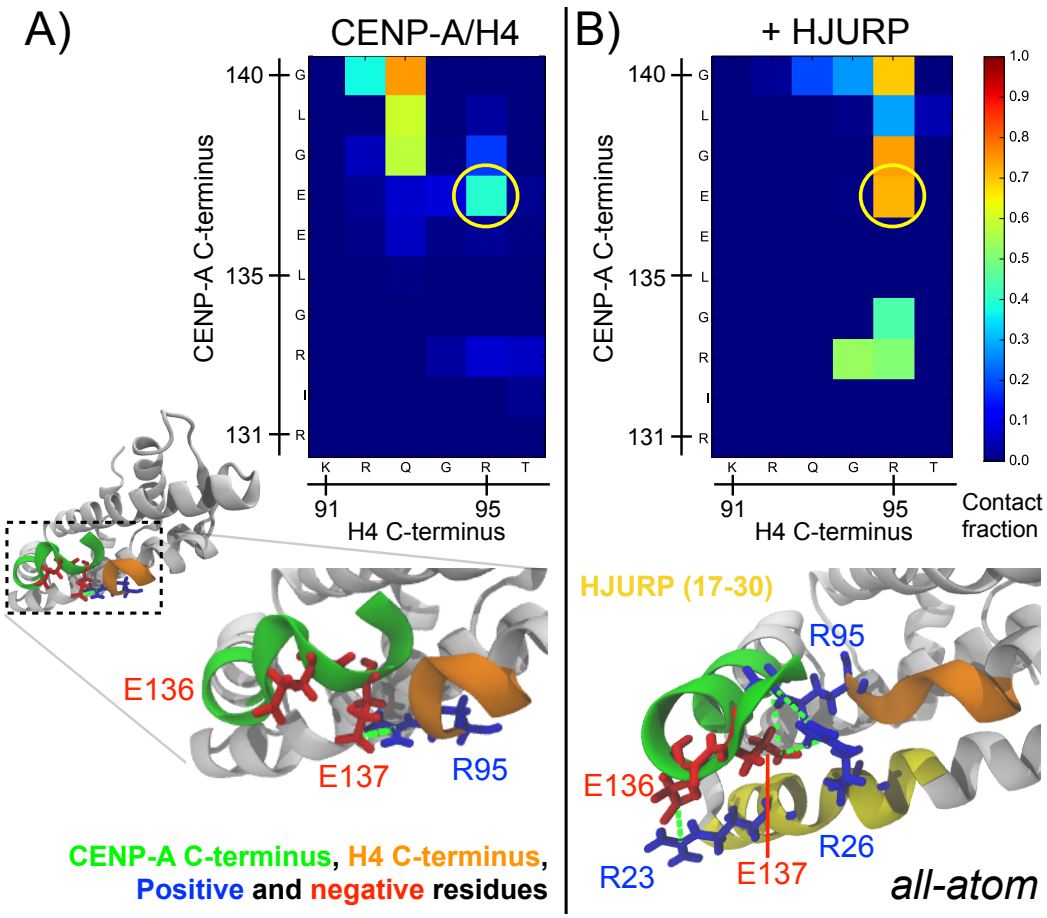

Figure 3.5: **The presence of HJURP rearranges interactions between the C-termini of CENP-A and H4.** Contact maps between the C-termini of CENP-A and H4, and representative simulation snapshots, in (A) the CENP-A/H4 dimer, and in (B) the CENP-A/H4 dimer in conjunction with CENP-A specific chaperone HJURP illustrate that HJURP facilitates electrostatic interactions that introduce greater helical structure to the C-terminus of CENP-A. The solid yellow circle highlights a potentially critical salt-bridge between CENP-A and H4.

CENP-A/H4 dimer through stabilizing the $\alpha 3$ helix of CENP-A.

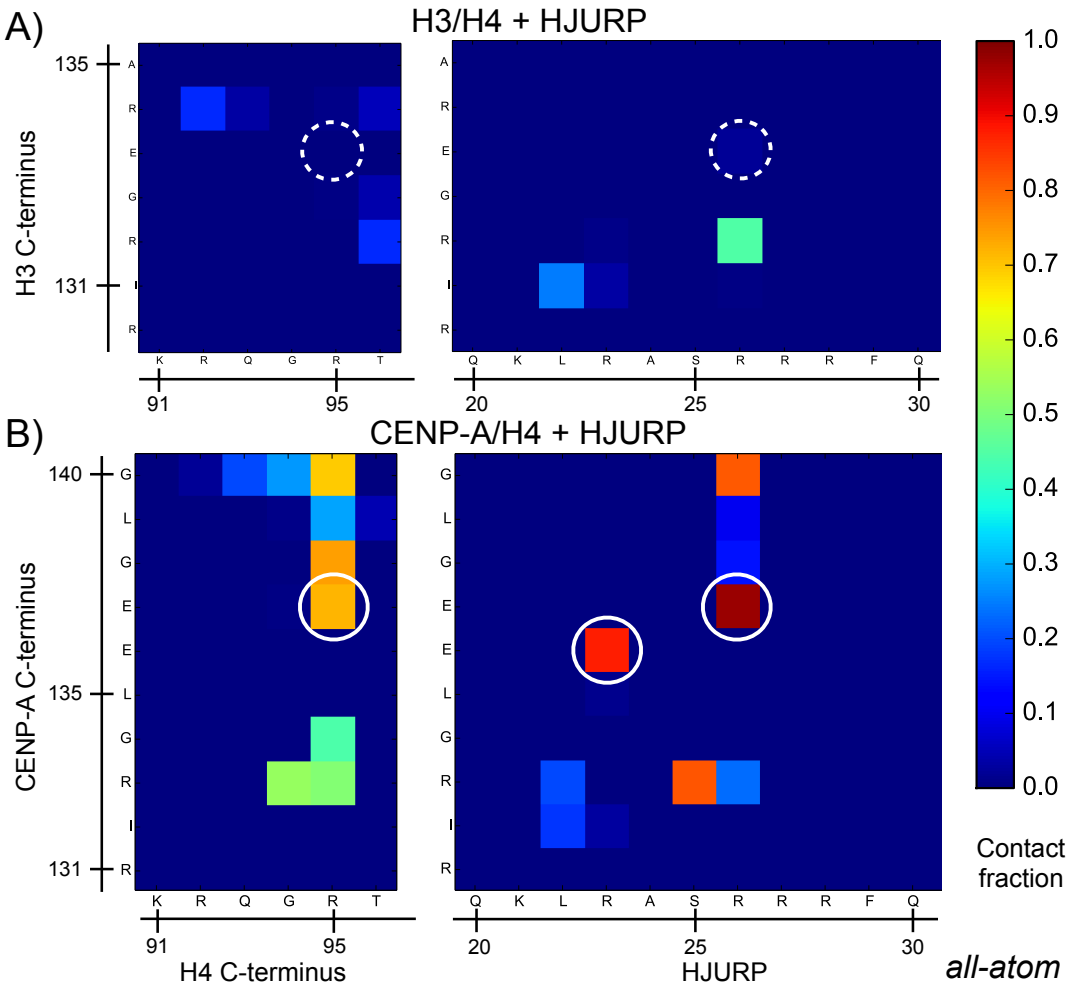

Figure 3.6: **HJURP forms electrostatic interactions with the C-termini of CENP-A/H4, but not H3/H4.** (A) The H3 C-terminus does not form significant interactions with the H4 C-terminus and α helix of HJURP in the H3/H4/HJURP all-atom trajectory. (B) Contact maps of the C-terminal region of CENP-A with the C-terminus of H4 and the α helix of HJURP in the all-atom simulation of CENP-A/H4/HJURP identify key electrostatic interactions. Solid white circles highlight specific salt-bridges, and dashed circles represent the lack thereof.

3.3.2 All-atom MD Results

3.3.2.1 HJURP facilitates forming a structure-inducing electrostatic network with CENP-A and H4

After analyzing global conformational features in CG-AWSEM simulations, we examined finer details of the interactions between CENP-A and H4, and those between HJURP and CENP-A, in all-atom simulations. First, we mapped the contacts between the C-termini of CENP-A and H4 in the absence and presence of HJURP (Figure 3.5.A,B). In the absence of HJURP, ∼40% of the time a contact forms between the oppositely charged H4 R95 and CENP-A E137 (Figure 3.5.A) and the α3 regions of CENP-A and H4 become partially unraveled. The C-terminal tail of CENP-A (the final 6 residues: 135-140) is ∼4% helical on average in the all-atom MD trajectory. The introduction of HJURP facilitates the formation of an electrostatic network between the C-termini of CENP-A and H4 and the α helix of HJURP, the contact between H4 R95 and CENP-A E137 increases to ∼70% (Figure 3.5.B), and the α3 regions of CENP-A and H4 retain their helical structure. The C-terminal tail of CENP-A increases to ∼35% helical on average in the presence of HJURP. Therefore, HJURP regulates the electrostatic interactions and drives the helicity in the CENP-A C-terminus. These results are consistent with the crystallographic information: except for the CENP-A/H4/HJURP complex, all other CENP-A-included crystal structures published thus far do not include the final six residues of CENP-A, because these residues remain disordered in these

structures. [26, 110, 111].

The C-terminal tail of CENP-A (-LEEGLG) carries an overall net charge of -2e and is three residues longer than the corresponding neutral tail of H3 (-ERA). The increased acidity and length of the CENP-A C-terminal tail compared to H3 could play an important role in differentiating assembly chaperones and binding partners for these two histones. Indeed, as can be seen in the contact maps analysis, several charged residues – including HJURP R23, R26, CENP-A E136, E137, and H4 R95 – form a network of interactions at the interface between the C-terminus of CENP-A, the C-terminus of H4 and the α helix of HJURP (Figure 3.6.B). In contrast, H3/H4 does not form analogous interactions upon the introduction of HJURP (Figure 3.6.A). Thus, the neighboring acidic residues near the C-terminus of CENP-A (E136 and E137) allow CENP-A to form key electrostatic interactions with basic residues of H4 (R95) and HJURP (R23 and R26).

3.3.2.2 CENP-A forms key interactions with the hydrophobic β domain of HJURP

On the other side, the N-terminal portion of the CENP-A histone-fold interacts with the hydrophobic β domain of HJURP. Previous experimental studies have focused on the role of CENP-A S68 in HJURP recognition, which has been challenged. [3, 112, 113] Here, we performed contact map analysis of the CENP-A/H4/HJURP all-atom simulations to examine the contribution of CENP-A S68 in atomistic detail. These analyses reveal that CENP-A S68 inserts well into the

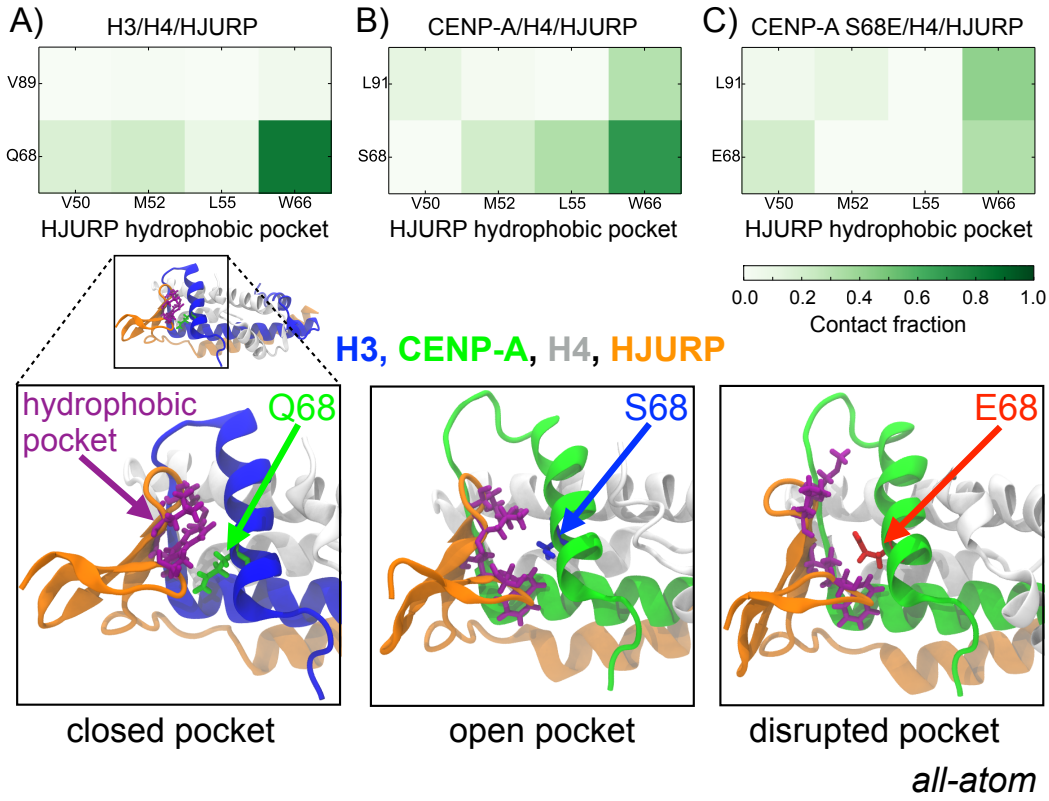

Figure 3.7: **CENP-A forms key interactions with the hydrophobic pocket of HJURP.** Contact maps between the hydrophobic pocket of HJURP (*i.e.* V50, M52, L55, and W66; in purple tubes) and key residues of (A) canonical H3, (B) CENP-A, and (C) CENP-A where S68 is replaced with E68 display different types of interactions. H3 Q68 almost exclusively interacts with HJURP W66, and HJURP's pocket becomes closed. CENP-A S68 forms contacts with multiple residues of the hydrophobic pocket, which remains open. When replacing CENP-A S68, E68 (shown in red tubes) disrupts the interactions between CENP-A and the hydrophobic pocket of HJURP. Colors identify H3 (blue), CENP-A (green), and HJURP (orange). Structure figures rendered in VMD.

hydrophobic pocket formed by the β domain of HJURP (V50, M52, L55, and W66) (Figure 3.7.B). On the contrary, H3 Q68 almost exclusively interacts with HJURP W66, leading to a closed hydrophobic pocket (Figure 3.7.A). While CENP-A S68 and L91 both form contacts with the hydrophobic pocket, there are virtually no interactions between these two CENP-A residues (only $\sim$2%). However, H3 Q68 interacts significantly with H3 V89 ($\sim$20%), which is the analogue of L91 in CENP-A. The data suggest that the shorter side chain of CENP-A S68 cannot reach CENP-A L91, whereas H3 Q68 is long enough to form contacts with H3 V89. Furthermore, since H3 Q68 and H3 V89 interact with each other, they cannot both insert simultaneously into the HJURP hydrophobic pocket (Figure 3.7.A). Between CENP-A S68 and CENP-A L91, S68 is more dominant in binding to HJURP: CENP-A S68 forms a contact with HJURP W66 $\sim$85% of the time, while the contact between CENP-A L91 and HJURP W66 is only present $\sim$35% of the time (Figure 3.7.B). Together, due to side chain lengths and strong to moderate hydrophobicities, CENP-A S68 and L91 permit CENP-A to form stronger interactions with HJURP than H3 Q68 alone.

To test our hypothesis that CENP-A S68 is required to bind with HJURP due to both the short length and some hydrophobicity (and electric neutrality) of its side chain, we performed *in vivo* experiments and all-atom simulations mutating this residue. Alanine (A), which is short and hydrophobic, and glutamic acid (E), which is long and negatively-charged, served as valuable replacement residues, denoted CENP-A S68A and S68E, respectively. In the experiment, we aimed to determine whether the S68-mutated CENP-A could still be functionally deposited

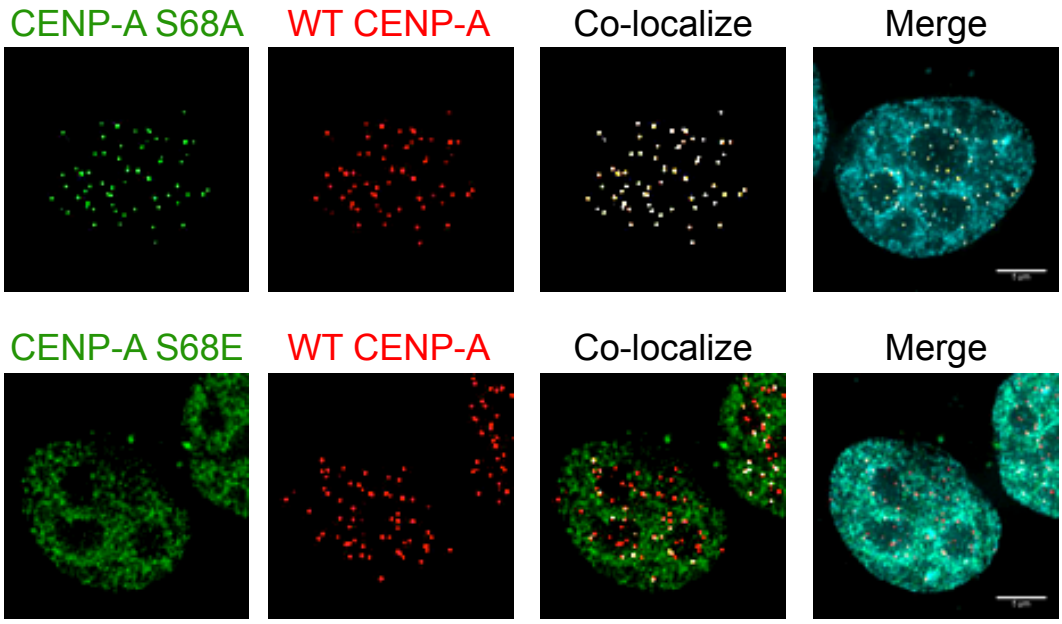

Figure 3.8: **CENP-A S68A localizes to the centromere whereas CENP-A S68E does not.** Residue S68 in CENP-A is mutated to alanine or glutamic acid respectively. Mutant are GFP-tagged, and co-expressed with mch-tagged WT CENP-A to assess co-localization. Co-localized foci appear as white dots in the co-localized column. Merge column shows the DAPI-stained DNA within the nucleus.

to the centromeric region by its chaperone HJURP *in vivo*. Successful binding with HJURP drives CENP-A deposition exclusively to the centromeres, whereas disrupted binding with HJURP is predicted to lead to the ectopic deposition of CENP-A. Site-directed mutagenesis experiments were conducted for CENP-A S68A and CENP-A S68E. These GFP-tagged CENP-A S68 mutants were co-expressed with mCh-tagged wild-type (WT) CENP-A under the control of a constitutive promoter, and the mutants' ability to localize to either the centromere or at ectopic regions was determined. Comparing the localization of mutated and WT CENP-A (Figure 3.8), it can be seen that the mutant CENP-A S68A results in robust centromeric localization while the mutant CENP-A S68E is not localized to the centromeres but displays ectopic incorporation.

To gain more biochemical insight into the specific role of S68, we performed all-atom simulations of CENP-A/H4/HJURP replacing CENP-A serine 68 with glutamic acid. The CENP-A S68E mutant disrupts the interactions between CENP-A and the hydrophobic pocket of HJURP (Figure 3.7.C). The longer side chain of E68 sterically clashes with HJURP's hydrophobic pocket, pushing it away from the CENP-A α1 helix. Once pushed away, the hydrophobic pocket becomes disrupted and loses its structural integrity. This explains why S68E CENP-A cannot successfully be recognized and loaded by chaperone HJURP in our *in vivo* experiments. Overall, our all-atom MD simulations and *in vivo* experiments demonstrate that CENP-A S68 is necessary to maintain the unique binding interface between CENP-A and the hydrophobic β domain of HJURP. All-atom simulation results indicate that the short length of S68's side chain is essential for CENP-A's recognition by the hydrophobic β domain of HJURP.

3.4 Discussions

3.4.1 Biological implications

In this report, coarse-grained and all-atom MD simulations provide a dual-resolution perspective of the effects of HJURP and CENP-A on histone dimer dynamics. These data reveal that the replacement of canonical H3 with CENP-A translates into increased conformational heterogeneity in histone dimer dynamics (Figure B.10). Furthermore, the chaperone HJURP plays a stabilizing role for the CENP-A/H4 dimer, and modifies the CENP-A dimer's overall shape (Figure 3.3)

as a potentially priming step in advance of the CENP-A loading. H4 remains stable and adopts native-like conformations in both CENP-A/H4 and H3/H4 (Figure B.9). This intriguing distinction is consistent with the fact that H4 remains conserved throughout eukaryotic evolution, whereas distinct variants of H3 exist for special roles in transcription and chromosome segregation. Thus, H4 could provide a consistent reinforcing structural framework for histone dimers, while the H3 family, including canonical H3 and the centromere-specific variant CENP-A, provides variability to the structure and function.

Our overarching aim is to investigate the fundamental dynamics of the histone dimers H3/H4 and CENP-A/H4. Therefore, only the histone-fold domains were previously considered, excluding the H3 (CENP-A) N-terminal helix and histone tails, based on the fact that those regions are primarily involved in the interactions with DNA or other histones, such as H2A/H2B (Figure S14). Nevertheless, in the nucleosome structure, the H4 C-terminal tail forms a few hydrophobic interactions with H3 (CENP-A) $\alpha 2$ and H4 $\alpha 3$, suggesting the possibility that the H4 C-terminal tail stabilizes histone dimers (Figure S14). In CG simulations, the angle between CENP-A $\alpha 2$ and H4 $\alpha 3$ is mostly stable in the absence of the H4 C-terminal tail (Figure S11). Further CG simulations demonstrate that including the H4 C-terminal tail increases the structural flexibility of the CENP-A/H4 dimer, compared to when the H4 C-tail is excluded (Figures S3.B,D and S15.B,D). It is feasible that H2A/H2B, together with H3(CENP-A) $\alpha 2$ and H4 $\alpha 3$, stabilizes the H4 C-terminal tail, as can be seen in the nucleosome crystal structure: β strands form between the H4 C-terminal tail (H4 T96 and Y98) and H2A T101 (Figure S14). Interestingly, even

with the H4 C-terminal tail included, H4 still adopts more native-like conformations than CENP-A (Figure S3.C, Figure S15.C). Investigating the precise role of histone tails in the CENP-A/H4/HJURP complex and the structural dynamics comparison between CENP-A/H4 and H3/H4 homotypic or heterotypic histone tetramers are important future directions.

The variability of CENP-A is due, in part, to its longer C-terminal residues (six in CENP-A versus three in H3), which maintains helical structural integrity only when in a complex with HJURP (Figure 3.5). The increased acidity of the CENP-A C-terminus (-2e) compared to the neutral charge of the corresponding C-terminus in H3 could contribute to HJURP's specificity to CENP-A [107]. The coarse-grained MD results demonstrate that HJURP reduces the conformational heterogeneity of the CENP-A/H4 dimer by modifying the dimer's overall shape and stabilizing the CENP-A α3 helix (Figures 3.3 and 3.4). Furthermore, all-atom MD simulations illustrate that HJURP forms a structure-inducing electrostatic network with the C-termini of CENP-A and H4 but not with H3/H4 (Figures 3.5 and 3.6). The two-residue-longer loop 1 region of CENP-A is subject to less fluctuations upon the introduction of HJURP (Figure S7), which indicates that HJURP stabilizes loop 1 region of CENP-A indirectly. Debate continues over the role of CENP-A S68 [3, 112, 114] and its post-translational modification [108] in CENP-A's interaction with HJURP and deposition into the nucleosome. Replacing CENP-A S68 with E68 *in vivo* and in all-atom MD simulations mimics S68 phosphorylation by elongating the side chain and introducing a negative charge. Recent studies suggest that phosphorylating S68 is sufficient to disrupt CENP-A–HJURP binding. In our

experiments (Figure 3.8), mutating this residue to glutamic acid resulted in ectopic CENP-A deposition *in vivo*. All-atom simulations provide a physical explanation of how S68 phosphorylation could disrupt the binding interface between CENP-A and HJURP: when replacing CENP-A S68, the longer E68 side chain sterically clashes with HJURP's hydrophobic pocket, pushing it away from the CENP-A α1 helix and disrupting the pocket's overall shape. Together, *in vivo* and all-atom simulation results support the previously proposed model in which CENP-A S68 phosphorylation (S68ph) must be tightly regulated and the eviction of CENP-A's chaperone HJURP must be orchestrated within a small window of the cell cycle in order to minimize the risk of ectopic CENP-A incorporation. [108]

Further analysis reveals that the introduction of HJURP to H3/H4 significantly disrupts the binding interface between H3 and H4 (Figure S12.B) and leads to a slightly larger average RMSD in CG-AWSEM simulations (Figure S12.A), compared to the H3/H4 dimer in isolation. In all-atom simulations of the same system, the introduction of HJURP destabilizes a key electrostatic interaction between the C-termini of H3 and H4 (Figure S13). These results may provide a partial explanation for experimental evidence suggesting that H3/H4 cannot bind HJURP *in vitro* [89, 92, 112].

Based on our observations above, it is possible that a currently under-appreciated role for chaperone HJURP may also be its ability to "lock" the C-terminus of CENP-A before it encounters another kinetochore protein. HJURP may work as a switch, turning on and off the binding availability of the CENP-A C-terminal tail. The presence of HJURP stabilizes the C-terminus of CENP-A before CENP-A's deposition,

and, after CENP-A is deposited, HJURP must release the intrinsically-disordered C-terminal tail of CENP-A, in order for it to become available to bind with another kinetochore protein, most critically, CENP-C [6, 107]. The structural alignment of CENP-A from different molecular contexts clearly shows the "on" and "off" states of its C-terminal tail (Figure S16). Plus, recent research by Tachiwana *et al.* illustrates that CENP-C recruitment requires direct interaction between CENP-C and HJURP [115]. Consequently, HJURP may be unique in that it functions as a protein-folding chaperone for CENP-A, stabilizing the CENP-A/H4 dimer, and also as a protein-binding chaperone for CENP-C and CENP-A, mediating CENP-C's recruitment to the CENP-A nucleosome. A related work previously reported on the interaction between the chaperone Chz1 and the H2A.Z/H2B dimer, wherein the chaperone Chz1 undergoes a disorder-to-order transition upon binding to H2A.Z/H2B [69], suggesting such transitions might be conserved in the structure-inducing mechanisms employed by histone chaperones. [116–118]

3.4.2 All-atom and coarse-grained: the dual-resolution MD method

The dual-resolution nature of this study provides a unique opportunity to directly compare and cross-validate the same results from both CG and all-atom simulations. Therefore, for each of the main CG results (monomer flexibility; dimer variability; global shape; and HJURP's effect on the angle between helices), we performed the same analysis on the all-atom MD trajectories, including the resulting figures in the Supplementary Information (Figures S2, S4, S8, and S9). Overall,

all-atom and CG methods reach the same consensus qualitatively. However, how the results of these two techniques differ is important to our work as well. When examining global properties including pairwise Q, interface Q, and the distances between histones, the results based on all-atom MD simulations remain close to the native state, and these properties do not vary much across different systems. On the other hand, the analysis of CG simulations reveals significant differences in the global properties of the systems studied, clearly illustrating the value added by including CG simulations. The strength of all-atom MD lies in its ability to probe specific interactions and native-state dynamics at high resolution. For example, when replacing CENP-A S68 with E68 in all-atom simulations, the glutamic acid sterically clashes with HJURP's hydrophobic pocket, pushing the pocket away from the CENP-A α1 helix (Figure 7). This detailed effect is not observed in CG-AWSEM MD simulations because it is mainly due to the long length of the glutamic acid side chain, a difficult property to capture in a three bead per amino acid model. Altogether, CG explores greater conformational space at a more global level, and all-atom MD investigates finer details close to the native state.

3.5 Conclusion

In summary, our dual-resolution MD simulations shed light on the differences between the structural dynamics of the CENP-A/H4 and H3/H4 dimers, providing insights into how HJURP primes the CENP-A/H4 dimer for deposition. Our results indicate that HJURP, while potentially acting as a disruptive force for H3/H4,

serves as a protein-folding chaperone for the CENP-A dimer and a protein-binding chaperone for CENP-C and the CENP-A dimer. Finally, this study makes predictions about the key histone-histone and CENP-A-HJURP interactions, one of which is confirmed by *in vivo* experiments, and provides new dynamic insights into the underlying mechanisms governing the HJURP-mediated assembly of CENP-A nucleosomes *in vivo*.

Chapter 4: The Oligomerization Landscapes of Canonical and CENP-A Histones

The chapter is based on the submitted work of the author: H. Zhao, D. Winogradoff, Y. Dalal, and G. Papoian; (2018)

4.1 Introduction

Eukaryotes wrap their DNA around histone proteins to form nucleosomes, the fundamental unit of chromatin. Typically, the histone component of a nucleosome is an octamer core, composed of a central $(H3/H4)_2$ tetramer plus two H2A/H2B dimers [2]. The histone tetramer initiates nucleosome assembly by forming a tetrasome with DNA [119, 120] which recognizes the nucleosome positioning signals [13], serving as the structural basis for nucleosomal or chromatin dynamics. Recent single-molecule experiments studied the spontaneous flipping behavior of DNA handedness on the tetrasome, and found that the iodoacetamide-treated residue mutations around the tetramer interface can result in the enhanced flexibility and in faster superhelical flipping kinetics of the wrapped DNA, indicating the critical importance of studying the histone tetramer to DNA supercoiling in chromatin. [121–123] Hence, a deep molecular understanding of histone tetramer

dynamics is crucial to understanding tetrasome or nucleosome assembly and may unveil innovative pathways that have evolved to deal with the mechanical stress associated with chromosome segregation.

Crystallographic studies have resolved consistent structures of the histone tetramer in different molecular contexts, including in the form of an octamer with H2A/H2B [2, 124, 125], and in the binding state with chaperone proteins such as FACT [126], Spt2 [127], TONSL and MCM2 [128,129]. Studies of dynamics in solution have been limited due to the tetramer's large heterogeneity. Early experiments utilized size-exclusion chromatography to demonstrate that the histone tetramer is in a dynamic equilibrium alternating between two H3/H4 dimers and an assembled tetramer [130, 131]. Through Electron Paramagnetic Resonance (EPR) spectroscopy, previous work shows that the histone tetramer exhibits greater structurally heterogeneity on its own than when sequestered in the octamer [132]. However, the structural details that would reveal the mechanisms governing those properties are not readily available to experimental techniques. Thus, computational investigation into histone tetramer dynamics is needed.

Centromere Protein A (CENP-A) is a centromere-specific histone H3 variant, which has been proposed to be the epigenetic mark of the centromere, ensuring proper chromosome separation during cell division. CENP-A has been studied extensively for its significant functional role [15,50,66,67,88,133] and interesting structural dynamics [51, 70, 134–137]. However, dissecting the dominant structure and dynamics of CENP-A nucleosomes [5,25,77,83–85], and their association with kinetochore partners [75,79,81,82,107], remains an exciting avenue of research. CENP-

A-containing nucleosomes are known to follow a different assembly pathway than those containing H3 *via* the unique chaperone HJURP [91–93, 115, 138]. CENP-A is over-expressed in cancer cells, and the redundant CENP-A can localize into ectopic (*i.e.* non-centromeric) regions *via* alternative chaperone pathways [24, 73, 135]. Thus, one outstanding question is whether CENP-A, in normal cells, can be efficiently regulated to avoid ectopic delivery. Another important question is whether replacing canonical H3 with variant CENP-A alters nucleosome's physical properties and overall dynamics. Conflicting studies suggest that: (1) CENP-A nucleosomes and CENP-A/H4 dimers are more flexible than their canonical H3 counterparts [6, 7, 19]; (2) *in vitro* deuterium exchange experiments indicate that the CENP-A tetramer is more rigid than the H3 tetramer [21, 25]; and (3) CENP-A- and H3-containing nucleosomes have virtually superimposable crystal structures, within a root-mean-square deviation (RMSD) of ∼2 Å [26–28]. Nucleosomes dissociate and re-associate during dynamic transitions in chromatin structure for fundamental biological processes such as transcription, replication, and repair. Hence, it is crucial to elucidate the assembly dynamics of canonical and variant tetramers, which are key intermediates along nucleosome assembly and disassembly pathways.

Molecular dynamics (MD) simulations are able to capture mechanistic details at the molecular level, complementing experimental approaches. Previously, we used atomistic MD to reveal that the CENP-A nucleosome exhibits greater flexibility than the canonical nucleosome [7], and its dynamics can be modulated by internal modifications [137]. Combining coarse-grained, atomistic simulations and *in vivo* mutation experiments, we reported that CENP-A dimer is structurally variable, and

chaperone HJURP prevents the promiscuous mis-assembly of the CENP-A dimer, protecting it from binding with other proteins [19].

Building upon these findings, we performed coarse-grained MD simulations using the AWSEM model [44] (CG-AWSEM; *i.e.* 3-beads per amino acid) to investigate the assembly mechanisms of histone tetramers and to determine how histones CENP-A and H3 differ at the tetramer level. In particular, we computed the free energy of association of two dimers into tetramer, demonstrating that CENP-A/H4 forms a more compact and stable tetramer with a more favorable free energy, while the free energy landscape of $(H3/H4)_2$ is significantly more rugged, indicating the structural lability. The latter is also evident from constant temperature simulations starting from pre-assembled tetramers, which reveal the rotational motion of H3 tetramer about its central interface. Furthermore, histone octamer simulations show that the addition of H2A/H2B dimers gently restrains the internal rotation of the $(H3/H4)_2$ tetramer, while also revealing significant incongruence between the preferred structures of CENP-A tetramer versus the octamer.

3.2 Methods

3.2.1 Simulation methods

Initial configurations of the simulation are obtained from the nucleosome crystal structures containing H3 (PDB: 1KX5 [139]) and CENP-A (PDB: 3AN2 [26]). Histone tails and DNA were not included in current study. All the sequences of simulated proteins are provided in the SI (Figure S6).

Associative-memory, Water-mediated, Structure and Energy Model (AWSEM) [44] was used, with three beads representing one amino acid and a water-mediated potential describing the water-protein interactions. Simulations were performed in the large-scale atomic/molecular massively parallel simulator 2016 (LAMMPS 2016), using the Nose-Hoover thermostat. Coupled replica-exchange and umbrella-sampling method was applied to enhance the phase space sampling. Samplings at 300 K were then collected for analysis after cutting out the beginning equilibration timesteps. PMFs were calculated using weighted histogram analysis method (WHAM) [4]. The relevant Jacobian factor correction term was subtracted from the free energy calculation [140]. All simulations were run in a 200-Å-long cubic box with periodic boundary conditions. The convergences of simulations were verified by the root-mean-squared inner-product (RMSIP) analysis (see SI section S2). More details are provided in the SI.

3.2.2 Trajectory analyses

All the trajectory analyses in this work, including the calculations of root-mean-square deviations (RMSD), radius-of-gyration (R_g), distances (R), dihedral angles θ, Q values, and contact analysis, were based on the Cα coordinates. Q values were calculated separately for the tetramer interface between two dimers, the whole dimer, the dimer interface between two monomers, and also for the monomers. The dihedral angle calculations in Figure 4 were obtained by measuring the dihedral angle formed by the first and last Cα atoms of the α2 helices. A contact in Figure 5

was considered to exist when the distance between two Cα atoms was shorter than 8 Å. More details about the dihedral angle calculation are provided in the SI.

3.3 Results

3.3.1 Binding free energy of the histone tetramer

Motivated by the previous observation of CENP-A dimer's flexibility [19], we first investigated the formation of two CENP-A and canonical H3 dimers, respectively. Via a mixed enhanced sampling methodology that couples replica-exchange with umbrella-sampling, we mapped their corresponding binding free energy landscapes. Basically, two CENP-A (or H3) dimers were put in the simulation with the distance between their centers-of-mass controlled by a umbrella potential $U = \frac{1}{2}k(R - R_o)^2$. K is the biasing strength for this harmonic potential and R_o is the controlled center distance for each window. Simulations of ten replicas with temperatures ranging from 280 K to 370 K were performed for thirty umbrella windows at various distances of two associating dimers from 20 Å to 50 Å. The binding free energy profile (FEP) was then calculated at 300 K and presented along different dimensions.

In Figure 3.1, we projected the free energy onto two reaction coordinates: the distance between centers-of-mass of the two dimers R_{COM} and another order parameter that quantifies the nativeness of the binding interface between the dimers, $Q_{interface}$. $Q_{interface}$ is the fraction of native interface contacts present, defined as, $\frac{1}{n}\sum_{i<j-2} \exp\left[-\frac{\left(r_{ij}-r_{ij}^{native}\right)^2}{2\sigma_{ij}^2}\right]$, where n is the total number of contacts, r_{ij} is the

distance between the Cα atoms of residues i and j, and σ_{ij} is given as $\sigma_{ij} = (1+|i-j|^{0.15})$. Q ranges from 0 to 1, where no common contacts between a conformation and the native state corresponds to 0 and complete similarity of contacts corresponds to 1. The Q calculations were performed with respect to the reference structure of hisone tetramers from the corresponding nucleosomes containing canonical H3 and variant CENP-A.

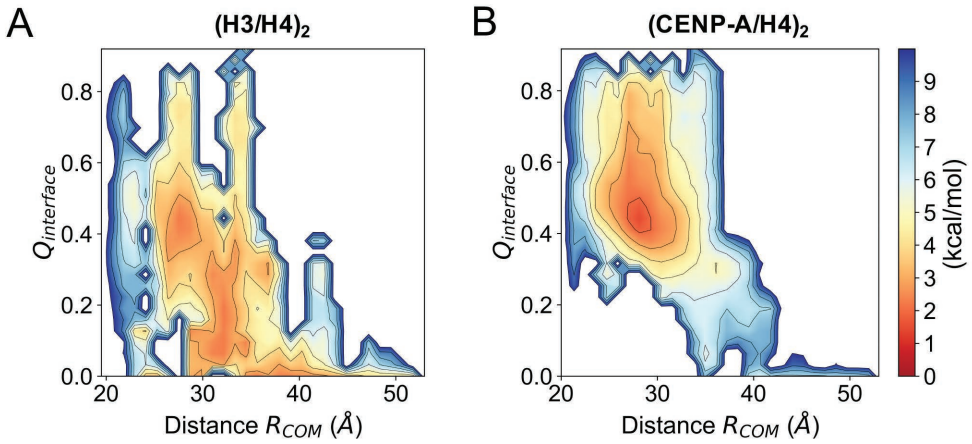

Figure 3.1: **The binding free energy landscapes of the H3 and CENP-A tetramers.** Two-dimensional free energy profiles are mapped as a function of the distance between two interacting dimers R_{COM} and of the quantification of the nativeness of their binding interface $Q_{interface}$, for $(H3/H4)_2$ (A) and $(CENP\text{-}A/H4)_2$ (B).

As seen in Figure 3.1, the binding free energy landscape for H3/H4 dimers (Figure 3.1A) is relatively rugged with multiple minima, at $Q_{interface}$ = 0.4, 0.1–0.2, and 0.0. These minima occupy a large portion of configuration space described in terms of R_{COM} and $Q_{interface}$, indicating the large heterogeneity of $(H3/H4)_2$ with a broad ensemble of accessible conformations. These data are consistent with the experimental result that histone H3 tetramer is unstable at moderate ionic strengths [130]. On the other hand, CENP-A/H4 has a well-funneled free energy

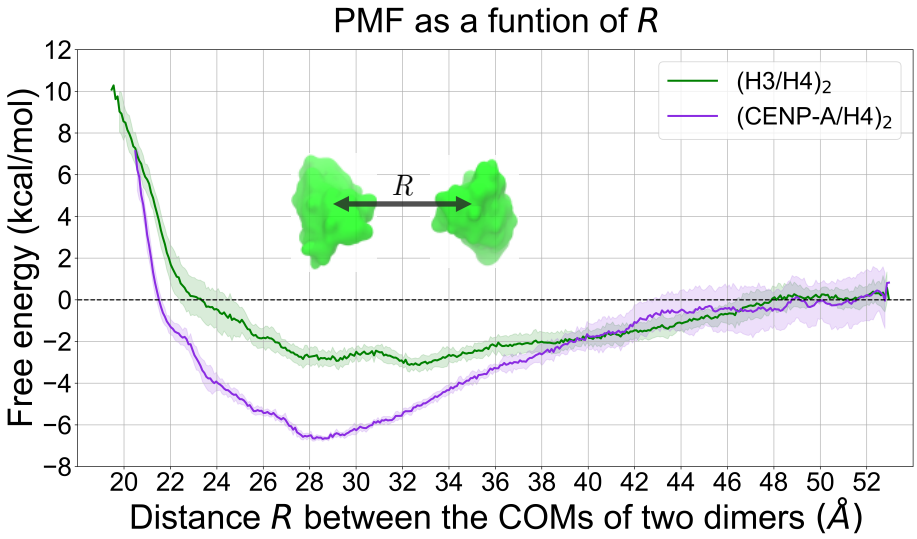

Figure 3.2: **$(\text{CENP-A/H4})_2$ has a deeper free energy profile than $(\text{H3/H4})_2$.** The potential of mean force (PMF) along the distance R between histone dimers is deeper for $(\text{CENP-A/H4})_2$ (purple) than for $(\text{H3/H4})_2$ (green). R is measured from the center-of-mass (COM) of one dimer to the other. The shaded areas illustrate the standard deviations of the curves.

landscape (Figure 3.1B). The minimum is at $R_{COM} = 29$ Å, $Q_{interface} = 0.4$, where there is a thermodynamically favorable binding state for $(\text{CENP-A/H4})_2$.

To further quantify and compare the binding free energies for H3/H4 and CENP-A/H4, we projected two computed FEPs along one dimension, R_{COM}, after aligning both the converged FE at the far-end to zero, at which we assume there is no interaction between the two dimers (*i.e.* when $R_{COM} > 50$ Å). Figure 3.2 presents the potential of mean force (PMF) as a function of the distance between the COMs of two H3 dimers or CENP-A dimers, indicating the FEP minimum for $(\text{CENP-A/H4})_2$ at appropriately -7 kcal/mol, and -3 kcal/mol for $(\text{H3/H4})_2$. The latter is in agreement with experimentally measured value [131], validating the accuracy of our simulation method. Since the overall FEP curve of CENP-A

dimers is deeper, we expect that, in the absence of DNA and other histone proteins, CENP-A/H4 dimers can more readily assemble into a tetramer than H3/H4 dimers. Furthermore, the free energy minimum is located at a distance of ∼28 Å between dimers of CENP-A/H4 and at ∼32 Å between dimers of H3/H4 (Figure 3.2), indicating that the thermodynamically favored CENP-A tetramer is more compact than the H3 tetramer. This result is in agreement with previous SAXS measurements that found CENP-A tetramer is substantially compacted relative to their H3 counterparts [110]. Overall, these free energy calculations suggest that the CENP-A/H4 homotypic tetramer is thermodynamically more stable, and more compact, than H3/H4. Additional free energy profiles using other one-dimensional and two-dimensional coordinates are provided in the SI (Supplementary Figures S4 and S5).

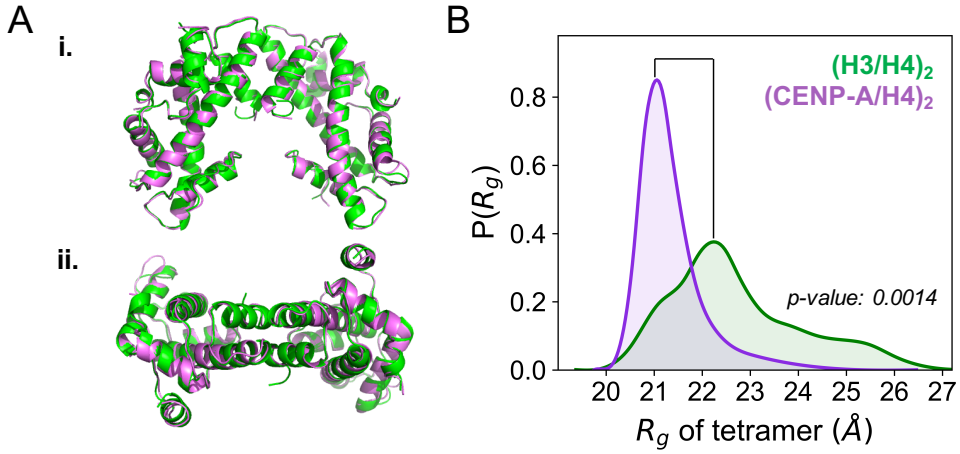

Figure 3.3: **$(CENP\text{-}A/H4)_2$ is more compact than $(H3/H4)_2$.** (A) The initial conformations of the H3 tetramer (green) and CENP-A tetramer (purple) were taken from their nucleosome crystal structures (PDB IDs: 1KX5 and 3AN2). Lateral view (i) and top view (ii) of aligned structures are displayed. (B) The CENP-A tetramer has a smaller radius-of-gyration R_g than the H3 tetramer, with a narrower distribution.

3.3.2 Histone tetramer geometries and swiveling dynamics

To explore the intrinsic dynamics of histone tetramers further, we performed microsecond-scale continuous constant temperature CG-AWSEM simulations for CENP-A and H3 tetramers at 300 K, starting from pre-assembled conformations, taken from the central tetramers of the corresponding octameric nucleosome crystal structures (Figure 3.3A). Overall, the results obtained from these constant temperature simulations were broadly consistent with enhanced sampling simulations, providing additional dynamics and insights. We present here some of the most salient observations: additional analyses including the principle component analysis (PCA) and distributions of other structural quantities including the root-mean-square deviation (RMSD), R_{COM}, and $Q_{interface}$ can be found in the SI (Supplementary Figure S6). R_g was used to quantify the molecules' degree of compaction, defined as $R_g = \sqrt{\frac{1}{N^2}\sum_{i=1,i<j}^{N}(\boldsymbol{r_i} - \boldsymbol{r_j})^2}$, where N is the total number of residues and $\boldsymbol{r_i}$ are the coordinates of ith residue. Figure 3.3B shows that the average R_g for (CENP-A/H4)$_2$ is 21 $\pm$ 0.7 Å and 23 $\pm$ 1.4 Å for (H3/H4)$_2$, implying that (CENP-A/H4)$_2$ samples more compact geometries with less R_g fluctuations. Plus, the distribution of (CENP-A/H4)$_2$ is unimodal, with a dominant central peak, while the H3 tetramer R_g samples a much broader distribution (Figure 3.3B). This result is consistent with our free energy calculations (Figure 3.2), and, together, these results indicate that CENP-A tetramer is more closely packed, structurally well-defined, and thermodynamically more favorable than H3.

In recent magnetic tweezer experiments [122, 123] , the DNA of H3-containing

tetrasomes were observed to flip between left- and right-handed superhelically-wound states, which may be initiated by conformational changes of the proteins inside. Considering these experiments, we then examined the overall orientation of the simulated tetramers by measuring the dihedral angle between the two dimers. To quantify the overall relative orientation of the two dimers, we measured the dihedral angle between two $\alpha 2$ helices, one from each copy of H3 or CENP-A, since they are the longest continuous helical structures.

Our results demonstrate that, compared to $(\text{CENP-A/H4})_2$, the two H3 dimers in $(\text{H3/H4})_2$ occupy a range of orientations as the distribution of the dihedral angle between them includes several populations (Figure 3.4B), one positive and two negative, three distinct states in total (Figure 3.4A,B i,ii,iii). This range of orientations for two histone dimers found in our simulations could explain the transitions in tetrasome handedness observed in experiment [123]. A positive dihedral angle corresponds to left-handed superhelically-wrapped DNA, and a negative angle corresponds to right-handed DNA wrapping (Figure 3.4C). Figure S7 illustrates how the angle evolves as a function of simulation time. It shows that, $(\text{H3/H4})_2$ frequently transits from one dihedral angle to another, undergoing swiveling motion around the binding interface, matching the spontaneous flipping behavior of DNA handedness in the tetrasome observed in experiment.

On the other hand, $(\text{CENP-A/H4})_2$ maintains a relatively fixed orientation, with no obvious rotational motions around the interface, as seen in $(\text{H3/H4})_2$. The dihedral angle between the scaffold helices is about 90° (Figure 3.4A,B iv), slightly less than the angle measured in crystal structures of the CENP-A tetramer in the

nucleosome or in the presence of other chaperones (110°). Indeed, from the simulation snapshots, as well as other measurements including overall R_g and R_{COM} between dimers, the two CENP-A/H4 dimers seem to pack more closely together in a twisted orientation, presenting a compact tetramer. Moreover, we observe that, in the absence of DNA and other histones, both H3 and CENP-A histone tetramers prefer not to stay in the same plane compared to the geometries of their respective nucleosome structures (Figure 3.4A). The $\alpha 2$ helices of CENP-A were found to be curved (Figure S13) as was also revealed from hydrogen exchange mass spectrometry [21]. H3 $\alpha 2$ helices were also found to be curved (Figure S12). The curvature of $\alpha 2$ helices could be a result of the absence of surrounding DNA and bracketing H2A/H2B, which provide necessary topological supports to the central tetramer.

3.3.3 Distinct dynamics at the binding interface

We then checked whether the observed difference between $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$ arises from the tetrameric interface (*i.e.* the interface between two dimers). The $Q_{interface}$ distribution for the CENP-A tetramer is centered at 0.5 (where 1.0 cooresponds to the nucleosomal structure), while the same distribution for the H3 tetramer contains three peaks, with an average value of 0.2 (Figure S6.B). This result indicates that $(CENP\text{-}A/H4)_2$ forms a tetrameric interface that is better defined and more native-like compared with $(H3/H4)_2$. Hence, in the context of the DNA-free tetramer, the four-helix bundle of $(CENP\text{-}A/H4)_2$ still maintains a well-connected and symmetric geometric arrangement (Figure 3.5B), despite some

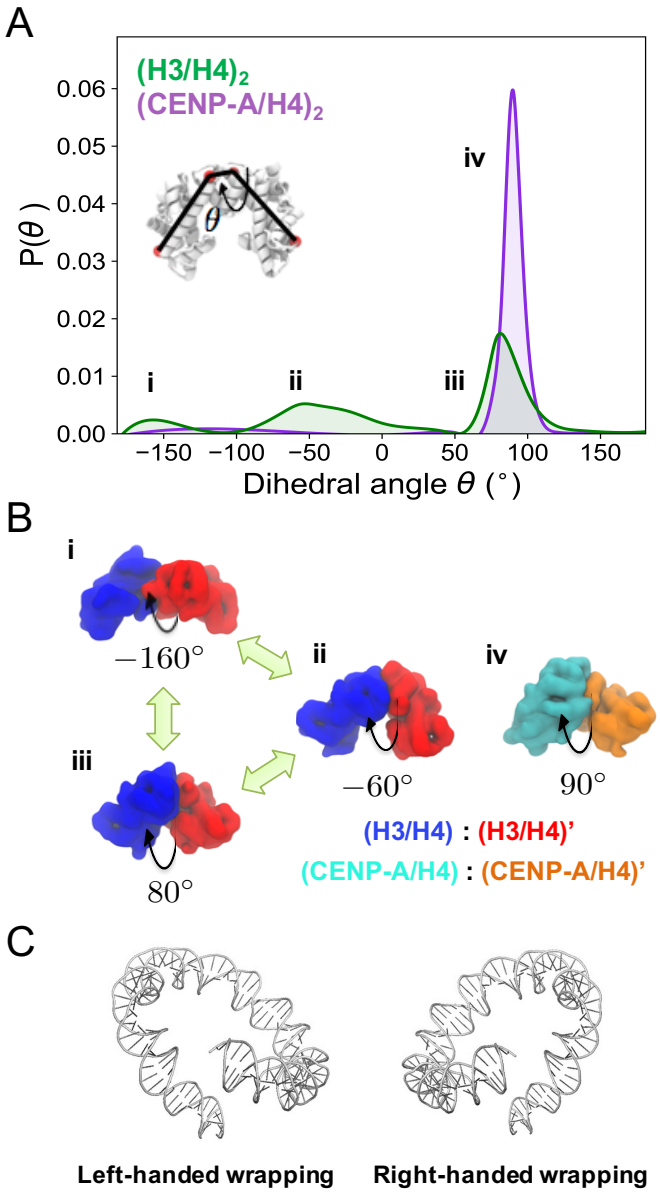

Figure 3.4: **The H3 tetramer swivels around its binding interface while the CENP-A tetramer remains relatively stable.** (A) The distribution for the dihedral angle between $\alpha 2$ helices features one prominent peak for $(\text{CENP-A/H4})_2$, and three smaller peaks for $(\text{H3/H4})_2$, indicating $(\text{CENP-A/H4})_2$ maintains a more fixed orientation than $(\text{H3/H4})_2$. (B) Representative conformations from each population are displayed. (C) Positive and negative dihedral angles of the histone tetramer correspond to left-handed and right-handed DNA superhelical wrapping in the tetrasome, respectively.

structural twisting.

Furthermore, contact analysis of the four-helix bundles demonstrates that there are more contacts, on average, in the corresponding region of $(\text{CENP-A/H4})_2$ ($\sim$27) than in the same region of $(\text{H3/H4})_2$ ($\sim$17) (Figure 3.5). Two peaks are found in the contacts histogram of $(\text{H3/H4})_2$, and only one narrower peak exists for $(\text{CENP-A/H4})_2$ (Figure 3.5A). Detailed residual pair interaction in AWSEM shows that residues Leu111, Gln127, and Arg131 of CENP-A contribute strong hydrophobic interactions to the four-helix bundle tetramer interface (Table S1), which H3 lacks. Large fluctuations observed at the αN sections of histone H3, mainly caused by hydrophobic interactions between Val46, Ala47, Leu48 in αN and Lue107, Ala111 in $\alpha 2$, may play an important role in disrupting the four-helix bundle at H3 tetrameric interface (Figure 3.5B, S11). The distances between the αN helices of both copies of H3 or CENP-A are shown in Figure S8. The distribution for the H3 tetramer features two prominent peaks (at about 20 and 32 Å apart), correlating well with the experimental finding that the H3 αN helix exhibits large structural heterogeneity. On the other hand, $(\text{CENP-A/H4})_2$ maintains a well-defined, native-like four-helix bundle throughout the simulation (Figure 3.5B). Furthermore, CENP-A αN helices remain outside the central interface region, and the distance between them remains relatively far, the distribution of which only includes a single peak (Figure S8). The αN helix of H3 is greater in hydrophobicity than that of CENP-A, which could explain, in part, the reason why H3 αN helices are more-likely to be found close together at the interior of the tetramer than the same helices of CENP-A.

3.3.4 Effects of H2A/H2B on histone tetramers

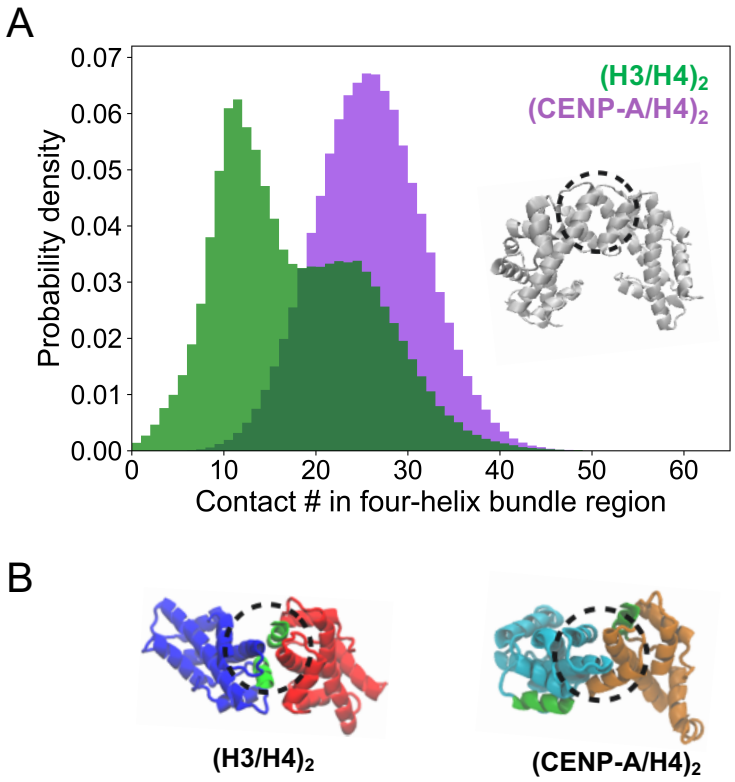

Figure 3.5: **$(CENP\text{-}A/H4)_2$ has a more stable four-helix bundle than $(H3/H4)_2$.** (A) $(H3/H4)_2$ (orange) forms fewer contacts than $(CENP\text{-}A/H4)_2$ (green) in the four-helix bundle region. The histogram of the number of contacts for $(H3/H4)_2$ has two peaks at 13 and 25 while $(CENP\text{-}A/H4)_2$ has a single peak at 27. (B) Corresponding representative structures demonstrate that the $(H3/H4)_2$ four-helix bundle becomes broken or disrupted, while the four-helix bundle of $(CENP\text{-}A/H4)_2$ remains stable throughout our simulations. Four-helix bundles between two histone dimers are circled with dashed lines. αN and $\alpha N'$ helices are marked in green. The dimers of the H3 tetramer are shown in blue and red, and those of the CENP-A tetramer in cyan and orange.

Lastly, to examine the effects of histone dimer H2A/H2B on the dynamics of histone tetramers $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$, we investigated canonical and variant histone octamers, in the absence of DNA. As done for tetramers, continuous constant T simulations and similar analyses were performed to explore the dynamical

features of the histone octamers.

For the H3 octamer, the distribution of both the tetrameric R_g and the distance R between H3/H4 pairs becomes more focused and Gaussian-like, compared to the solo tetramer situation (Figure 3.6). The standard deviation decreases from 3.8 Å to 1.9 Å for R, and from 1.4 Å to 0.7 Å for R_g. The distribution of the dihedral angles between two primary helices of H3 features a dominant peak at 90° (Figure 3.6B), similar to that measured in the case of CENP-A. The other two populations observed in solo H3/H4 tetramer simulations were significantly diminished. These data establish that the swiveling motion around the binding interface was reduced due to the bracketing histone dimers H2A/H2B on either side of the tetramer. Analogous stabilizing effects were not found in the CENP-A octamer case. Interestingly, for the CENP-A octamer, a shoulder and a tail are present in the distributions of R and R_g of the CENP-A tetramer, indicating new conformational flexibility of $(\text{CENP-A/H4})_2$ in the context of an octamer. In particular, the second most populated state has a larger R_g and R than the dominant values observed in the solo CENP-A tetramer (Figure 3.6B). In turn, this implies that the addition of H2A/H2B dimers leads to a less compact association of CENP-A dimers, encouraging the CENP-A tetramer to adopt a geometry closer to that found in the octameric nucleosome. This frustration between the intrinsic compactness of the solo CENP-A tetramer and the expansion and structural twisting induced by the addition of H2A/H2B dimers explains well the observed computational and experimental findings of that CENP-A-containing histone nucleosomes or octamers being structurally more flexible and heterogeneous than their canonical counterparts [6,7].

3.4 Discussions

3.4.1 Tetramers vs. dimers vs. octamers

Previously, we reported that, in the context of a dimer, histone H4 is more native-like than its binding partner H3 or CENP-A, and that the CENP-A/H4 dimer is more dynamic than its canonical counterpart H3/H4 [19]. In this work, in the context of a tetramer, analyses of the dimer and monomer components of the central tetramer (see SI section S10) yielded consistent results with the previous study. For instance, Q_{dimer} and $Q_{interface}$ for the H3 dimer are larger, on average, than for the CENP-A dimer (Figure S9), indicating that H3 dimers adopt more native-like conformations than CENP-A dimers. The average Q value of H4 (Figure S10) is larger than that of H3 or CENP-A in all tetramer simulations, which implies the noticeable stability of histone H4.

However, compared to the structural variabilities at the dimer and monomer level, the movements between dimers forming the tetramer are on a larger scale, with an RMSD of 10-15 Å for the tetramer (Figure S4A, S5A, S6A) compared to an RMSD of 3-4 Å for the dimer (Figure S9B, and Figure 2 in [19]). Therefore, the dynamics observed here could be challenging for atomistic simulations to reach due to practical sampling limitations.

Moreover, in our previous study, the CENP-A nucleosome was revealed to be more flexible than the H3 nucleosome, leading to a shearing motion at the tetramer interface [141]. Here, in the context of an octamer with H2A/H2B dimers, the

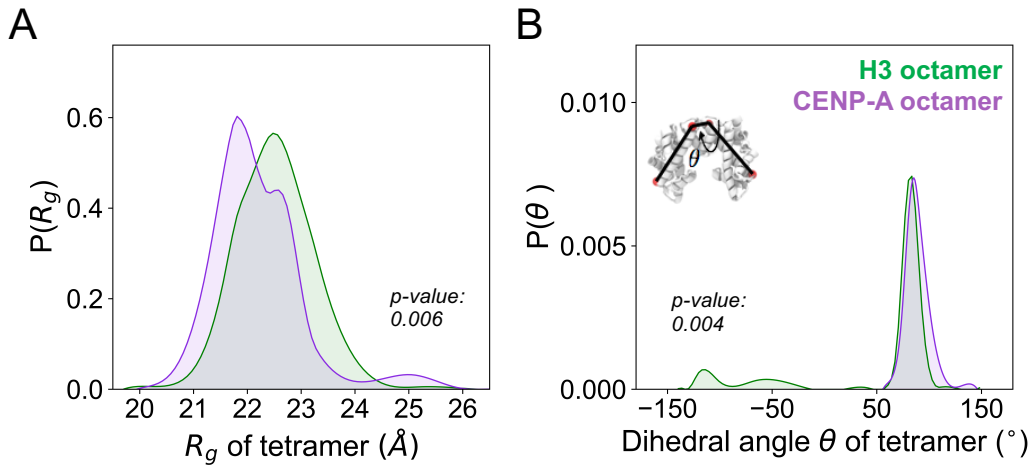

Figure 3.6: **H2A/H2B stabilizes $(H3/H4)_2$ but not $(CENP\text{-}A/H4)_2$.** (A) The probability distribution of H3 tetramer R_g features a more focused peak in the context of an octamer compared to that of the solo H3 tetramer (Figure 3.3C), while one peak and one shoulder exist in the same distribution for the CENP-A tetramer in the context of an octamer. (B) Distributions of the dihedral angle between $\alpha 2$ helices demonstrate that, in the presence of H2A/H2B, $(H3/H4)_2$ becomes more similar to $(CENP\text{-}A/H4)_2$; both curves feature a prominent peak around 80°.

CENP-A tetramer occupies two distinct conformational states: one is similar to that of the isolated tetramer conformation while the other state is less compact, close to the H3 (or CENP-A) tetramer resolved in the octamer. Hence, the CENP-A tetramer is intrinsically compact. This intrinsic compactness and overall shape of the tetramer tend to be tuned or corrected when interacting with (H2A/H2B)s in an octamer. As a result, disrupting the energetically stable interface of the CENP-A tetramer could be an underlying source of the shearing motion observed for octameric nucleosomes. The two-state nature of the CENP-A tetramer in the octamer may be a coarse-grained view of CENP-A's shearing motion in nucleosomes and explains why the CENP-A nucleosome appears to be actively dynamic while canonical H3 is more static.

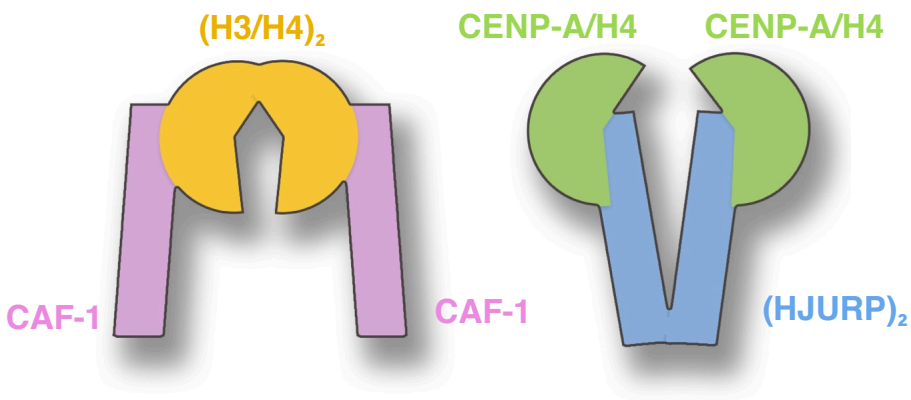

Figure 3.7: **Suggested models for histones and their chaperones during deposition.** (A) H3/H4 may be deposited in the form of a tetramer with each external side bracketed by a CAF-1 chaperone, which may stabilize the tetramer. (B) CENP-A may be deposited as dimers; each dimer loaded by one HJURP chaperone.

3.4.2 Biological implications

We would like to suggest several potential biological implications of our investigation. First, this work emphasizes the importance of structural context for the canonical H3 tetramer, which, *in vivo*, interacts with DNA, histones, or chaperone proteins. The canonical tetramer may have evolved to be unstable, and this instability may relate to its role in the nucleosome. In other words, the structural dependence of the H3 tetramer may be key to ensure the fidelity and stability of genetic material. On the other hand, CENP-A, as a functional variant histone, has intrinsic stability in its tetramer form, and is therefore less dependent on DNA or other proteins, which may be needed for its diverse functions and unique assembly pathway.

Our results support the possibility that the stably formed CENP-A tetramer could regulate the availability of individual CENP-A dimers, which we previously

found to be flexible and could easily encounter other proteins [19]. The rigidity of the CENP-A tetramer could prevent CENP-A from associating with non-centromeric proteins, so as to avoid the ectopic localization which is often observed in cancer cells. A speculative proposal is that the $(\text{CENP-A/H4})_2$ tetramer may serve as a sequestration channel, needed to maintain CENP-A homeostasis.

Another hypothesis based on this research is that the tetramerization of two CENP-A dimers could be nearly irreversible, so that the CENP-A tetramer, once formed, may not be able to separate into two dimers afterwards, even in the presence of chaperone HJURP (more simulation results about HJURP are in section S11). In this scenario, the DNA-free protein tetramer might serve as a kinetic trap for excess CENP-A. This hypothesis sheds light on the unique assembly/disassembly pathway of the CENP-A nucleosome. The CENP-A tetramer may be just one state along the assembly pathway of CENP-A nucleosome, after being delivered by HJURP, given the experimental evidence that the CENP-A–CENP-A′ interface is a requirement for stable chromatin incorporation [112].

The CENP-A-specific chaperone HJURP, by competing for the same binding site, the internal side of CENP-A tetramer, blocks CENP-A dimers from self-associating into a tetramer. It has been shown that two HJURP chaperones [**?**] and the dimerization of HJURP [142] is required for proper CENP-A nucleosome assembly. Therefore, a HJURP dimer may interact with two isolated CENP-A dimers, instead of with a CENP-A tetramer (Figure 3.7B). On the contrary, as in the structure of H3 and its chaperone CAF-1 [143,144], CAF-1 binds with an H3 dimer at the external side, without the possibility of preventing it from forming a tetramer. In-

deed, the kinetically less stable tetramer of H3 may need the enhanced stabilization *via* binding with CAF-1 chaperones at either side (Figure 3.7A). Taken all together, we propose two different chaperone models for CENP-A and H3 assembly, CENP-A/H4–$(HJURP)_2$–CENP-A/H4 vs CAF-1–$(H3/H4)_2$–CAF-1 (Figure 3.7), with a subtle yet important difference: in the former, two copies of HJURP would prevent two CENP-A dimers from forming a tetramer in pre-assembly complexes, whereas, in the latter, CAF-1 proteins would stabilize a pre-formed H3 tetramer in preparation for subsequent nucleosome assembly. Our results support the possibility that canonical H3- and CENP-A-containing nucleosomes may have orthogonal assembly pathways: while H3 could be deposited as a tetramer, CENP-A may be loaded in the form of a dimer.

3.5 Conclusion

In summary, this work establishes that variant histone CENP-A thermodynamically favors a tetramer formation while the canonical H3 presents remarkable rotation dynamics about the tetramer interface, which results in a rugged yet shallow binding free energy landscape. Moreover, H2A/H2B dimers restrain the internal motion of $(H3/H4)_2$ and lead to multiple states for $(CENP\text{-}A/H4)_2$, providing a possible physical explanation for the shearing motion observed for the CENP-A nucleosome [141]. These findings offer a new perspective on the structural debate over CENP-A, and may shed light on CENP-A's unique dynamics. Based on our results and related research, we suggest two different chaperone models for H3 and

CENP-A. Lastly, we propose that the $(\text{CENP-A/H4})_2$ tetramer may serve as a sequestration channel *in vivo*, which would provide another layer of regulation to ensure the proper homeostasis of CENP-A.

Chapter 5: Summary and Future Prospects

This thesis reported on our investigations of the biophysical mechanisms of the assembly dynamics of histone proteins, from individual dimers to tetramers and octamers. Computer simulations are the primary methodology used. Besides the assembly principles that govern histone oligomers, this series of works elucidated the differences in the dynamics of canonical and variant CENP-A histones, revealing the nucleosomal dynamics needed for properly functioning centromere. Overall, through molecular simulations, this work deepened our understanding of the structure-function relation of histone proteins, which play an essential role in organizing the genetic information of eukaryotic organisms.

One of the overarching goals for our research is to provide a molecular understanding, at the basic level, for the higher-order structural dynamics of chromatin packaging and chromatin-related dynamics. Having this as a long-term aim, we started from histone proteins, the scaffold carrier proteins required for DNA wrapping. Though current simulations do not involve DNA, the protein dynamics revealed here provide insightful details and knowledge for the subsequent study, which may include more complicated protein-DNA interactions, both computationally and experimentally.

The research in this thesis demonstrates that the AWSEM model is a proper and efficient coarse-grained protein model to simulate histone protein systems, even on the scale of a histone octamer. With more computational resources and better parallel algorithms, the AWSEM model is expected to be utilized for larger-scale molecular simulations, such as nucleosomal organizations, nucleosome-nucleosome association and ultimately nucleosome arrays. Future computational work could first focus on the DNA interaction with histone proteins. For example, one future work extended from the work in Chapter 2 is to use an enhanced sampling method with AWSEM to obtain the free energy profile of the histone-DNA association, where the energy minima may follow the topology of the histone-fold geometry.

In previous studies, due to the lack of DNA model, all histone tails were excluded based on the fact that DNA is essential for stabilizing histone tails. Therefore, two technical challenges may arise: the combination of DNA and the current protein model, and the model for simulating histone tails, which are typically intrinsically-disordered proteins. The 3SPN DNA model, shown to be a good candidate for DNA modeling, was previously utilized to study the nucleosome. The 3SPN DNA model [145], which was previously utilized to study the nucleosome disassembly [146], has been showed to be a good candidate for DNA modeling. However, when merging 3SPN model with the AWSEM protein model, some details like the DNA charges need to be reparameterized in order to precisely calculate the electrostatic interaction between DNA and protein. The recently developed AWSEM-IDP model in the Papoian group holds promise to be added for the purpose of modeling IDPs. With well-resolved DNA and IDP modeling, potential projects extended

from Chapter 4 include investigating the dynamics of the tetrasome, consisting of the histone tetramer and DNA. One interesting question is to study the topology of the wrapping DNA. Previous optical tweezer experiments [147] showed that the supercoiling state of DNA determines the handedness of both H3 and CENP-A nucleosomes. However, in our simulation, the H3 tetramer by itself displayed different flipping behavior, potentially governing the handedness of DNA wrapping. Hence, it would be interesting to compare if the dominating role of the DNA topology is due to the DNA itself or the protein dynamics, and compare in this regard between the canonical and variant histones like CENP-A. Another noteworthy point is the role of histone tails in mediating DNA-histone dynamics. In the supplemental work of chapter 3, we found that histone tails did not lead to a significant difference in the stability of H3 and CENP-A dimers. But different results may be observed when considering the effects of tails in a molecular system that includes DNA.

Another highlight of this thesis is the illustration of the differences in the dynamics of canonical and the centromere protein A (CENP-A) related histone assemblies. We found that the CENP-A-containing nucleosome and CENP-A/H4 dimer presented greater heterogeneity than canonical H3-containing nucleosome and dimer, respectively. This plasticity of the whole CENP-A nucleosome revealed here, and in another work from the Papoian group [7], may inspire the computational and experimental studies of the kinetochore assembly process. For instance, in Chapter 3, we found that the CENP-A C-terminus is protected by the chaperone HJURP during CENP-As delivery to the centromeric region, and we hypothesized that the binding availability of CENP-As C-terminus to regions outside the centromere is

carefully regulated by chaperone HJURP. Studies from both computational and experimental sides can examine the binding behavior of CENP-A with many other centromeric proteins, such as CENP-C and CENP-N, with and without the presence of HJURP, so as to provide more insights into kinetochore assembly.

Finally, the current study disclosed the crucial functions of chaperones in regulating the canonical nucleosome assembly as well as CENP-A-involved assemblies. For example, in Chapter 4, we suggest that during the formation of tetrasome, the H3 tetramer may need two CAF-1 chaperones from each external side to stabilize the central tetramer, allowing the whole tetramer to be deposited onto the DNA. In chapter 2, we suggest that CENP-A is deposited in the form of a dimer while H3 in the form of a tetramer. Indeed, many other histone chaperones, such as chaperones HIRA, Daxx, and Asf1a, play essential roles in regulating histone and nucleosome assemblies. The mechanistic insights on how these chaperones work on canonical and different variant histones are currently lacking. Specifically, how do these chaperones work in symphony to maintain the overall homeostasis of CENP-A in the cell since the overexpressed CENP-A could lead to ectopic deposition and subsequent cancer or tumor development? Experimental and computational works could investigate interactions between histone proteins and the aforementioned chaperones when their crystal structures become available in the future. Theoretical work can also start from the perspective of systems biology to investigate the complex regulating network of CENP-A. Altogether, the combination of theoretical, computational, and experimental methods is expected to bring comprehensive understanding of the histone variant chromatin landscape in the genome.

Appendix A: Supplementary Information for Chapter 3

A.0.1 RMSIP calculation

The root-mean-squared inner-product (RMSIP) is a measure of the amount of overlap between two samples. RMSIP is a normalized parameter, where 1 indicates completely overlapping sets and 0 means completely independent sets. To evaluate convergence, we calculated the RMSIP between the data sets corresponding to two halves of increasingly higher percentages of the entire simulation trajectories. The first ten eigenvectors with largest eigenvalues were used (Equation S1), based on the x, y, and z positions of the Cα atoms. For each point along one simulation trajectory, we divided the preceding time into 2 halves and calculated the RMSIP value between these two subspaces. In CG simulations, the RMSIP for every individual run was computed, starting by analyzing the first 10 ns, then the first 20 ns, and so forth. For all-atom simulations, we considered the trajectories starting from 400 ns: we calculated RMSIP first for 400 ns to 430 ns, then for 400 ns to 460 ns, and so on. All the RMSIP values are over 0.8 in CG simulations after 10 ns, and in all-atom it stays over 0.75 after 400 ns, indicating adequate convergence for both CG and all-atom MD simulations.

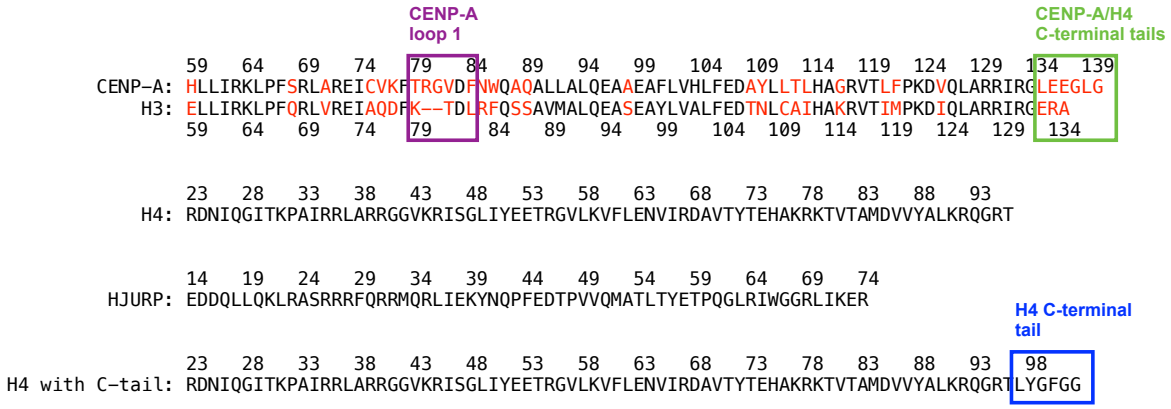

Figure A.1: **Amino acid sequence alignment.** The amino acid sequences of CENP-A, H3, H4, and the CENP-A-specific chaperone HJURP provide the primary level of description for the protein structures of the CENP-A/H4 and H3/H4 dimers, with and without HJURP, considered in this study. The amino acid sequence alignment of CENP-A and canonical H3 reveal CENP-A contains a longer loop 1 (purple box) and C-terminal tail (green box) than its canonical counterpart. Specific residue differences between CENP-A and canonical H3 are shown in red. Results from simulations including "H4 with C-tail" are only included here in the Supplementary Information, where the additional residues considered are identified by the blue box.

$$RMSIP = \left(\frac{1}{10} \sum_{i=1}^{10} \sum_{j=1}^{10} (\vec{\eta}_i \cdot \vec{\nu}_j) \right)^{\frac{1}{2}}, \tag{A.1}$$

where $\vec{\eta}_i, \vec{\nu}_j$ are the ith and jth eigenvector of the first and second half of the considered trajectory, respectively. The first ten eigenvectors with significant eigenvalues are used.

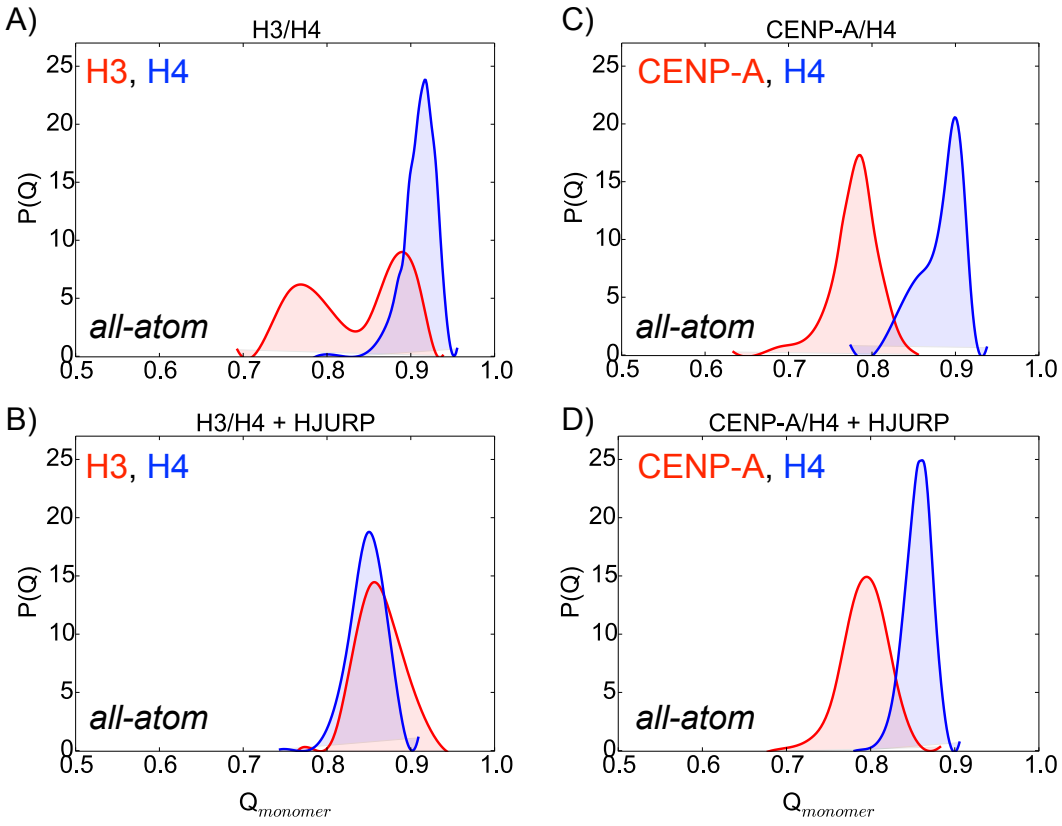

Figure A.2: **H4 is structurally consistent in all-atom MD simulations.** $Q_{monomer}$ analysis of the all-atom MD trajectories of (A) isolated H3/H4, (B) H3/H4 in conjunction with HJURP, (C) isolated CENP-A/H4, and (D) CENP-A/H4 in a complex with chaperone HJURP reveals qualitative agreement with the AWSEM coarse-grained MD trajectories. H4 adopts conformations closer to the native state (*i.e.* the experimentally determined crystal structure) than CENP-A or canonical H3 for every all-atom system studied except for H3/H4 in conjunction with HJURP, where histones H3 and H4 are equally close to their respective native states.

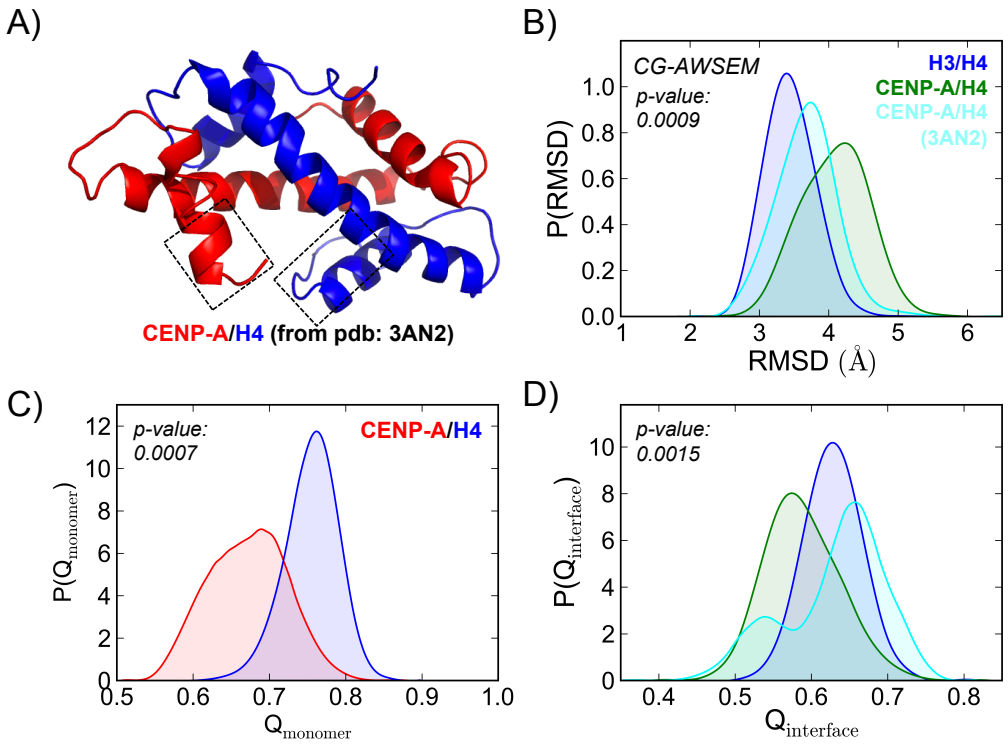

Figure A.3: **CG simulation of the CENP-A/H4 dimer from the CENP-A nucleosome crystal structure.** (A) The CENP-A α3 helix (Box on red) is not fully resolved in CENP-A nucleosome crystal structure (PDB ID: 3AN2). However, this structure does include the H4 C-terminal tail (Box on blue). (B) Without the fully-extended CENP-A α3 helix (*i.e.* the CENP-A α3 helix resolved in the CENP-A/H4/HJURP structure, PDB ID: 3R45), the H4 C-terminal tail does not increase the RMSD of CENP-A/H4. (C) $Q_{monomer}$ analysis illustrates that H4 still adopts more native-like conformations than CENP-A. (D) The binding interface of CENP-A/H4^{3AN2} (cyan) has two peaks, compared to one for CENP-A/H4^{3R45} (green), demonstrating that the H4 C-terminal tail is unstable and disrupts the binding interface of CENP-A/H4.

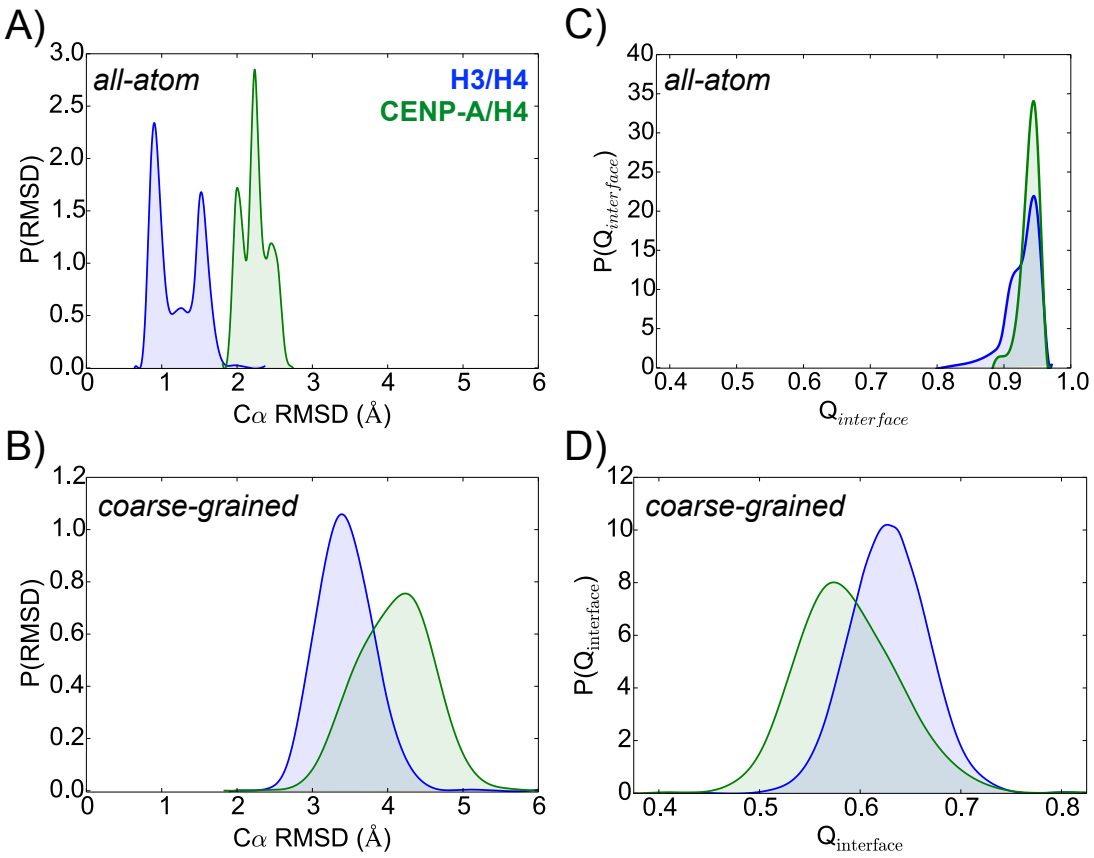

Figure A.4: **All-atom and CG-AWSEM MD results qualitatively agree and play complementary roles in analysis.** (A) Probability distribution functions of the Cα root-mean-square deviations (RMSD) for the all-atom MD simulation trajectories reveal that replacing H3 with CENP-A leads to greater structural variability in the heterodimer. (B) Probability distributions of the interface Q indicate that both the CENP-A and H3 dimers adopt conformations close to the native state (*i.e.* Q=1) in all-atom simulations. (C) RMSD probability distribution functions for the CG-AWSEM simulations demonstrate that CENP-A/H4 is more conformationally variable than H3/H4, an example of the overall qualitative agreement between all-atom and CG-AWSEM MD. (D) Centered at lower averages, with wider variances, compared to all-atom results, the Q interface probability distributions for CG-AWSEM illustrate that coarse-grained MD explores more conformational space further from the native state than all-atom MD.

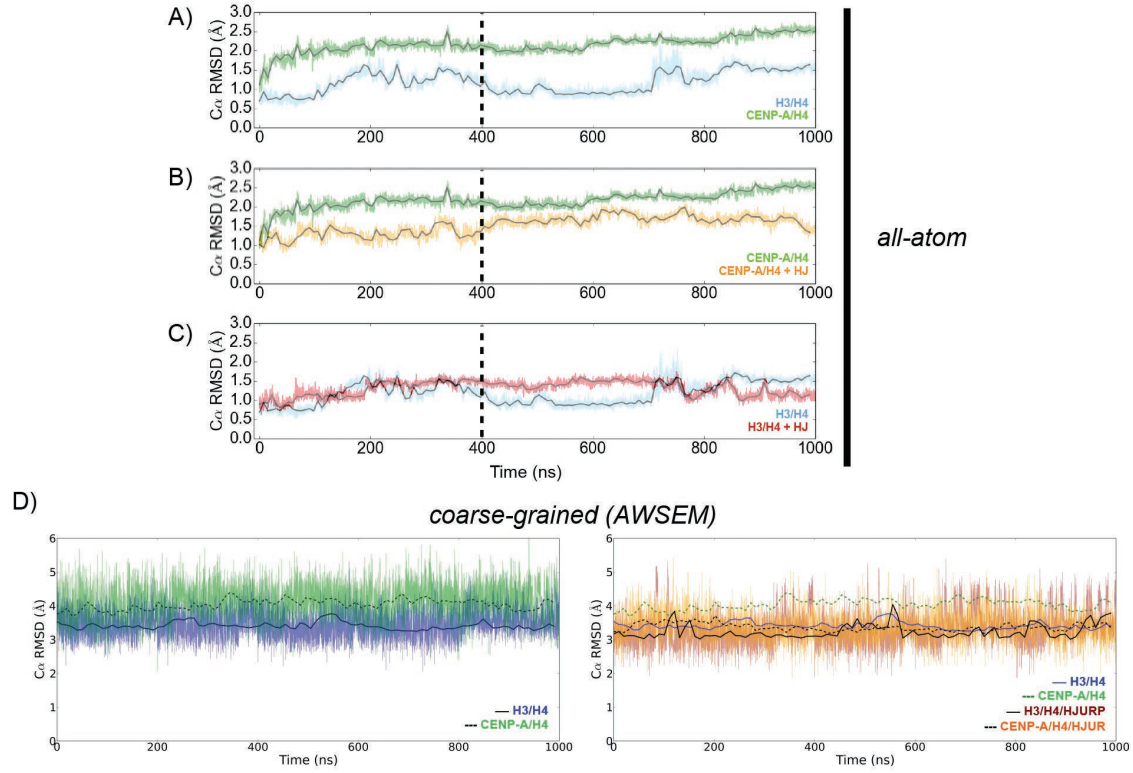

Figure A.5: **RMSD illustrates that both all-atom and CG simulations reached convergence.** We first examined the convergence of the all-atom and AWSEM coarse-grained MD trajectories by calculating the Cα RMSD of the simulation snapshots with respect to their positions in the experimentally-determined crystal structures as functions of simulation time. In the all-atom simulations, (A) the isolated CENP-A/H4 dimer is more structurally variable than H3/H4 in isolation; (B) the introduction of HJURP reduces the structural variation of the CENP-A/H4 dimer, bringing it closer to the native state, and (C) the presence of HJURP is not an important factor in determining the structural heterogeneity of the canonical H3/H4 dimer. Every all-atom system studied reaches convergence by 400 ns of simulation time (represented by the dashed, vertical lines), therefore only the final 600 ns are used for analysis. (D) Five independent 200 ns CG-AWSEM simulation trajectories were performed for each system, summing to 1000 ns of total CG simulation time. The CG-AWSEM simulations rapidly reach equilibration, therefore we combined those trajectories for further analysis after removing the first 10 ns from each 200 ns run.

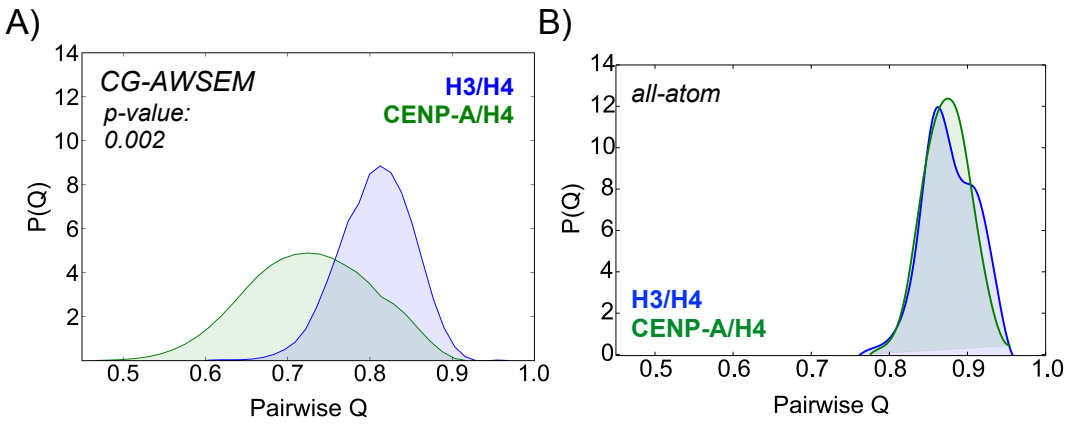

Figure A.6: **Pairwise Q value demonstrates that CENP-A/H4 has greater conformational heterogeneity than H3/H4 in CG-AWSEM simulations.** Pairwise Q is when the Q value is calculated between every two conformations from the same simulation, Instead of comparing the simulation conformations to the experimentally determined crystal structure. For each simulation, pairwise Q is calculated pairs of 1000 snapshots, chosen every 500,000 timesteps, corresponding to 1ns. (A) In CG-AWSEM simulations, the pairwise Q distribution for CENP-A/H4 (green) is broader and lower on average than that of H3/H4 (blue), implying that CENP-A/H4 is more conformationally heterogeneous than H3/H4. (B) On the other hand, in all-atom simulations, pairwise Q for both H3/H4 and CENP-A/H4 are high and narrowly distributed, implying that all-atom simulation probes dynamics near the native-state and samples relatively limited conformational space compared to CG simulation.

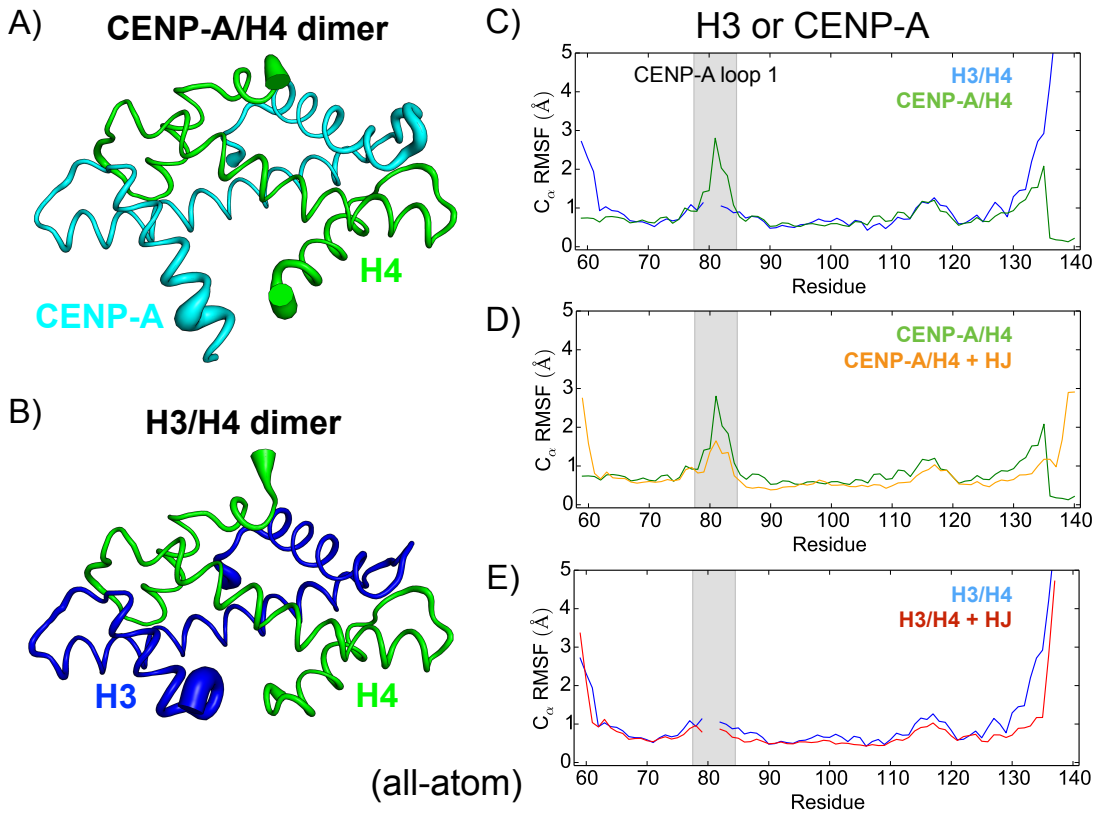

Figure A.7: **All-atom local mobility by RMSF.** Root-mean-squared fluctuations (RMSF) are a measure of local mobility. Cα RMSF, with respect to the geometric centers, of the all-atom MD simulation snapshots projected onto the crystal structures of (A) the CENP-A/H4 dimer, and (B) the H3/H4, where the tube width is proportional to RMSF, reveals that CENP-A loop 1 exhibits greater local mobility than the same region of canonical H3. (C) In isolated dimers with H4, CENP-A local mobility is only significantly greater than that of H3 at loop 1, except for the highly variable terminal regions. (D) The introduction of HJURP slightly reduces the local flexibility of CENP-A, stabilizing CENP-A loop 1. (E) The presence of HJURP has only a minimal effect on the local mobility of canonical H3.

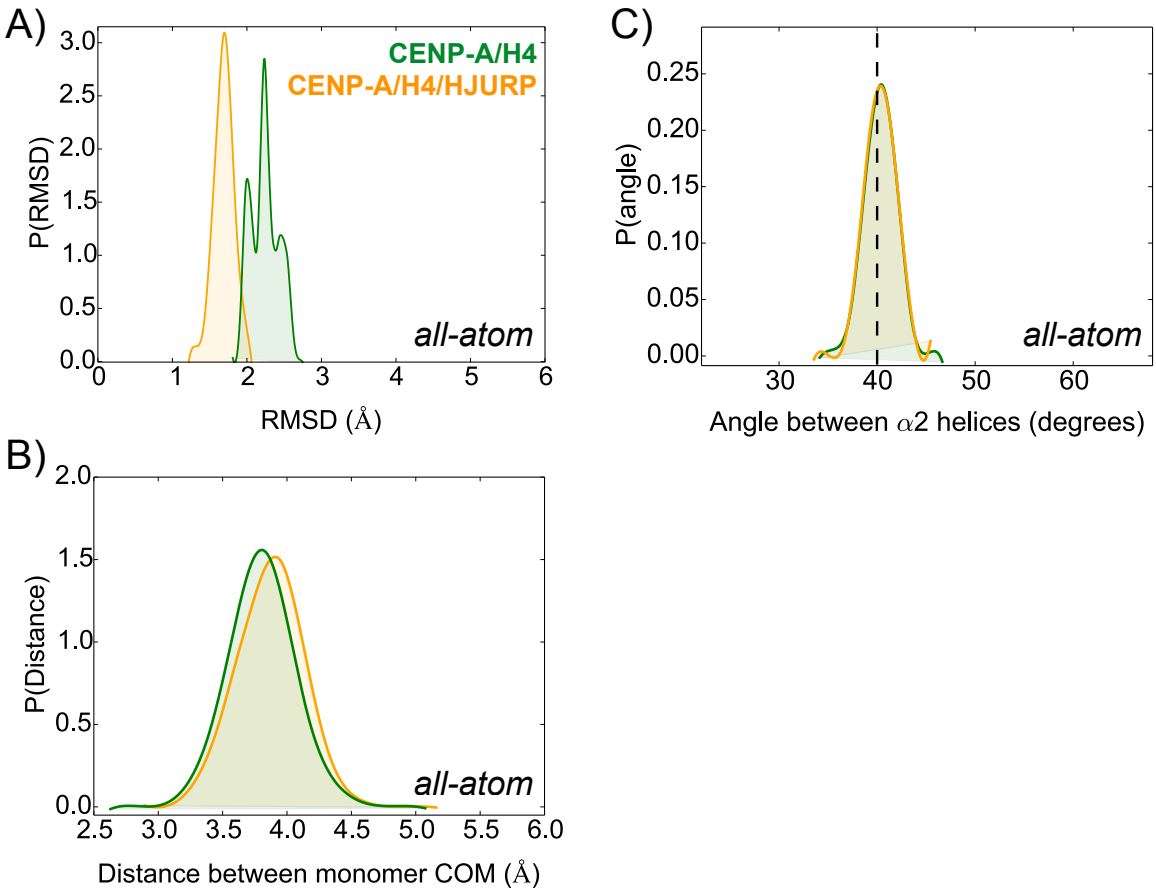

Figure A.8: **Global preferences do not change significantly upon the introduction of HJURP to the CENP-A/H4 dimer in all-atom simulations.** (A) Upon the introduction of HJURP, the Cα RMSD of CENP-A/H4 decreases, adopting a conformation closer to the 3R45 crystal structure conformation. However, for all-atom MD, the introduction of HJURP does not significantly influence (B) the distance between histone centers-of-mass or (C) the angle between the central α2 helices. Namely, adding HJURP does not change the global preferences of CENP-A/H4 in all-atom MD simulations.

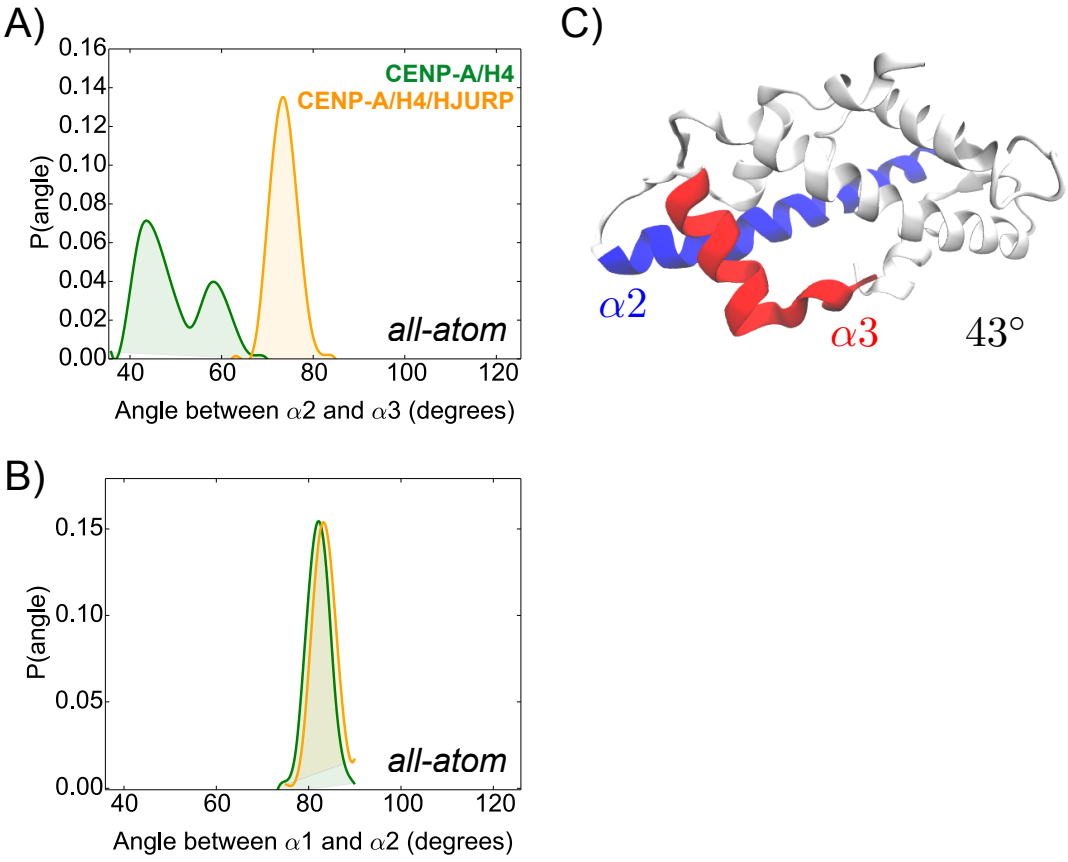

Figure A.9: **The introduction of HJURP stabilizes CENP-A α3 in all-atom simulations.** (A) In the absence of HJURP, the angle between CENP-A helices α2 and α3 adopts a bimodal distribution, with two peaks at about 43 and 60 degrees. Upon the introduction of HJURP, this angle becomes relatively fixed, in qualitative agreement with CG-AWSEM MD results. (B) Furthermore, the angle between CENP-A helices α1 and α2 remain the same whether HJURP is present or not, also agreeing with the results from CG simulation. (C) A representative all-atom simulation snapshot of the first peak in the α2-α3 angle distribution reveals that CENP-A α3 becomes partially unraveled in the absence of HJURP.

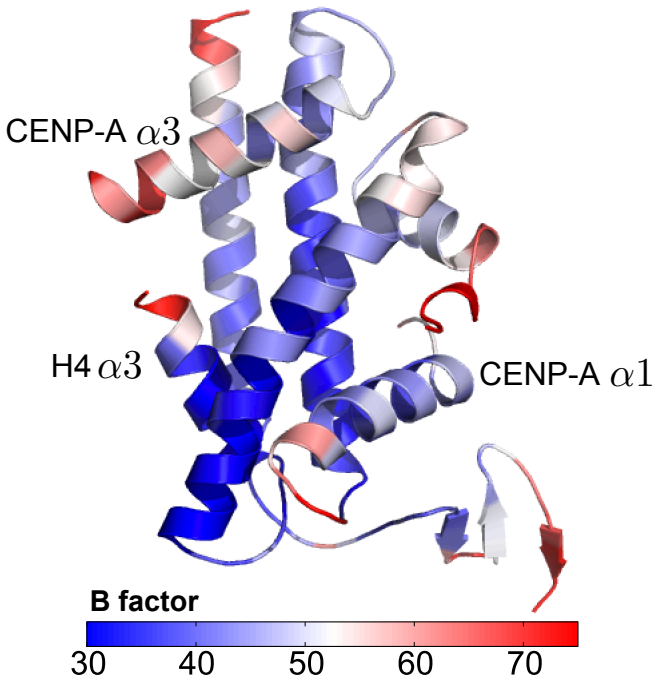

Figure A.10: **B-factor-colored crystal structure highlights CENP-A $\alpha 3$ and H4 C-terminal residues as regions of high local mobility.**

A.0.2 Angle Analysis

To obtain the angle between two α helices, we first calculate the orientation vector for each selected helix, using the coordinates of C_α. A variance matrix V is created:

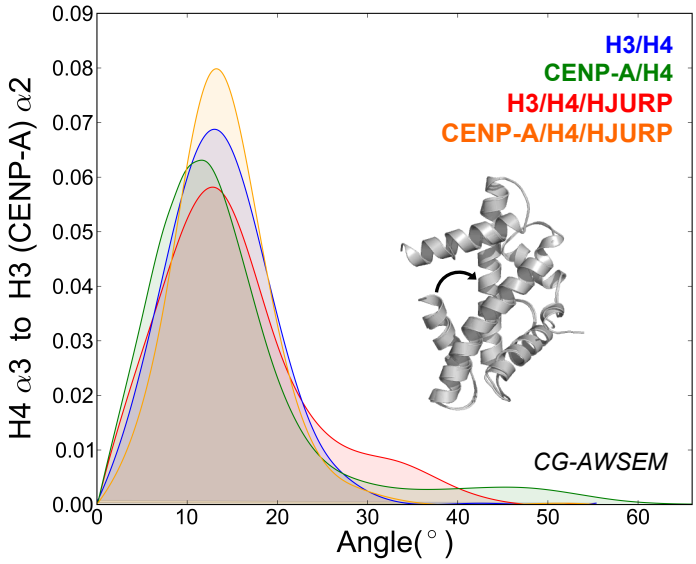

Figure A.11: **The angle between H4 α3 and H3 (CENP-A) α2 helices is mostly stable in the absence of H4 C-terminal tail in CG simulations.** For all CG simulations of H3/H4 (blue), CENP-A/H4 (green), H3/H4/HJURP (red), CENP-A/H4/HJURP (orange), the angle between the H4 α3 and H3 (CENP-A) α2 helices is mostly stable. Notice that, due to the flexible C-terminal, the angle distribution for CENP-A/H4 has a slight shoulder based on the interactions between CENP-A C-terminal and the C-terminal end of H4 α3, consistent with all-atom contact analysis (Figure 5). Furthermore, upon the introduction of HJURP, this shoulder disappears, in agreement with the role of HJURP revealed in this paper: stabilizing and regulating the CENP-A C-terminal. Lastly, when introducing HJURP to canonical H3/H4, the angle between H4 α3 and H3 α2 adopts a broader probability distribution, suggesting that HJURP may disrupt the binding interface between H3 and H4. The curved arrow shown with the structure identifies the angle measured.

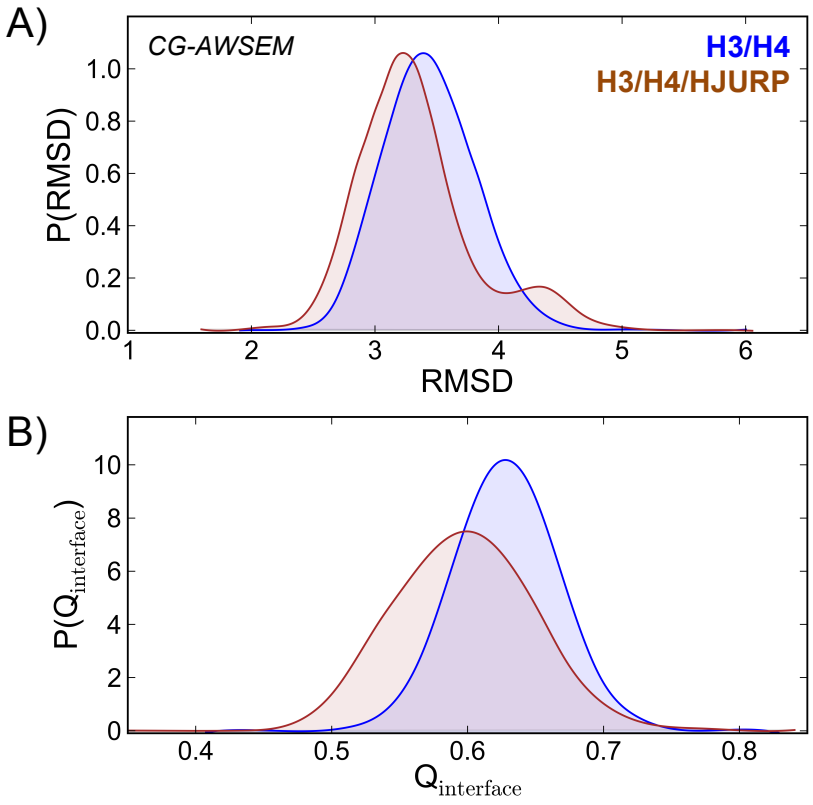

Figure A.12: **HJURP disrupts the binding stability of H3 and H4 in CG simulations.** (A) RMSD probability distributions demonstrate that the introduction of HJURP slightly increases the average overall deviation of the canonical H3/H4 dimer from the experimentally determined crystal structure, and leads to a subpopulation of conformations further from the native state (at ∼4.5 Å RMSD). (B) Furthermore, upon the introduction of HJURP, the binding interface between H3 and H4 becomes less native-like, adopting lower $Q_{interface}$ values on average than when HJURP is absent.

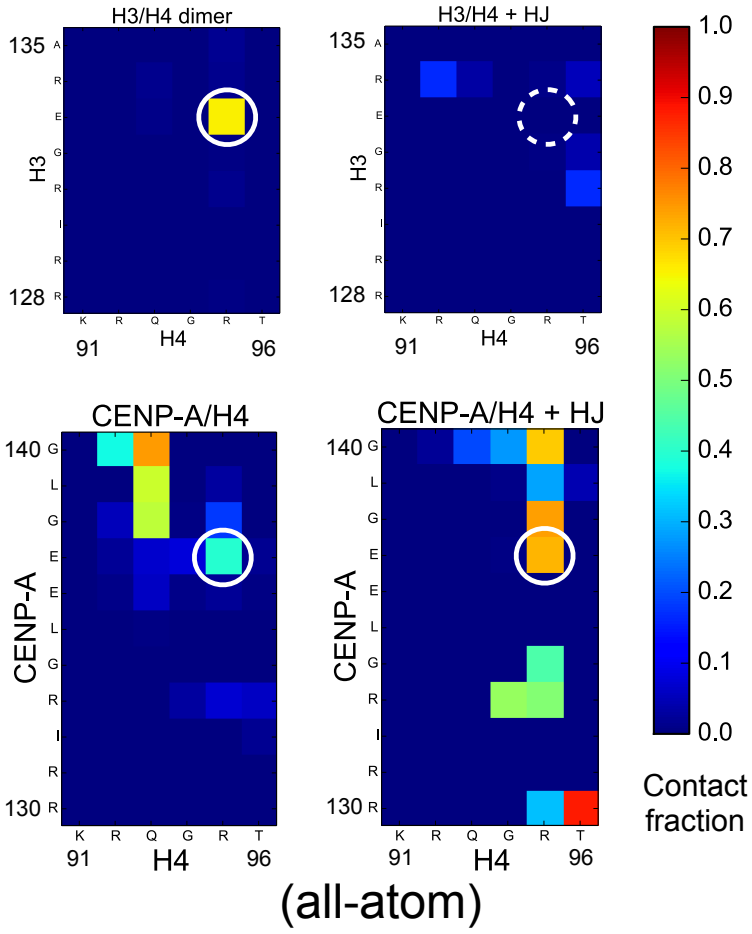

Figure A.13: **HJURP stabilizes interactions between the C-termini of CENP-A and H4, but not between H3 and H4, in all-atom simulations.** In isolation, one salt-bridge dominates the interactions between the C-termini of H3 and H4, H3 E133 to H4 R95, whereas the C-termini of CENP-A and H4 form several different contacts, including a salt-bridge between CENP-A E137 and H4 R95. Upon the introduction of HJURP, the C-termini of CENP-A and H4 form even more interactions, while the contacts between the C-termini of H3 and H4 become disrupted.

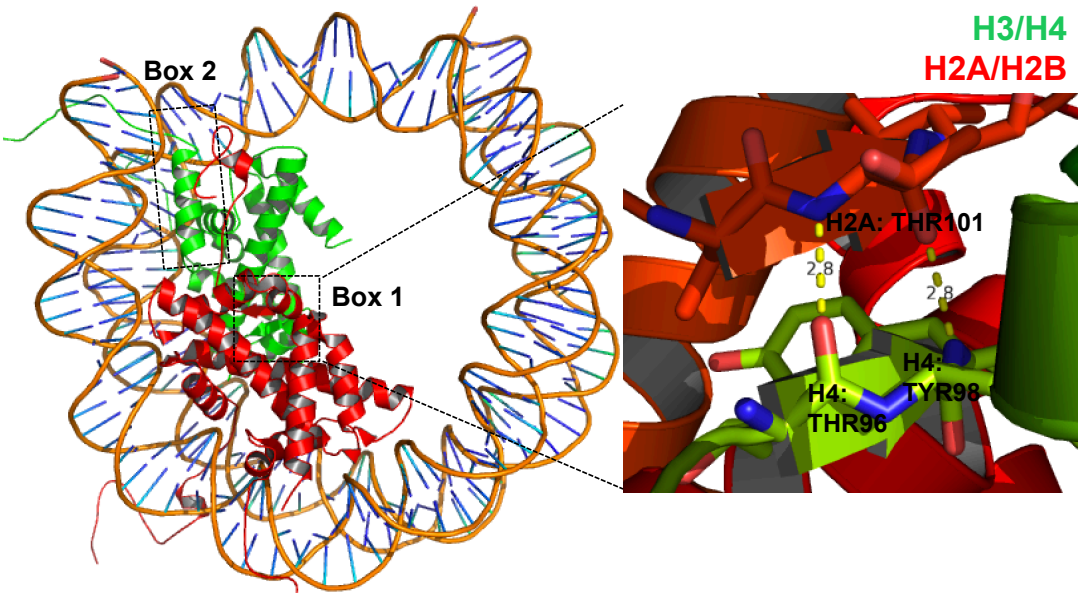

Figure A.14: **Histone tails and H3 (CENP-A) αN helix primarily interact with DNA and other histones.** Colors identify histone dimers H3/H4 (green) and H2A/H2B (red) in a typical nucleosome structure (PDB ID: 1KX5). In this structure, the H3 αN helix (Box 2) largely interacts with DNA and the H2A histone tail. Additionally, in the nucleosome context, the H4 C-terminal tail region forms a β strand, between H4 THR96/TYR98 and H2A' THR101, shown in Box 1 and in the zoomed-in view.

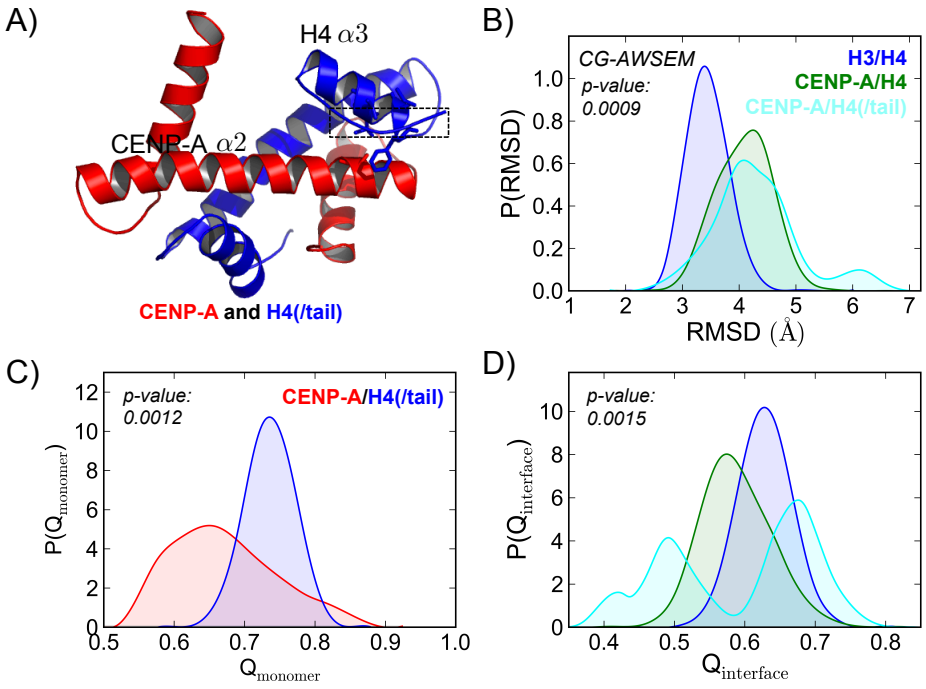

Figure A.15: **Including the H4 C-terminal tail increases the structural flexibility of the CENP-A/H4 dimer.** The initial conformation of the "CENP-A/H4(w/tail)" simulation is composed of CENP-A from the CENP-A/H4/HJURP structure (PDB ID: 3R45) and H4 from the CENP-A nucleosome structure (PDB ID: 3AN2) after structural alignment. (A) The boxed area in the structure figure illustrates where the H4 C-terminal tail has hydrophobic interactions with H4 $\alpha 3$ and CENP-A $\alpha 2$. (B) CENP-A/H4 with the H4 C-terminal tail included has a larger average RMSD with respect to the crystal structure than that of the CENP-A/H4 dimer structure excluding the H4 C-terminal tail. (C) $Q_{monomer}$ analysis demonstrates that, even with the H4 C-terminal tail included, H4 still remains more native-like than CENP-A. (D) Lastly, the Q interface probability distributions show that, when adding the H4 C-terminal tail, the binding interface of CENP-A/H4 is no longer stable, clearly adopting multiple different conformational states.

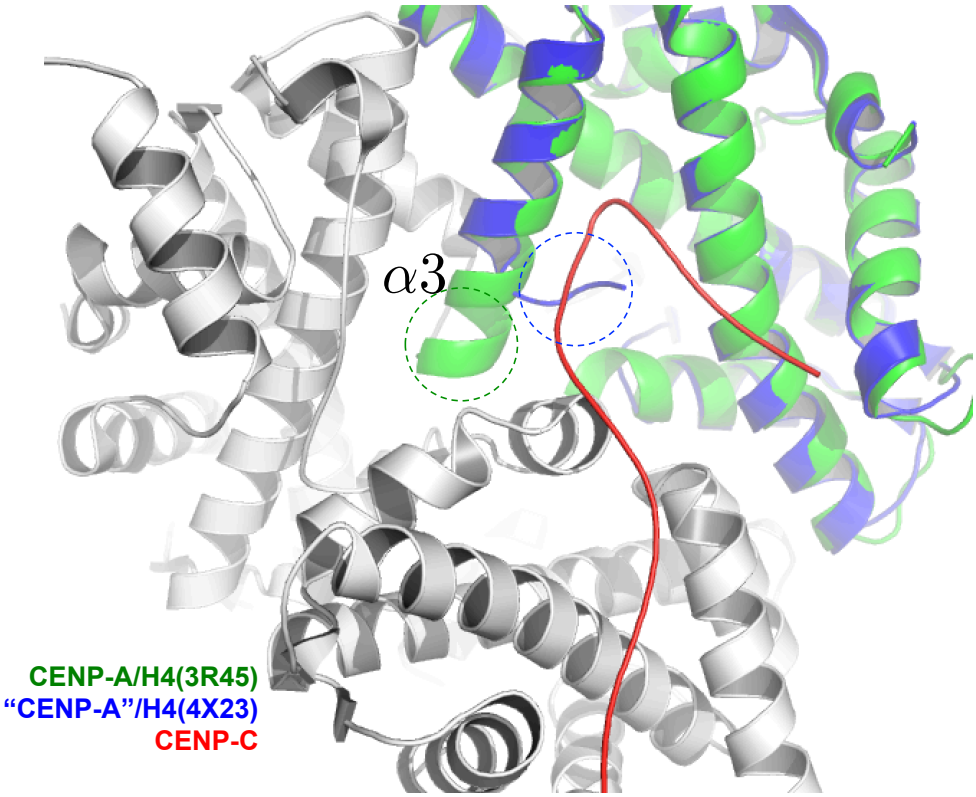

Figure A.16: **Alignment of CENP-A structures from different contexts show the "on" and "off" states of its C-terminal tail.** The CENP-A/H4 structure from CENP-A/H4/HJURP crystallography (PDB ID: 3R45) is shown in green, featuring an ordered C-terminal tail (circled in green) at the end of $\alpha 3$, corresponding to the "off" binding state. The blue, gray and red structures are all from the chimeric-nucleosome/CENP-C complex (PDB ID: 4X23). One chimeric "CENP-A/H4" is shown in blue, containing the C-terminal residues of CENP-A (circled in blue) and the remainder of H3. The rest of the histone core is colored gray. The C-terminal tail of CENP-A, at the end of $\alpha 3$, is disordered in the nucleosomal context, in the "on" binding state, interacting with CENP-C (red).

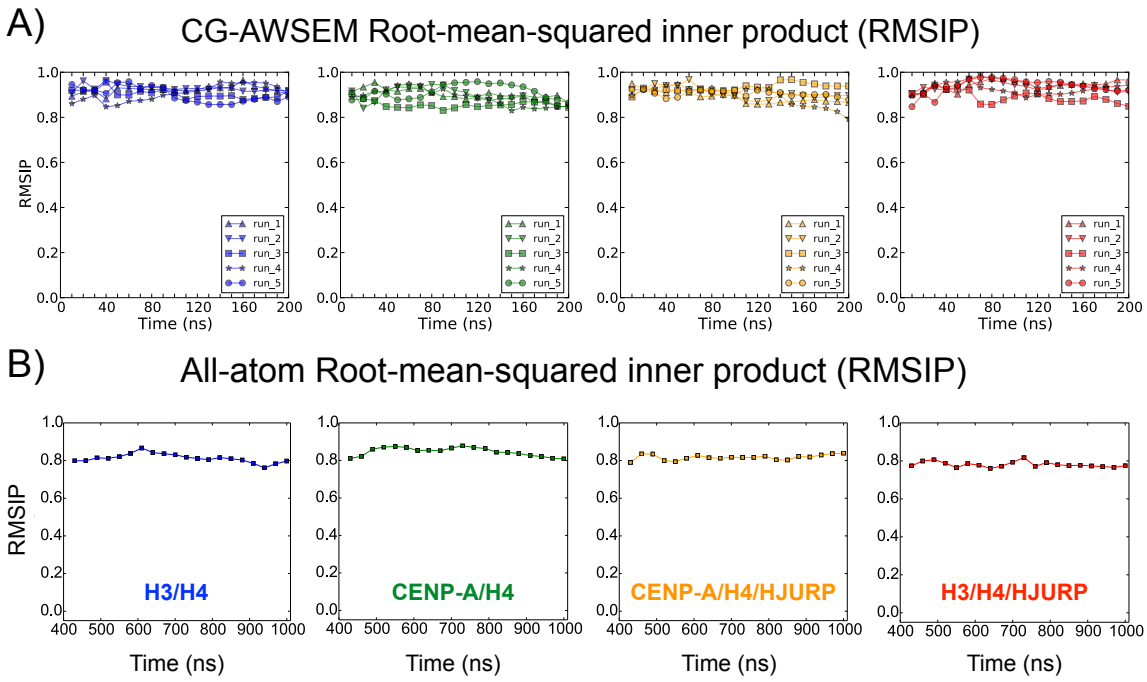

Figure A.17: **RMSIP demonstrates the convergence of CG and all-atom simulations.** To extend our evaluation of convergence, we calculated the root-mean-squared inner product (RMSIP). RMSIP is a sum of every dot product between the first ten eigenvectors of the first half of the trajectory and first ten eigenvectors of the second half. It is a normalized measure, where 1 indicates identical eigenvectors, and the simulation reaches convergence when RMSIP is close to 1. (A) CG simulations reached convergence (RMSIP > 0.8) after 10 ns, and, (B) convergence was achieved for the final 600 ns of all-atom MD simulation which are considered for analysis (RMSIP > 0.75).

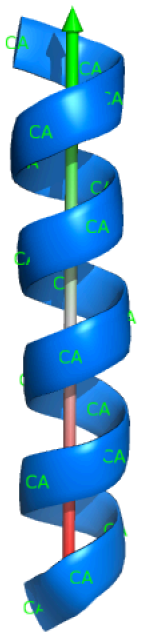

Figure A.18: **Orientation vector sketch for one α helix.**

$$V = \begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ . & . & . \\ . & . & . \\ x_i - x_0 & y_i - y_0 & z_i - z_0 \\ . & . & . \end{bmatrix},$$

where (x_i, y_i, z_i) represents the position of the ith Cα, and (x_0, y_0, z_0) is the coordinates of the geometric center of the selected helix. Then we use singular value decomposition (SVD) to determine all the eigenvalues of matrix V. The eigenvector corresponding to the biggest eigenvalue provides the orientation vector. A diagrammatic sketch is shown in Figure A.18.

Appendix B: Supplementary Information for Chapter 4

B.1 Simulation Method Details

The Associative-memory, Water-mediated, Structure and Energy Model (AWSEM) has been successfully applied to study protein folding [44], binding [**?**, **?**], aggregation [**?**,53,54], membrane proteins [55,56], and protein-DNA association [57,59,146]. Here, we used the AWSEM model as the force field to perform molecular dynamics simulations. In AWSEM, both physically-motivated potential terms, such as water-mediated potential, the hydrogen bonding potential, and a bioinformatically-based local structure biasing term are included. Three beads represent one amino acid and a water-mediated potential describes the water-protein interactions. Written in Eq. S1, the AWSEM Hamiltonian includes a backbone term $V_{backbone}$, a contact term $V_{contact}$, a many body burial term V_{burial}, a hydrogen-bonding term V_{HB}, and the bioinformatical term, called the fragment-memory or associative-memory potential V_{AM}. The protein-like backbone is maintained by the term $V_{backbone}$, a combination of harmonic potentials based on the positions of C_α, C_β (H for Glycine) and O atoms. $V_{contact}$ and V_{burial} deal with the water-mediated or protein-mediated residual interactions. The V_{HB} term defines hydrogen-bonding networks that are responsible for the formation of α helices or β hairpins. The bioinformatic term, called the

fragment-memory or associative-memory potential V_{AM} is a Go-like potential, but uses fragments of the target sequence to bias the local structure formation.

$$V_{AWSEM} = V_{backbone} + V_{contact} + V_{burial} + V_{HB} + V_{AM} \tag{B.1}$$

In our simulations, single memories were set, exclusively, found in the corresponding nucleosomal crystal structures. Weak biases were applied between the centers-of-mass of two monomers in a dimer and between two dimers in a tetramer to control the overall concentration of monomers (k = 0.02 kcal/(mol Å^2)). AWSEM simulations were performed in the large-scale atomic/molecular massively parallel simulator (LAMMPS), using the Nosé-Hoover thermostat with a timestep of 5 fs.

For free energy calculations, in general, the coupled replica-exchange and umbrella-sampling method was applied to enhance the sampling conformational space explored. First, harmonic potentials centered at different R_0 (*i.e.* the distance between two histone dimers) were used as the umbrella biasing potential (Eq. S2), with a biasing spring constant k_R of 5 kcal/(mol Å^2)). At the same time, ten temperature replicas, from 280 K to 370 K, were run in parallel for each umbrella window. Exchanges between replicas were attempted every 400 steps. Lastly, sampling data at 300 K from different windows were collected and the free energy profile, namely here the potential of mean force over coordinate R, was then calculated using the weighted histogram analysis method (WHAM) [4]. A relevant Jacobian factor correction term, $k_B T \ln[4\pi R^2]$, was subtracted from the free energy calculation since it makes unphysical contributions to the configurational partition function [140].

$$U_{umbrella} = \frac{1}{2}k_R(R - R_0)^2 \tag{B.2}$$

To calculate the binding free energy of dimer-dimer interaction, thirty umbrella windows are set along the distance between two histone dimers, ranging from 20 Å to 50 Å (see SI section S5). Each replica was run for 2 million steps. The first 0.5 million steps were not included for analysis to allow the systems to reach equilibration.

Separately, ten independent constant temperature simulations were carried out for tetramers of $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$, for 30 million timesteps each, 300 million steps in total (1500 ns in the coarse-grained timescale). 10 million timesteps were run for each octamer simulation of H3 and CENP-A, with a totaling 100 million timesteps for each octamer system (500 ns in the coarse-grained timescale). Simulations were performed in a 200-Å-long cubic box with periodic boundary conditions. Trajectories were combined for later data analysis after removing the first 10 ns in every run to account for thermal equilibration. The convergences of simulations were verified by the root-mean-squared inner-product (RMSIP) analysis (see section S2 for details).

B.2 Trajectory Analysis Details

All the trajectory analyses in this work, including the calculations of root-mean-square deviations (RMSD), radius-of-gyration (R_g), distances (R), dihedral angles θ, Q values, and contact analysis, were based on the Cα coordinates. Q

values were calculated separately for the tetramer interface between two dimers, the whole dimer, the dimer interface between two monomers, and also for the monomers. The dihedral angle calculations in Figure 4 were obtained by measuring the dihedral angle formed by the first and last Cα atoms of the α2 helices. A contact in Figure 5 was considered to exist when the distance between two Cα atoms was shorter than 8 Å.

The angle between two α helices is calculated by the orientation vector for each selected helix, based on the coordinates of C_α atoms. We then built the variance matrix V_{helix}, composed of all the C_α coordinates and coordinates of the geometric center of the helix. The singular value decomposition (SVD) [**?**] was applied to determine the eigenvalues of the matrix. The eigenvector corresponding to the biggest eigenvalue provided the orientation vector. The variance matrix V_{helix} was defined as:

$$V = \begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ . & . & . \\ . & . & . \\ x_i - x_0 & y_i - y_0 & z_i - z_0 \\ . & . & . \end{bmatrix},$$

where (x_i, y_i, z_i) represents the position of the ith Cα, and (x_0, y_0, z_0) is the coordinates of the geometric center of the selected helix.

B.3 Error Analysis and P-value

Error analysis for the free energy profiles (FEPs) consisted of two parts (Eq. S4). The first part is more related with the convergence of the simulation data. We determined this part of the statistical errors by calculating the FEP variances from independent simulation interval blocks. For example, in Figure 2, we divided the entire simulation trajectory into 3 non-overlapping blocks along the time series, and calculated the free energy for each block independently. The standard deviation of the free energy for each reaction coordinate window determined from the three blocks were taken to be the statistical error from the ensemble. In Eq. S3, N is 3 and F_i is the FEP of the ith internal block. F_0 is the FEP of the whole simulation data. The second error part is concerned more with the stochasticity of the data. We estimated this part using the Monte Carlo bootstrap error analysis in WHAM [**?**,4]. The basic idea is, for each simulation window, to use the computed cumulant of the histogram of the real data to randomly generate a new histogram, with the same number of points and then perform WHAM iterations on the set of newly-generated histograms until it is converged, storing the average normalized probability and free energy for each bin in the histogram. The statistical uncertainty is then obtained accordingly.

$$Error(FEP) = \sqrt{\frac{\sum_{i=1}^{N}(F_i - F_0)^2}{N-1}} + Error_{WHAM} \tag{B.3}$$

We used the p-value from a t-test [**?**] to verify whether the differences of our

samples were statistically significant. T-statistic for mean is given by ($\mid A_1 - A_2 \mid$) / $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, where A_1, A_2 are the mean values of the distributions and s_1, s_2 are the standard deviations and n_1, n_2 the number of the data in each distribution. The same method was used in for probability density function figures of the main text and of the SI.

B.4 Simulation Convergence

We calculated the root-mean-squared inner-product (RMSIP) [**?**] to verify the convergence of all performed simulations. RMSIP, as defined in the below equation, quantifies the overlap between essential subspaces through the inner product of the first ten principal eigenvectors of Cα atom coordinates. It is a normalized parameter, where an RMSIP closer to 1.0 indicates greater overlap between data sets.

$$\text{RMSIP} = \Big(\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}(\boldsymbol{\eta_i}\cdot\boldsymbol{\nu_j})^2\Big)^{1/2} \tag{B.4}$$

In our simulations, the RMSIP for every individual run was computed between the data sets corresponding to two halves of increasingly higher percentages of the entire simulation trajectory, starting with the first 10 ns, then the first 20 ns, and so forth. All the RMSIP values are over 0.7, indicating adequate convergence of corresponding simulations.

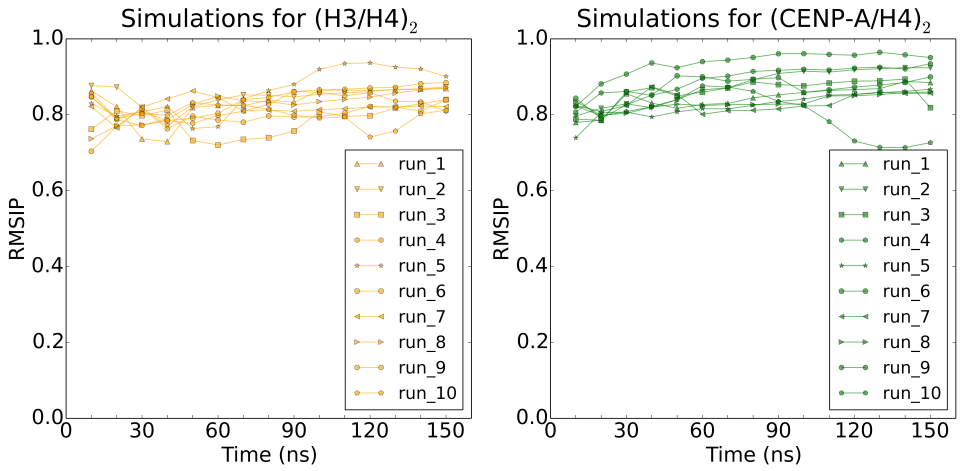

Figure B.1: **RMSIP analysis shows the convergence of CG-AWSEM simulations.** RMSIP are calculated for every independent simulation of both H3/H4 and CENP-A/H4 tetramer. All calculated RMSIP are over 0.7, indicating adequate convergences.

B.5 Umbrella Sampling Histograms

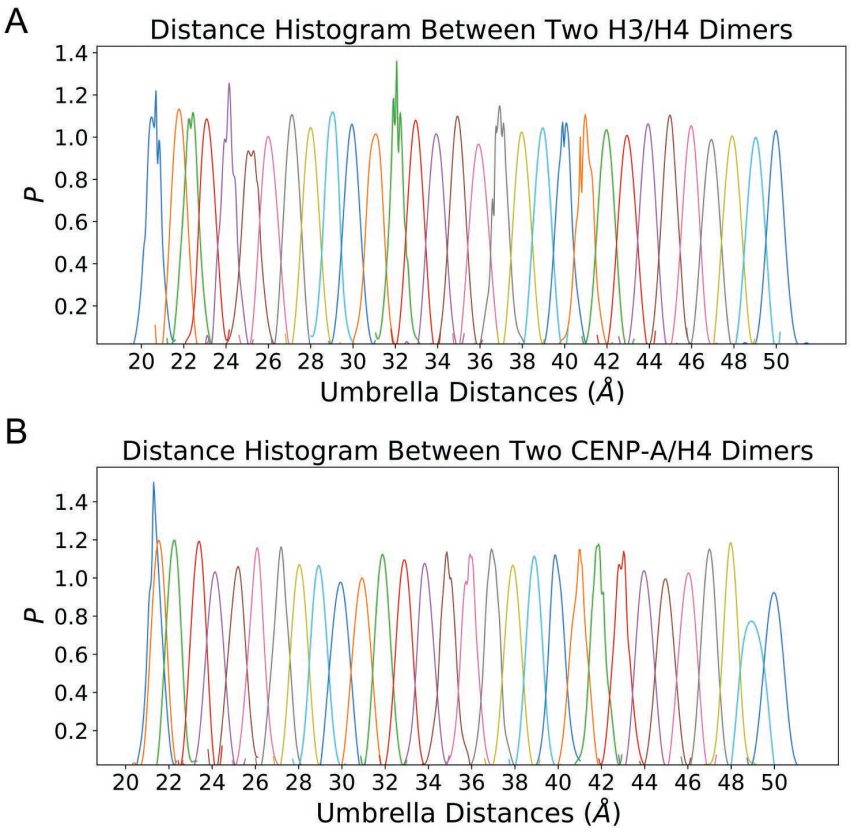

Figure B.2: **Sufficient overlaps of reaction coordinate R between adjacent umbrella windows ensure the convergence of WHAM [4] calculation.** Distances from all umbrella windows at replica 300K are collected and histogramed for (A) two H3/H4 dimers and (B) two CENP-A/H4 dimers. PMFs were then calculated from these data using WHAM.

H3:

VALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLCAIHAKRVTIMPKDIQLARRIRGER

H4:

NIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

H3':

LREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

H4':

DNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

CENP-A:

GWLKEIRKLQKSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRG

H4:

NIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

CENP-A':

LKEIRKLQKSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRGL

H4':

DNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

Figure B.3: **Amino acid sequence alignment.** The amino acid sequences of H3, H4, and CENP-A, provide the primary level of description for the protein structures of the $(\text{CENP-A/H4})_2$ and $(\text{H3/H4})_2$ tetramers investigated in this study. Sequences of the four-helix bundle region and of the $\alpha 2$ helix which are particularly emphasized in this work are marked in red and underlined.

B.6 Amino Acid Sequences

Histone tails are excluded in the simulations due to the reason that they are mainly stabilized by the DNA.

B.7 Extended 2D and 1D Free Energy Profiles from Enhanced Sampling Method

Through calculating the unbiased probability distribution and re-histogramming over different collective variables, we projected the calculated free energy profile onto different coordinates, either two-dimensional or one-dimensional. All the results consistently demonstrate that the H3 histone tetramer occupies a more rugged free energy landscape while CENP-A has a well-funneled landscape topology, indicating that the CENP-A tetramer favors a stable thermodynamic state while H3 does not. The one-dimensional free energy profiles from the coupled replica-exchange and umbrella sampling method can be qualitatively compared to the probability distribution of the same coordinate (Figure 3, Figure S7) from the later long-timescale constant temperature simulations, after using the Boltzmann relation. The consistency between the two results proves the efficiency and convergence of both methods.

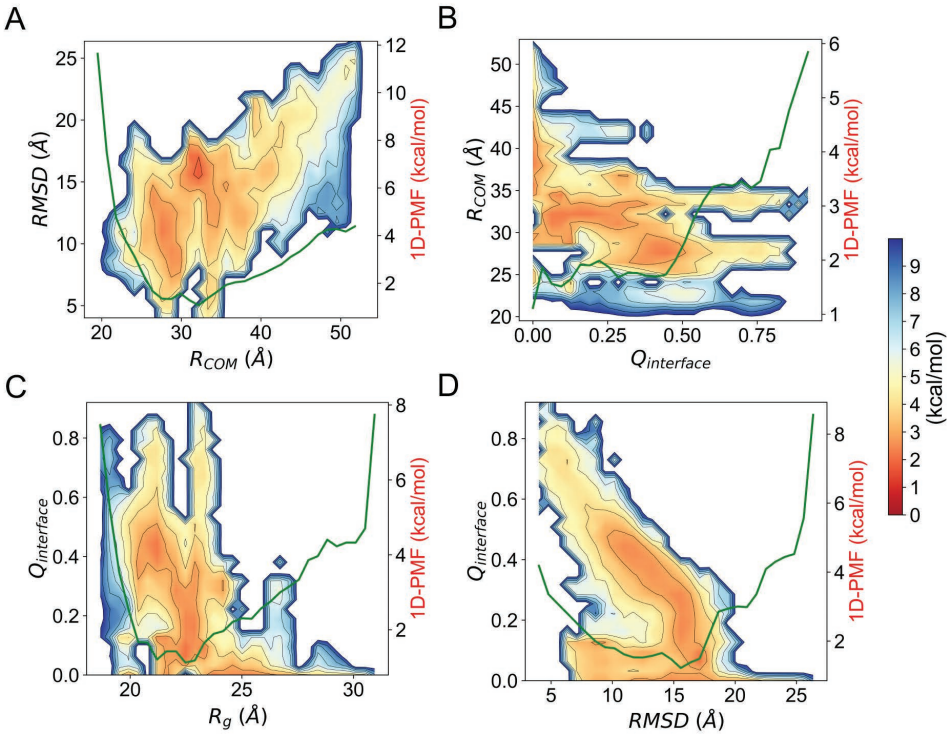

Figure B.4: **Extended 2D and 1D free energy profiles for histone $(H3/H4)_2$.** Free energy profiles calculated from the enhanced sampling for $(H3/H4)_2$ tetramer are projected on the 2D reaction coordinates of the distance between centers of masses of each dimer R_{COM} and the measure of overall structural fluctuation root-mean-square-deviation RMSD (A), the tetramer interface contact quantification parameter $Q_{interface}$ value and R_{COM} (B), the geometry measurement radius of gyration R_g and $Q_{interface}$ (C), the RMSD and $Q_{interface}$ (D). 1D free energy projection on the dimension (marked as green line) of R_{COM}, $Q_{interface}$, R_g, and RMSD are shown on the right side of each panel of (A, B, C, D) accordingly.

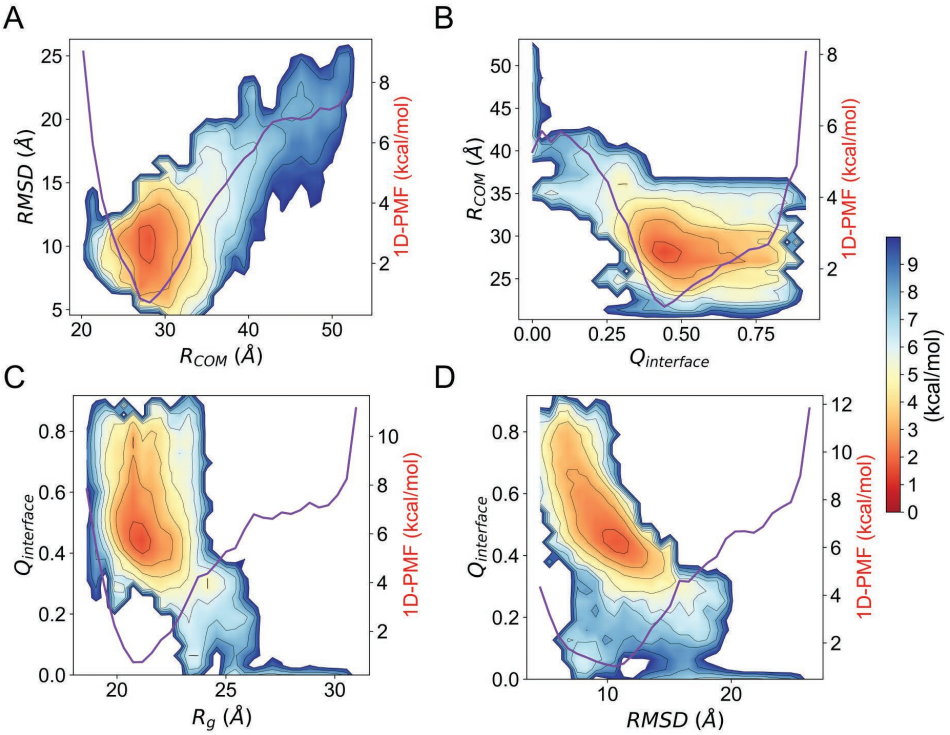

Figure B.5: **Extended 2D and 1D free energy profiles for (CENP-A/H4)$_2$ histone tetramer.** Free energy profiles calculated from the enhanced sampling for CENP-A tetramer are projected on the 2D reaction coordinates of the distance between centers of masses of each dimer R_{COM} and the measure of overall structural fluctuation root-mean-square-deviation RMSD (A), the tetramer interface contact quantification parameter $Q_{interface}$ value and R_{COM} (B), the geometry measurement radius-of-gyration R_g and $Q_{interface}$ (C), the RMSD and $Q_{interface}$ (D). 1D free energy projection on the dimension (marked as purple line) of R_{COM}, $Q_{interface}$, R_g, and RMSD are shown on the right side of each panel of (A, B, C, D) accordingly.

B.8 Extended Distributions of Structural Measures and Measurement with Time in Constant T Simulations

For the long-timescale constant temperature simulations, we also calculated the probability distribution for different structural measures, including the root-mean-square deviation (RMSD), the distance between two internal dimers R_{COM}, and the interface contact resemblance $Q_{interface}$. Locations of the distribution peaks observed from constant T simulations agree with the minima locations in free energy profiles from the enhanced sampling (Figures 1/2, Figures S4/S5), demonstrating the convergence of both methods.

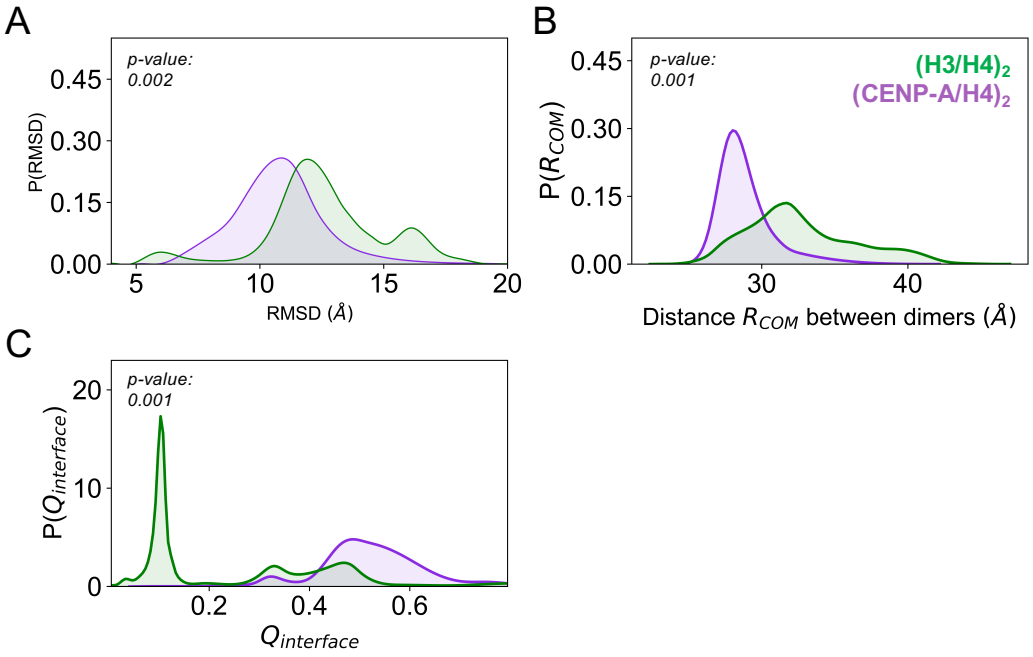

Figure B.6: **Distributions of different structural measures confirm the conformational heterogeneity of the H3 tetramer (green) and the homogeneity of the CENP-A tetramer (purple).** (A) The RMSD distribution features multiple populations for the H3 tetramer and a single population for the CENP-A tetramer. (B) The distance between dimers R_{COM} is shorter on average for (CENP-A/H4)$_2$, with a much narrower distribution, than that of (H3/H4)$_2$. (C) $Q_{interface}$ distributions indicate that the interface of (CENP-A/H4)$_2$ remains more stable and closer to its native state than (H3/H4)$_2$. Locations of the peaks in these panels agree with the minima locations in free energy profiles calculated from the enhanced sampling simulations.

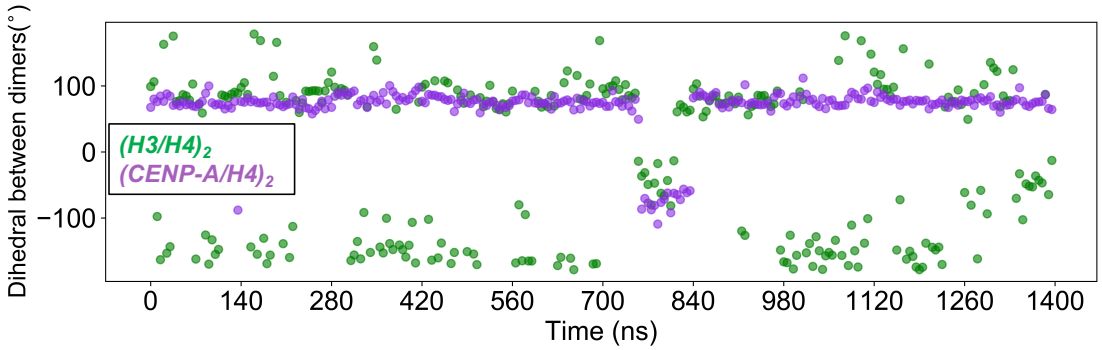

Figure B.7: **The dihedral angle between $\alpha 2$ helices measured as a function of time emphasizes the rotational dynamics of the H3 tetramer.** The tetramer dihedral angle of H3 (green) frequently transitions between 90°, −150°, and −50°, while the dihedral of CENP-A (purple) remains constant throughout most of simulation, with only one dihedral angle transition observed.

B.9 Principle Component Analysis

To extract the dominant modes of motion from the long-timescale constant temperature MD simulations, we performed principal component analysis (PCA). Overall translational and rotational motion of the MD trajectories were eliminated by a translation to the average geometric center and by alignment to the energy-minimized structure. Then the simulation trajectories are projected onto the first two principal components to illustrate the corresponding free energies. The result (Figure S6) is qualitatively consistent with the free energy profile computed from the enhanced the sampling (Figure 1, Figures S4/S5).

The PCA method is described in detail below. Using the Cartesian coordinates of all n Cα atoms over N simulation snapshots (t_i represents an individual time), we created a trajectory position matrix $\mathbf{Q}$,

$$q_i = (x_1, y_1, z_1, ..., x_n, y_n, z_n), \mathbf{Q} = (q(t_1), q(t_2), ..., q(t_N)) \tag{B.5}$$

$$Q_{ij} = q_i(t_j) \tag{B.6}$$

From this trajectory matrix $\mathbf{Q}$, we constructed a covariance matrix $\mathbf{C}$. Let N be the number of snapshots, n the number of Cαs, and $\mathbf{Q}$ $(3n \times N)$ be the trajectory position matrix. Hence, we have the covariance matrix $\mathbf{C}$ $(3n \times 3n)$ defined in Eq. S7. We then diagonalize the covariance matrix $\mathbf{C}$,

$$C_{j,k} = \frac{1}{N-1} \sum_{i=1}^{N} (Q_{ji} - \langle Q_j \rangle)(Q_{ki} - \langle Q_k \rangle). \tag{B.7}$$

$$\mathbf{CM} = \mathbf{M}\Lambda \tag{B.8}$$

m_i, the columns of $\mathbf{M}$, are orthonormal eigenvectors representing the principal components, and the diagonal values along Λ_{ii} are the eigenvalues associated with each principal component. We arranged the eigenvalues from highest to lowest, meaning the first principal component captures the most variance within our dataset, the second principal component captures the second most variance, and so forth. Next, we projected the trajectory matrix $\mathbf{Q}$ onto the first 2 principal components, the eigenvectors corresponding to the 2 highest eigenvalues:

$$\nu_1 = \mathbf{Q} \cdot m_1, \qquad \nu_2 = \mathbf{Q} \cdot m_2. \tag{B.9}$$

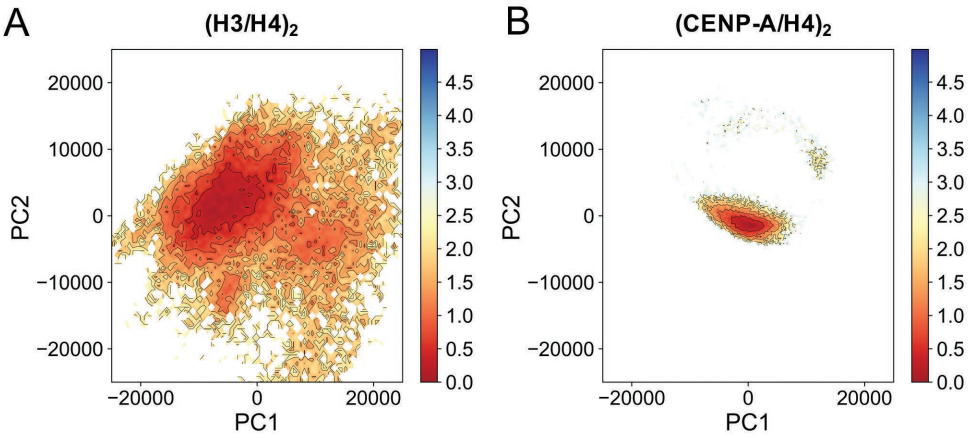

Figure B.8: **Free energy projections along the first two PCA components show that $(H3/H4)_2$ has a more rugged free energy landscape than $(CENP\text{-}A/H4)_2$.** (A) Free energy projection of $(H3/H4)_2$ reveals a broad landscape with multiple conformations basins. (B) Free energy projection of $(CENP\text{-}A/H4)_2$ has only one single and deep basin.

We then separated the (ν_1, ν_2) space covered equally into a grid and obtain joint probabilities, $P(\nu_1, \nu_2)$ for each box within the grid. Finally, the free energy landscape was projected along the first two principal components:

$$F(\nu_1, \nu_2) = -k_B T \ln P(\nu_1, \nu_2) - F_{\min}. \tag{B.10}$$

B.10 Analyses on the Level of Dimers and Monomers

To compare the results with our previously-published independent dimer study [**?**], we performed similar analysis of the tetramer simulations, including RMSD, Q of dimer, $Q_{interface}$ of dimer, and Q for histone monomers. The results here show that, even in tetramer, the CENP-A dimer is still more heterogeneous than H3 dimer, and the H4 monomer is more native-like than its binding partner H3 or CENP-A.

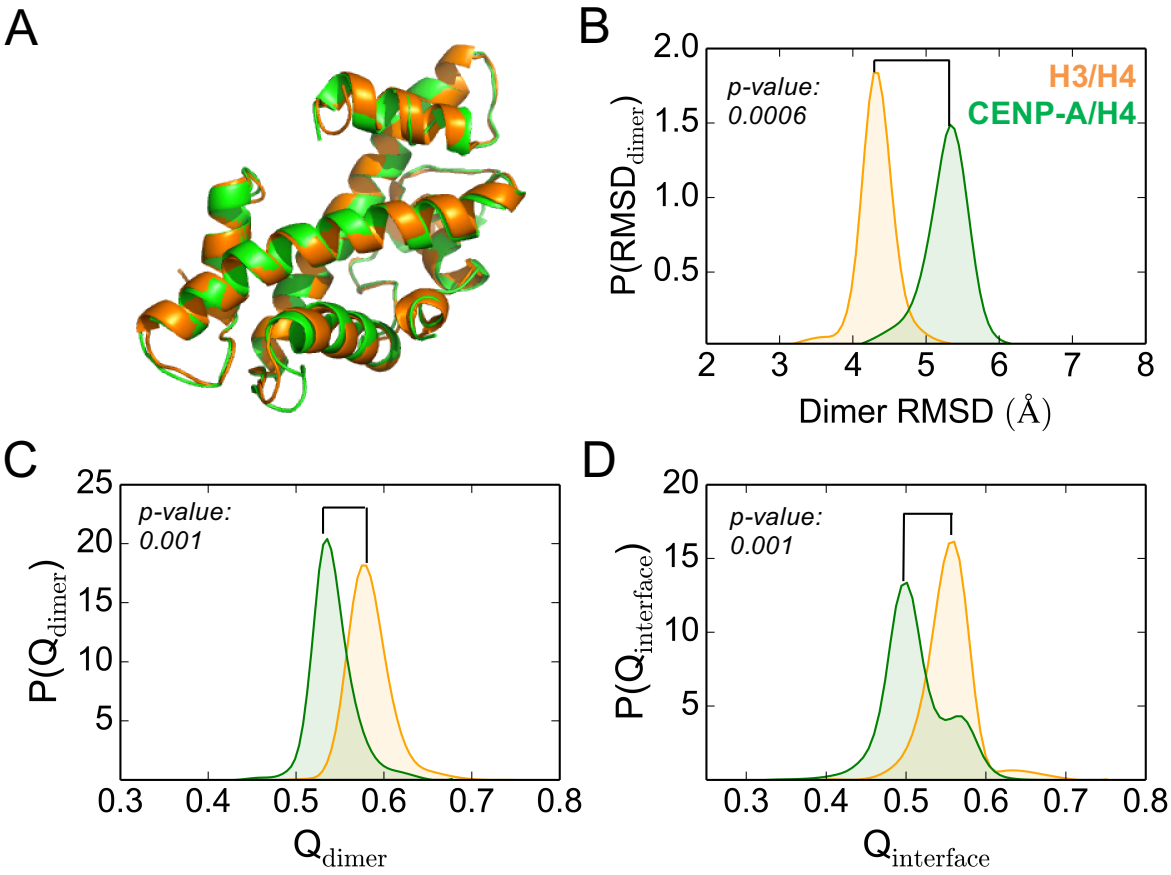

Figure B.9: **The H3/H4 dimer is structurally more stable than the CENP-A/H4 dimer in tetramer simulations.** Q_{dimer} and $Q_{interface}$ of the dimer, characterize a dimer's overall structural resemblance or the resemblance of the monomeric interface to its native state (A) respectively. Analyses on the dimer level demonstrate that, in tetramer simulations, CENP-A/H4 exhibits a larger root-mean-square deviation (RMSD) and lower Q_{dimer} (C) and $Q_{interface}$ (D), on average, than H3/H4. This implies the high variability or elasticity of CENP-A in general, which agrees with previous experimental [5, 6] and computational studies [**?**, 7].

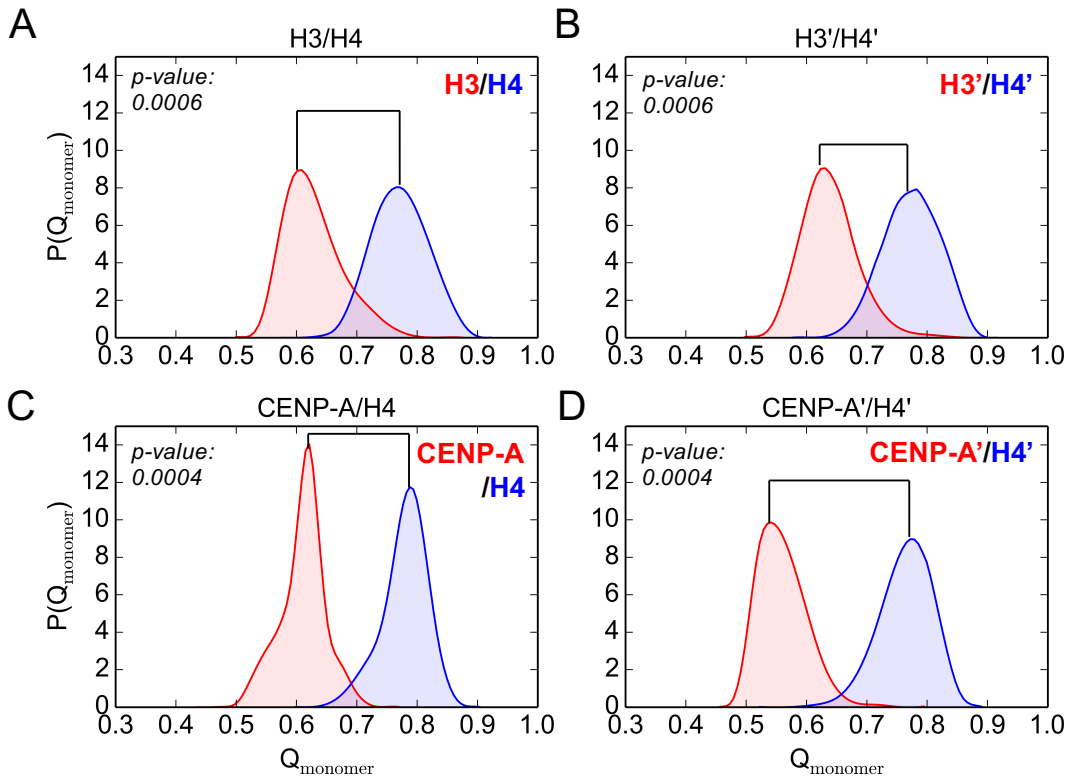

Figure B.10: **The H4 monomer maintains a more native-like conformation than its binding partner, either H3 or CENP-A, in all tetramer AWSEM simulations.** $Q_{monomer}$ describes a histone monomer's overall structural similarity with respect to the crystal structures of the corresponding H3 nucleosome (PDB ID: 1KX5) and CENP-A nucleosome (PDB ID: 3AN2). $Q_{monomer}$ was calculated for individual histone proteins of both the first and second H3/H4 dimer (A, B), and for both the first and second CENP-A/H4 dimer (C, D). It shows that H4 has a higher $Q_{monomer}$ than H3, or CENP-A, meaning that H4 maintains a more stable and native-like structure. This result is in accordance with the previous histone dimer study [**?**].

B.11 AWSEM Energy Analysis of the Tetramer Interface

For a better understanding of the histone tetramer interface dynamics at residual level, we analyzed the energy terms in AWSEM for the corresponding interfaces. The sum of $V_{contact}$ and V_{bural} terms (details in section S1), E_{pair}, is collected for every pair of residue interactions between the two dimers inside a tetramer. The sorted large pair interactions are shown in table S1 and S2. The cutoff was chosen as 0.65 kcal/mol for the absolute value of E_{pair}.

The shown two tables (table S1, S2) are the AWSEM energy outputs for residual pair interactions around the tetramer interface. The first table provides the energies for representative conformations from AWSEM simulations. The included energies in the second table are the AWSEM outputs for the snapshot of each system that is closest to the initial conformation. Both of these two tables demonstrate that CENP-A:CENP-A' have more interacting residue pairs at the binding interface. Detailed residue positions are shown in the structure figures (Figure S11 for H3 and S12 for CENP-A).

Table B.1: **Key residue interactions of the tetramer interface in AWSEM (representative conformations)**

H3	H3'	E_{pair} (kcal/mol)	CENP-A	CENP-A'	E_{pair} (kcal/mol)
A111	A111	-0.9720	Y110	L111	-1.0000
N108	A111	-0.8886	L111	T113	-1.0000
R116	A111	-0.6897	L111	R131	-1.0000
A111	R116	-0.6885	T113	R131	-1.0000
			Q127	L111	-1.0000
			Y110	Q127	-0.9998
			T113	Q127	-0.9997
			T113	L111	-0.9784
			Q127	Y110	-0.9773
			R131	Y110	-0.9759
			Y110	R131	-0.9755
			R131	T113	-0.9072
			L111	Q127	-0.9071
			L111	L111	-0.906
			L111	Y110	-0.897
			R131	A109	-0.7634
			Y110	Y110	-0.7006
			H115	H115	-0.6895

Table B.2: **Key residual interactions of the tetramer interface in AWSEM (initial conformations)**

H3	H3'	E_{pair} (kcal/mol)	CENP-A	CENP-A'	E_{pair} (kcal/mol)
Q125	R129	-1.0000	Y110	R131	-1.0000
R129	Q125	-1.0000	L111	R131	-1.0000
L126	R129	-0.9999	Q127	L111	-1.0000
R129	L126	-0.9999	Q127	R131	-1.0000
Q125	Q125	-0.9918	R131	L111	-0.9999
N108	R129	-0.9832	R131	Q127	-0.9994
Q125	A111	-0.9549	L111	L111	-0.9991
A111	Q125	-0.9263	L111	Q127	-0.9989
N108	Q125	-0.8416	L128	R131	-0.9878
H113	H133	-0.7130	Q127	Y110	-0.9747
			R131	R131	-0.9349
			Y110	Q127	-0.8929
			L111	Y110	-0.8346
			R131	Y110	-0.6405

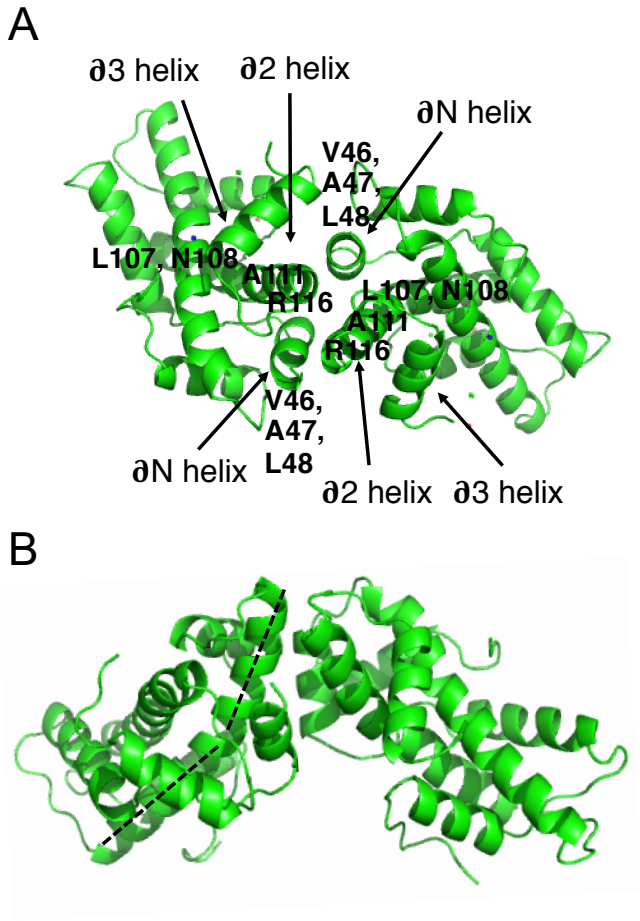

Figure B.11: **Representative structure of H3 tetramer with interface interaction details.** (A) Top view of the structure highlights the $(H3/H4)_2$ interface is a disrupted four-helix bundle region. αN helix competes with the $\alpha 3$ helix, to interact with $\alpha 2$ helix, forming hydrophobic interactions between V46, A47, L48 of αN and A111, L107 of $\alpha 2$. (B) Side view of the representative H3 tetramer shows that the $\alpha 2$ helix in H3 can be curved, which is illustrated by the dash line.

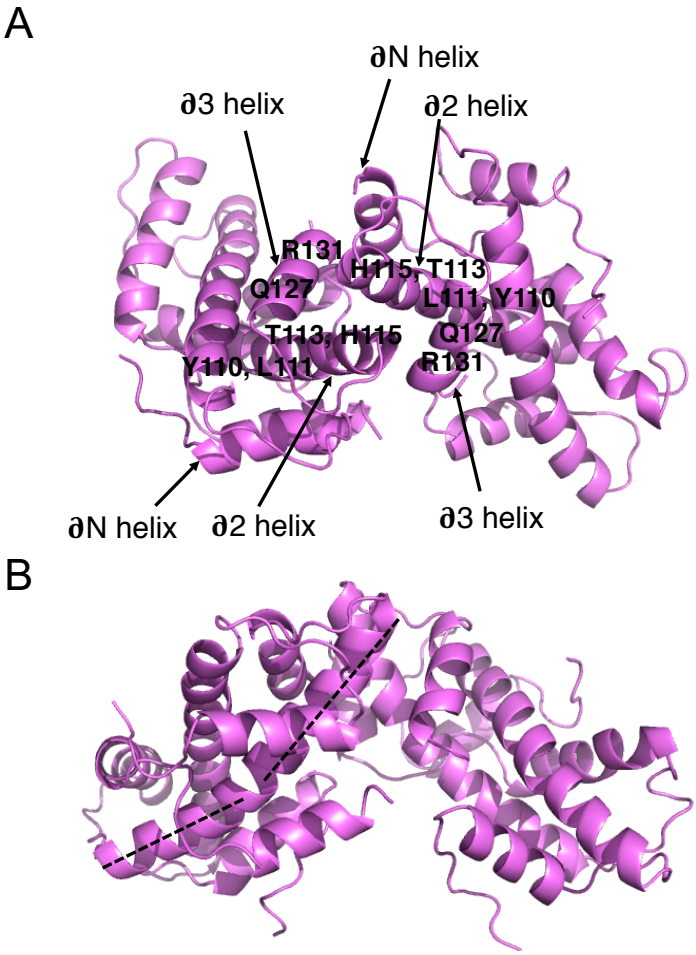

Figure B.12: **Representative structure of CENP-A tetramer with interface interaction details.** (A) Top view of the interface highlights a well-formed four-helix bundle region, composed of $\alpha 2$ and $\alpha 3$ helices. (B) Side view of the representative CENP-A tetramer shows that the $\alpha 2$ helix in CENP-A is curved, illustrated by the dash line.

Bibliography

[1] Ruxandra I Dima and D Thirumalai. Asymmetry in the shapes of folded and denatured states of proteins. *The Journal of Physical Chemistry B*, 108(21):6564–6570, 2004.

[2] Karolin Luger, Armin W Mader, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome resolution core particle at 2 . 8 A. *Nature*, 389(6648):251–260, 1997.

[3] 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.

[4] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

[5] 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.

[6] 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.

[7] David Winogradoff, Haiqing Zhao, Yamini Dalal, and Garegin A Papoian. Shearing of the cenp-a dimerization interface mediates plasticity in the octameric centromeric nucleosome. *Scientific Reports*, 5:17038, 2015.

[8] Craig Venter and Daniel Cohen. The century of biology. *New Perspectives Quarterly*, 21(4):73–77, 2004.

[9] Francis S Collins, Michael Morgan, and Aristides Patrinos. The human genome project: lessons from large-scale biology. *Science*, 300(5617):286–290, 2003.

[10] Chen Chen, Hong Hwa Lim, Jian Shi, Sachiko Tamura, Kazuhiro Maeshima, Uttam Surana, and Lu Gan. Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Molecular biology of the cell*, 27(21):3357–3368, 2016.

[11] Mikhail Eltsov, Kirsty M MacLellan, Kazuhiro Maeshima, Achilleas S Frangakis, and Jacques Dubochet. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences*, 105(50):19732–19737, 2008.

[12] Lu Gan, Mark S Ladinsky, and Grant J Jensen. Chromatin in a marine picoeukaryote is a disordered assemblage of nucleosomes. *Chromosoma*, 122(5):377–386, 2013.

[13] Feng Dong and KE Van Holde. Nucleosome positioning is determined by the (h3-h4) 2 tetramer. *Proceedings of the National Academy of Sciences*, 88(23):10596–10600, 1991.

[14] 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.

[15] Burcu Biterge and Robert Schneider. Histone variants: key players of chromatin. *Cell and tissue research*, 356(3):457–466, 2014.

[16] David T Brown. Histone variants: are they functionally heterogeneous? *Genome biology*, 2(7):reviews0006–1, 2001.

[17] Harmit S Malik and Steven Henikoff. Phylogenomics of the nucleosome. *Nature Structural & Molecular Biology*, 10(11):882–891, 2003.

[18] Harmit S Malik and Steven Henikoff. Phylogenomics of the nucleosome. *Nature Structural & Molecular Biology*, 10(11):882–891, 2003.

[19] Haiqing Zhao, David Winogradoff, Minh Bui, Yamini Dalal, and Garegin A Papoian. Promiscuous histone mis-assembly is actively prevented by chaperones. *Journal of the American Chemical Society*, 138(40):13207–13218, 2016.

[20] K Yoda, S Ando, S Morishita, K Houmura, K Hashimoto, K Takeyasu, and T 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.

[21] 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.

[22] Minh Bui, Marcin P Walkiewicz, Emilios K Dimitriadis, and Yamini Dalal. The CENP-A nucleosome: a battle between Dr Jekyll and Mr Hyde. *Nucleus*, 4(1):37–42, 2014.

[23] 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.

[24] Rajbir K Athwal, Marcin P Walkiewicz, Songjoon Baek, Song Fu, Minh Bui, Jordi Camps, Thomas Ried, Myong-Hee Sung, and Yamini Dalal. Cenp-a nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics & chromatin*, 8(1):2, 2015.

[25] 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.

[26] 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.

[27] Hiroaki Tachiwana and Hitoshi Kurumizaka. Structure of the CENP-A nucleosome and its implications for centromeric chromatin architecture. *Genes Genet. Syst.*, 86(6):357–364, 2011.

[28] Hitoshi Kurumizaka, Naoki Horikoshi, Hiroaki Tachiwana, and Wataru Kagawa. Current progress on structural studies of nucleosomes containing histone H3 variants. *Curr. Opin. Struct. Biol.*, 23(1):109–115, February 2013.

[29] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin A Papoian. Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The Journal of Physical Chemistry B*, 116(29):8494–8503, 2012.

[30] Cyrus Levinthal. How to fold graciously. *Mossbauer spectroscopy in biological systems*, 67:22–24, 1969.

[31] Martin Karplus. The levinthal paradox: yesterday and today. *Folding and design*, 2:S69–S75, 1997.

[32] Joseph D Bryngelson, Jose Nelson Onuchic, Nicholas D Socci, and Peter G Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195, 1995.

[33] Peter G Wolynes. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proceedings of the National Academy of Sciences*, 94(12):6170–6175, 1997.

[34] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G Wolynes. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry*, 48(1):545–600, 1997.

[35] Chung-Jung Tsai, Sandeep Kumar, Buyong Ma, and Ruth Nussinov. Folding funnels, binding funnels, and protein function. *Protein Science*, 8(6):1181–1190, 1999.

[36] Buyong Ma, Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Folding funnels and binding mechanisms. *Protein engineering*, 12(9):713–720, 1999.

[37] Sandeep Kumar, Buyong Ma, Chung-Jung Tsai, Neeti Sinha, and Ruth Nussinov. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Science*, 9(1):10–19, 2000.

[38] Yaakov Levy, Peter G Wolynes, and José N Onuchic. Protein topology determines binding mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(2):511–516, 2004.

[39] Yaakov Levy, Garegin A Papoian, José N Onuchic, and Peter G Wolynes. Energy landscape analysis of protein dimers. *Israel journal of chemistry*, 44(1-3):281–297, 2004.

[40] Yaakov Levy, Samuel S Cho, José N Onuchic, and Peter G Wolynes. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *Journal of molecular biology*, 346(4):1121–1145, 2005.

[41] IE Fermi, P Pasta, S Ulam, and M Tsingou. Studies of the nonlinear problems. Technical report, Los Alamos Scientific Lab., N. Mex., 1955.

[42] Berni J Alder and T E Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[43] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585, 1977.

[44] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin A Papoian. Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B*, 116(29):8494–8503, 2012.

[45] Weihua Zheng, Nicholas P Schafer, Aram Davtyan, Garegin A Papoian, and Peter G Wolynes. Predictive energy landscapes for protein–protein association. *Proceedings of the National Academy of Sciences*, 109(47):19244–19249, 2012.

[46] Garegin A Papoian, Johan Ulander, Michael P Eastwood, Zaida Luthey-Schulten, and Peter G Wolynes. Water in protein structure prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10):3352–3357, 2004.

[47] Yuzo Ueda, Hiroshi Taketomi, and Nobuhiro Gō. Studies on protein folding, unfolding, and fluctuations by computer simulation. ii. a. three-dimensional lattice model of lysozyme. *Biopolymers*, 17(6):1531–1548, 1978.

[48] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[49] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[50] Steven Henikoff, Takehito Furuyama, and Kami Ahmad. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics*, 20(7):320–326, 2004.

[51] Yamini Dalal, Takehito Furuyama, Danielle Vermaak, and Steven Henikoff. Structure, dynamics, and evolution of centromeric nucleosomes. *Proceedings of the National Academy of Sciences*, 104(41):15974–15981, 2007.

[52] Vassiliki Karantza, Andreas D Baxevanis, Ernesto Freire, and Evangelos N Moudrianakis. Thermodynamic studies of the core histones: ionic strength and ph dependence of h2a-h2b dimer stability. *Biochemistry*, 34(17):5988–5996, 1995.

[53] Weihua Zheng, Nicholas P Schafer, and Peter G Wolynes. Free energy landscapes for initiation and branching of protein aggregation. *Proc. Natl. Acad. Sci. U.S.A.*, 110(51):20515–20, Dec 2013.

[54] Mingchen Chen and Peter G Wolynes. Aggregation landscapes of huntingtin exon 1 protein fragments and the critical repeat length for the onset of huntingtons disease. *Proceedings of the National Academy of Sciences*, 114(17):4406–4411, 2017.

[55] Bobby L Kim, Nicholas P Schafer, and Peter G Wolynes. Predictive energy landscapes for folding -helical transmembrane proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 111(30):11031–6, Jul 2014.

[56] Ha H Truong, Bobby L Kim, Nicholas P Schafer, and Peter G Wolynes. Predictive energy landscapes for folding membrane protein assemblies. *J. Chem. Phys.*, 143(24):243101, Dec 2015.

[57] Davit A Potoyan, Weihua Zheng, Elizabeth A Komives, and Peter G Wolynes. Molecular stripping in the nf-κb/iκb/dna genetic regulatory network. *Proc. Natl. Acad. Sci. U.S.A.*, 113(1):110–115, 2016.

[58] Bin Zhang, Weihua Zheng, Garegin A Papoian, and Peter G Wolynes. Exploring the free energy landscape of nucleosome assembly. *Journal of the American Chemical Society*, 138(26):8126–8133, 2016.

[59] MinYeh Tsai, Bin Zhang, Weihua Zheng, and Peter G Wolynes. The molecular mechanism of facilitated dissociation of fis protein from dna. *J. Am. Chem. Soc.*, 138(41):13497–13500, 2016.

[60] David Fushman, Sean Cahill, and David Cowburn. The main-chain dynamics of the dynamin pleckstrin homology (ph) domain in solution: analysis of 15n relaxation with monomer/dimer equilibration 1. *Journal of molecular biology*, 266(1):173–194, 1997.

[61] Paul J Flory. *Principles of polymer chemistry*. Cornell University Press, 1953.

[62] Richard L Fahrner, Duilio Cascio, James A Lake, and Alexei Slesarev. An ancestral nuclear protein assembly: crystal structure of the methanopyrus kandleri histone. *Protein Science*, 10(10):2002–2007, 2001.

[63] Vikram Alva, Moritz Ammelburg, Johannes Söding, and Andrei N Lupas. On the origin of the histone fold. *BMC structural biology*, 7(1):17, 2007.

[64] Andreas D Baxevanis, Gina Arents, Evangelos N Moudrianakis, and David Landsman. A variety of dna-binding and multimeric proteins contain the histone fold motif. *Nucleic acids research*, 23(14):2685–2691, 1995.

[65] Gina Arents and Evangelos N Moudrianakis. The histone fold: a ubiquitous architectural motif utilized in dna compaction and protein dimerization. *Proceedings of the National Academy of Sciences*, 92(24):11170–11174, 1995.

[66] Kavitha Sarma and Danny Reinberg. Histone variants meet their match. *Nature reviews Molecular cell biology*, 6(2):139–149, 2005.

[67] Daniël P Melters, Jonathan Nye, Haiqing Zhao, and Yamini Dalal. Chromatin dynamics in vivo: a game of musical chairs. *Genes*, 6(3):751–776, 2015.

[68] Catherine Volle and Yamini Dalal. Histone variants: the tricksters of the chromatin world. *Current opinion in genetics & development*, 25:8–14, 2014.

[69] Xiakun Chu, Yong Wang, Linfeng Gan, Yawen Bai, Wei Han, Erkang Wang, and Jin Wang. Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone chz1 and histone h2a. z-h2b. *PLoS Comput. Biol.*, 8(7):e1002608, 2012.

[70] Rohinton T Kamakaka and Sue Biggins. Histone variants: deviants? *Genes & development*, 19(3):295–316, 2005.

[71] Paul B Talbert and Steven Henikoff. Histone variants—ancient wrap artists of the epigenome. *Nature reviews Molecular cell biology*, 11(4):264–275, 2010.

[72] Takeshi Tomonaga, Kazuyuki Matsushita, Seiko Yamaguchi, Tatsuya Oohashi, Hideaki Shimada, Takenori Ochiai, Kinya Yoda, and Fumio Nomura. Overexpression and mistargeting of centromere protein-a in human primary colorectal cancer. *Cancer research*, 63(13):3511–3516, 2003.

[73] Nicolas Lacoste, Adam Woolfe, Hiroaki Tachiwana, Ana Villar Garea, Teresa Barth, Sylvain Cantaloube, Hitoshi Kurumizaka, Axel Imhof, and Geneviève Almouzni. Mislocalization of the centromeric histone variant cenh3/cenp-a in human cells depends on the chaperone daxx. *Molecular cell*, 53(4):631–644, 2014.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

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

[81] 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.

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

[83] 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.

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

[85] 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.

[86] 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.

[87] 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.

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

[89] Ali Hamiche and Muhammad Shuaib. Chaperoning the histone h3 family. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(3):230–237, 2012.

[90] Francesca Mattiroli, Sheena D'Arcy, and Karolin Luger. The right place at the right time: chaperoning core histone variants. *EMBO reports*, page e201540840, 2015.

[91] 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.

[92] 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.

[93] Ewelina Zasadzińska, Meghan C Barnhart-Dailey, P Henning JL Kuich, and Daniel R Foltz. Dimerization of the cenp-a assembly factor hjurp is required for centromeric nucleosome deposition. *The EMBO journal*, 32(15):2113–2124, 2013.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

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

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

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

[102] Andrea Amadei, Marc A Ceruso, and Alfredo Di Nola. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins*, 36(4):419–424, 1999.

[103] Berk Hess. Convergence of sampling in protein simulations. *Phys. Rev. E*, 65(3):031910, 2002.

[104] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: structure, function, and genetics*, 23(4):566–579, 1995.

[105] Giorgio Parisi. Order parameter for spin-glasses. *Physical Review Letters*, 50(24):1946, 1983.

[106] Hidenori Kato, Bing-Rui Zhou, Hanqiao Feng, and Yawen Bai. An evolving tail of centromere histone variant cenp-a. *Cell Cycle*, 12(19):3133–3134, 2013.

[107] 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.

[108] Z Yu, X Zhou, W Wang, W Deng, J Fang, H Hu, Z Wang, S Li, L Cui, J Shen, and et al. Dynamic phosphorylation of cenp-a at ser68 orchestrates its cell-cycle-dependent deposition at centromeres. *Developmental cell*, 32(1):68–81, 2015.

[109] Christopher W Carroll, Kirstin J Milks, and Aaron F Straight. Dual recognition of cenp-a nucleosomes is required for centromere assembly. *The Journal of cell biology*, 189(7):1143–1155, 2010.

[110] 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, 2010.

[111] 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. *Scientific reports*, 4, 2014.

[112] Emily A Bassett, Jamie DeNizio, Meghan C Barnhart-Dailey, Tanya Panchenko, Nikolina Sekulic, Danielle J Rogers, Daniel R Foltz, and Ben E Black. Hjurp uses distinct cenp-a surfaces to recognize and to stabilize cenp-a/histone h4 for centromere assembly. *Developmental cell*, 22(4):749–762, 2012.

[113] Glennis A Logsdon, Evelyne J Barrey, Emily A Bassett, Jamie E DeNizio, Lucie Y Guo, Tanya Panchenko, Jennine M Dawicki-McKenna, Patrick Heun, and Ben E Black. Both tails and the centromere targeting domain of cenp-a are required for centromere establishment. *J. Cell Biol.*, 208(5):521–531, 2015.

[114] Delphine Quénet and Yamini Dalal. The cenp-a nucleosome: a dynamic structure and role at the centromere. *Chromosome research*, 20(5):465–479, 2012.

[115] Hiroaki Tachiwana, Sebastian Müller, Julia Blümer, Kerstin Klare, Andrea Musacchio, and Geneviève Almouzni. Hjurp involvement in de novo cenh3 cenp-a and cenp-c recruitment. *Cell reports*, 11(1):22–32, 2015.

[116] Christine M English, Melissa W Adkins, Joshua J Carson, Mair EA Churchill, and Jessica K Tyler. Structural basis for the histone chaperone activity of asf1. *Cell*, 127(3):495–508, 2006.

[117] Gaku Mizuguchi, Xuetong Shen, Joe Landry, Wei-Hua Wu, Subhojit Sen, and Carl Wu. Atp-driven exchange of histone h2az variant catalyzed by swr1 chromatin remodeling complex. *Science*, 303(5656):343–348, 2004.

[118] Francesca Mattiroli, Sheena D'Arcy, and Karolin Luger. The right place at the right time: chaperoning core histone variants. *EMBO reports*, page e201540840, 2015.

[119] Minou Bina-Stein and Robert T Simpson. Specific folding and contraction of dna by histones h3 and h4. *Cell*, 11(3):609–618, 1977.

[120] Barbara Sollner-Webb, Rafael D Camerini-Otero, and Gary Felsenfeld. Chromatin structure as probed by nucleases and proteases: Evidence for the central role of hitones h3 and h4. *Cell*, 9(1):179–193, 1976.

[121] Rifka Vlijm, Mina Lee, Orkide Ordu, Anastasiya Boltengagen, Alexandra Lusser, Nynke H Dekker, and Cees Dekker. Comparing the assembly and handedness dynamics of (h3. 3-h4) 2 tetrasomes to canonical tetrasomes. *PloS one*, 10(10):e0141267, 2015.

[122] Rifka Vlijm, Mina Lee, Jan Lipfert, Alexandra Lusser, Cees Dekker, and Nynke H Dekker. Nucleosome assembly dynamics involve spontaneous fluctuations in the handedness of tetrasomes. *Cell reports*, 10(2):216–225, 2015.

[123] Orkide Ordu, Leopold Kremser, Alexandra Lusser, and Nynke H Dekker. Modification of the histone tetramer at the h3-h3 interface impacts tetrasome conformations and dynamics. *The Journal of Chemical Physics*, 148(12):123323, 2018.

[124] Gina Arents, Rufus W Burlingame, Bi-Cheng Wang, Warner E Love, and Evangelos N Moudrianakis. The nucleosomal core histone octamer at 3.1 a resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences*, 88(22):10148–10152, 1991.

[125] Christopher M Wood, James M Nicholson, Stanley J Lambert, Laurent Chantalat, Colin D Reynolds, and John P Baldwin. High-resolution structure of the native histone octamer. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(6):541–545, 2005.

[126] Yasuo Tsunaka, Yoshie Fujiwara, Takuji Oyama, Susumu Hirose, and Kosuke Morikawa. Integrated molecular mechanism directing nucleosome reorganization by human fact. *Genes & development*, 30(6):673–686, 2016.

[127] Shoudeng Chen, Anne Rufiange, Hongda Huang, Kanagalaghatta R Rajashankar, Amine Nourani, and Dinshaw J Patel. Structure–function studies of histone h3/h4 tetramer maintenance during transcription by chaperone spt2. *Genes & development*, 29(12):1326–1340, 2015.

[128] Giulia Saredi, Hongda Huang, Colin M Hammond, Constance Alabert, Simon Bekker-Jensen, Ignasi Forne, Nazaret Reverón-Gómez, Benjamin M Foster, Lucie Mlejnkova, Till Bartke, et al. H4k20me0 marks post-replicative chromatin and recruits the tonsl–mms22l dna repair complex. *Nature*, 534(7609):714, 2016.

[129] Hongda Huang, Caroline B Strømme, Giulia Saredi, Martina Hödl, Anne Strandsby, Cristina González-Aguilera, Shoudeng Chen, Anja Groth, and Dinshaw J Patel. A unique binding mode enables mcm2 to chaperone histones h3–h4 at replication forks. *Nature Structural and Molecular Biology*, 22(8):618, 2015.

[130] Douglas D Banks and Lisa M Gloss. Equilibrium folding of the core histones: the h3- h4 tetramer is less stable than the h2a- h2b dimer. *Biochemistry*, 42(22):6827–6839, 2003.

[131] Douglas D Banks and Lisa M Gloss. Folding mechanism of the (h3–h4) 2 histone tetramer of the core nucleosome. *Protein science*, 13(5):1304–1316, 2004.

[132] Andrew Bowman, Richard Ward, Hassane El-Mkami, Tom Owen-Hughes, and David G Norman. Probing the (h3-h4) 2 histone tetramer structure using pulsed epr spectroscopy combined with site-directed spin labelling. *Nucleic acids research*, 38(2):695–707, 2009.

[133] Kizhakke M Sathyan, Daniele Fachinetti, and Daniel R Foltz. α-amino trimethylation of cenp-a by nrmt is required for full recruitment of the centromere. *Nature Communications*, 8:14678, 2017.

[134] Ana Stankovic, Lucie Y Guo, João F Mata, Dani L Bodor, Xing-Jun Cao, Aaron O Bailey, Jeffrey Shabanowitz, Donald F Hunt, Benjamin A Garcia, Ben E Black, and Lars ET Jansen. A dual inhibitory mechanism sufficient to maintain cell-cycle-restricted cenp-a assembly. *Molecular Cell*, 2016.

[135] Yael Nechemia-Arbely, Daniele Fachinetti, Karen H Miga, Nikolina Sekulic, Gautam V Soni, Dong Hyun Kim, Adeline K Wong, Ah Young Lee, Kristen Nguyen, Cees Dekker, Bing Ren, Ben E Black, and Don W Cleveland. Human centromeric cenp-a chromatin is a homotypic, octameric nucleosome at all cell cycle points. *J Cell Biol*, 216(3):607–621, 2017.

[136] Kehui Wang, Zhouliang Yu, Yuting Liu, and Guohong Li. Ser68 phosphorylation ensures accurate cell-cycle-dependent cenp-a deposition at centromeres. *Developmental Cell*, 40(1):5–6, 2017.

[137] Minh Bui, Mary Pitman, Arthur Nuccio, Serene Roque, Paul Gregory Donlin-Asp, Aleksandra Nita-Lazar, Garegin A Papoian, and Yamini Dalal. Internal modifications in the cenp-a nucleosome modulate centromeric dynamics. *Epigenetics & chromatin*, 10(1):17, 2017.

[138] Meghan C Barnhart-Dailey, Prasad Trivedi, P Todd Stukenberg, and Daniel R Foltz. Hjurp interaction with the condensin ii complex during g1 promotes cenp-a deposition. *Molecular Biology of the Cell*, 28(1):54–64, 2017.

[139] 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 a resolution. *J. Mol. Biol.*, 319(5):1097–1113, June 2002.

[140] Stefan Boresch and Martin Karplus. The jacobian factor in free energy simulations. *The Journal of chemical physics*, 105(12):5145–5154, 1996.

[141] 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.

[142] Ewelina Zasadzińska, Meghan C Barnhart-Dailey, P Henning JL Kuich, and Daniel R Foltz. Dimerization of the cenp-a assembly factor hjurp is required for centromeric nucleosome deposition. *The EMBO journal*, 32(15):2113–2124, 2013.

[143] Wallace H Liu, Sarah C Roemer, Yeyun Zhou, Zih-Jie Shen, Briana K Dennehey, Jeremy L Balsbaugh, Jennifer C Liddle, Travis Nemkov, Natalie G Ahn, Kirk C Hansen, Jessica K Tyler, and Mair EA Churchill. The cac1 subunit of histone chaperone caf-1 organizes caf-1-h3/h4 architecture and tetramerizes histones. *eLife*, 5:e18023, 2016.

[144] Francesca Mattiroli, Yajie Gu, Tejas Yadav, Jeremy L Balsbaugh, Michael R Harris, Eileen S Findlay, Yang Liu, Catherine A Radebaugh, Laurie A Stargell, Natalie G Ahn, Iestyn Whitehouse, and Karolin Luger. Dna-mediated association of two histone-bound caf-1 complexes drives tetrasome assembly in the wake of dna replication. *eLife*, 6:e22799, 2017.

[145] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J De Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of dna: Structure, thermodynamics, and dynamics of hybridization. *The Journal of chemical physics*, 139(14):10B604_1, 2013.

[146] Bin Zhang, Weihua Zheng, Garegin A Papoian, and Peter G Wolynes. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.*, 138(26):8126–8133, 2016.

[147] R Vlijm, SH Kim, PL De Zwart, Y Dalal, and C Dekker. The supercoiling state of dna determines the handedness of both h3 and cenp-a nucleosomes. *Nanoscale*, 9(5):1862–1870, 2017.