Biophysical Journal
Article

Biophysical Society

The Oligomerization Landscape of Histones

Haiqing Zhao,[1,2] David Winogradoff,[3] Yamini Dalal,[2,*] and Garegin A. Papoian[1,3,4,*]
[1]Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland; [2]Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; [3]Chemical Physics Program, Institute for Physical Science and Technology; and [4]Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland

ABSTRACT In eukaryotes, DNA is packaged within nucleosomes. The DNA of each nucleosome is typically centered around an octameric histone protein core: one central tetramer plus two separate dimers. Studying the assembly mechanisms of histones is essential for understanding the dynamics of entire nucleosomes and higher-order DNA packaging. Here, we investigate canonical histone assembly and that of the centromere-specific histone variant, centromere protein A (CENP-A), using molecular dynamics simulations. We quantitatively characterize their thermodynamical and dynamical features, showing that two H3/H4 dimers form a structurally floppy, weakly bound complex, the latter exhibiting large instability around the central interface manifested via a swiveling motion of two halves. This finding is consistent with the recently observed DNA handedness flipping of the tetrasome. In contrast, the variant CENP-A encodes distinctive stability to its tetramer with a rigid but twisted interface compared to the crystal structure, implying diverse structural possibilities of the histone variant. Interestingly, the observed tetramer dynamics alter significantly and appear to reach a new balance when H2A/H2B dimers are present. Furthermore, we found that the preferred structure for the $(CENP\text{-}A/H4)_2$ tetramer is incongruent with the octameric structure, explaining many of the unusual dynamical behaviors of the CENP-A nucleosome. In all, these data reveal key mechanistic insights and structural details for the assembly of canonical and variant histone tetramers and octamers, providing theoretical quantifications and physical interpretations for longstanding and recent experimental observations. Based on these findings, we propose different chaperone-assisted binding and nucleosome assembly mechanisms for the canonical and CENP-A histone oligomers.

INTRODUCTION

Eukaryotes wrap their DNA around histone proteins constituting the fundamental unit of chromatin, the nucleosome. Inside each nucleosome, histones typically exist as an octamer, composed of a central tetramer $(H3/H4)_2$ plus one H2A/H2B dimer on either side (1). Nucleosomes dynamically dissociate and reassociate in chromatin structure for fundamental biological processes such as DNA transcription, replication, and repair. By initiating nucleosome assembly through forming a tetrasome with DNA, the histone tetramer serves as the structural basis for nucleosomal or chromatin dynamics (2,3). Thus, it is crucial to elucidate the dynamics of histone tetramers, which are key intermediates along nucleosome assembly and disassembly pathways. Recent single-molecule experiments studied the spontaneous flipping behavior of DNA handedness of the tetrasome, finding that iodoacetamide-treated residue mutations on the tetramer can cause the enhanced flexibility and faster superhelical flipping kinetics of the wrapped DNA (4–6). Therefore, a deep molecular understanding of histone tetramer dynamics is not only essential to understanding subnucleosomal or nucleosomal assembly but may also suggest innovative pathways for higher-order DNA packaging.

The centromere-specific histone H3 variant, centromere protein A (CENP-A), has been extensively studied in recent decades for: 1) its significant functional role as the epigenetic mark of centromere ensuring proper chromosome separation during cell division (7–12) and 2) its unique structural dynamics (13–15), particularly dissecting the dominant structure of CENP-A nucleosomes (16–21) and their special association with kinetochore partners (22–26). Unlike canonical H3 nucleosomes, CENP-A-containing nucleosomes follow a different assembly pathway via the unique chaperone HJURP (27–31). Also, in cancer cells, CENP-A is overexpressed, and the redundant CENP-A can localize into ectopic (i.e., noncentromeric) regions via alternative pathways (32,33). Thus, one outstanding question is whether CENP-A, in normal cells,

Submitted October 4, 2018, and accepted for publication March 14, 2019.
*Correspondence: dalaly@mail.nih.gov or gpapoian@umd.edu
Editor: Tamar Schlick.
https://doi.org/10.1016/j.bpj.2019.03.021

can be efficiently regulated to avoid ectopic delivery. Another related question is whether replacing canonical H3 with CENP-A alters its physical properties and overall dynamics.

Conflicting studies have suggested the following: 1) in vitro chromatography and deuterium exchange experiments indicate that the soluble CENP-A/H4 forms a more compact and rigid tetramer than the conventional H3 complex (34), and partially truncated CENP-A tetramers adopt compact conformations in crystals and in solution (16); 2) CENP-A- and H3-containing nucleosomes have nearly identical crystal structures (35,36); and 3) recent computational and experimental studies reveal that CENP-A dimers (37) and nucleosomes (38,39) are more flexible than their canonical H3 counterparts. On the other hand, canonical histone tetramers present consistent crystal structures in different molecular contexts, including as a tetramer in a nucleosome (1,40), in an octamer (41–43), and in complexes with chaperones such as FACT (44), Spt2 (45), TONSL, and MCM2 (46,47). Early size-exclusion chromatography experiments demonstrate that there is a dynamic equilibrium between two H3/H4 dimers and an assembled tetramer (48,49), and this equilibrium is responsive to changes in ionic strength (50). Through electron paramagnetic resonance (EPR) spectroscopy, a previous study shows that the canonical histone tetramer exhibits greater structurally heterogeneity on its own than when sequestered in the octamer (51). However, dynamical structural details that would reveal the mechanisms governing observed properties are not readily amenable to existing experimental techniques. Here, we apply computational modeling to study both H3 and CENP-A oligomers aiming to provide a comprehensive theoretical quantification that can explain and unify these experimental observations that might seem incompatible.

Among computational approaches, 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 around their native states (38), and its dynamics can be modulated by internal modifications (52). Combining coarse-grained, atomistic simulations and in vivo mutation experiments, we reported that the CENP-A dimer is structurally variable, and chaperone HJURP prevents the promiscuous misassembly of the CENP-A dimer, protecting it from binding with other proteins (37).

Building upon these findings, we performed coarse-grained MD simulations using the associative-memory, water-mediated structure and energy model (AWSEM) (53,54) to investigate the assembly mechanisms of histone oligomers and asked whether histones CENP-A and H3 differ at the tetramer and octamer levels. We computed the association free energy of two dimers forming a tetramer, finding that CENP-A dimers form a more compact and stable tetramer with a more favorable free energy, whereas the landscape of H3 dimers is more rugged, indicating its structural lability. In particular, simulations starting from preassembled tetramers reveal a swiveling motion around the H3 tetrameric interface. Furthermore, histone octamer simulations suggest that the addition of H2A/H2B dimers gently restrains the internal rotation of the H3 tetramer. In contrast, H2A/H2B addition causes the CENP-A tetramer to adopt multiple conformational states, demonstrating a significant incongruence between the preferred structures of the CENP-A tetramer versus the octamer. Finally, we put forward a speculative model for canonical and variant histone assembly and propose that the CENP-A tetramer may serve as a critical sequestration channel, preventing the assembly of excess CENP-A into chromatin, thereby regulating CENP-A homeostasis in vivo.

MATERIALS AND METHODS

Initial structures

Despite diverse structural environments, the canonical histone tetramer adopts a consistent configuration in the crystal structures of histone octamer, nucleosome, and protein complex with chaperone protein(s) (detailed comparisons are provided in Fig. S4). In this work, we took the tetramer structure from a nucleosome crystal structure containing H3 (Protein Data Bank (PDB): 1KX5 (40)). Initial configurations for the CENP-A tetramer were obtained from the CENP-A-containing nucleosome (PDB: 3AN2 (35)) and the αN-helices-truncated CENP-A tetramer crystal structure (PDB: 3NQJ (16)). This study does not include histone tails and DNA. Their effects are discussed in the Discussion. More structural information of the excluded nucleosomal DNA and histone tails (Fig. S5) and information about simulated protein length and their sequences (Fig. S6) are covered in the Supporting Materials and Methods.

Simulation methods and trajectory analyses

In this work, we used AWSEM to carry out computational simulations for both canonical and variant CENP-A histone oligomers. AWSEM is a coarse-grained protein model with three beads (C_α, C_β, and O) representing one amino acid. The total potential function includes the $V_{backbone}$ term for protein backbone formation, residue-residue and residue-water interaction terms $V_{contact}$ and V_{burial}, hydrogen bonding term V_{HB}, and the associate memory term V_{AM} (Eq. 1). Details of each potential term are described in the Supporting Material of (54) and the Supporting Materials and Methods of this work. This model is based on not only the physical interactions like $V_{backbone}$ and V_{HB} but also the bioinformatics-inspired fragment memory term V_{AM} (Fig. S1). Here, we use the respective histone monomer structures to build the biasing structural fragment memory database, wherein each fragment is 3- to 12-residues long. It is important to note that no intermolecular information between either monomers or dimers was provided to the force field. So, from this perspective, AWSEM is used as a predictive protein model. In the constant temperature simulations of two dimers, a weak distance constraint in the harmonic potential form (spring constant $k = 0.02$ kcal/mol/Å^2) is applied between the centers of mass of each dimer. Constraints of the same magnitude are applied between the two H2A/H2B dimers and between each H2A/H2B dimer and the tetramer. These constraints ensure the two objects are within a reasonable distance of each other for possible interactions.

$$V_{AWSEM} = V_{backbone} + V_{contact} + V_{burial} + V_{HB} + V_{AM} \quad (1)$$

All simulations were performed in the large-scale atomic/molecular massively parallel simulator 2016, using the Nosé-Hoover thermostat. We applied umbrella sampling together with replica exchange (55) to enhance

the phase space sampling for further free-energy calculations. For instance, in the case of H3 dimer association, two H3 dimers were put in the simulation with the distance between their centers of mass controlled by an umbrella constraint. A typical harmonic potential is used for this purpose as shown in Eq. 2, where k_R is the biasing strength and R_o is the controlled center distance for each window. Here, $k_R = 5$ kcal/mol/Å^2, and R_o ranges from 20 to 50 Å, averagely spaced by 1 Å. Simulations for each umbrella window, 30 in total, were performed using 10 replicas with temperatures linearly ranging from 280 to 370 K. After the simulation reached convergence (see Supporting Materials and Methods), data from different windows were then collected, and the weighted histogram analysis method (56) was applied to calculate the potentials of mean force and construct free-energy landscapes onto different coordinates. A relevant Jacobian factor correction term, $k_BT\ln[4\pi R^2]$, was subtracted from the free-energy calculation because a sampling space based on the distance R_{COM} makes nonphysical contributions to the configurational partition function (57). The time step was set at 5 fs in all simulations. Each replica was run for two million steps. Exchanges between replicas were attempted every 400 steps. The first 0.5 million steps were not used in the analysis for system's equilibration.

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

Separately, 10 independent constant temperature simulations were carried out for tetramers $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$, with 30 million time steps each and 300 million steps in total (1500 ns in the coarse-grained timescale). Weak biases in the form of harmonic potential were applied as mentioned above. Simulations and analyses for H3 and CENP-A tetramers that exclude αN helices were performed using the same setup. Octamer simulations for $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$ with two (H2A/H2B) s were run for 10 million time steps, totaling 100 million time steps for each octamer system. 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. Note that the coarse-graining timescale cannot be directly converted into real timescale because it could be at least 10 times larger than that in the atomistic MD simulations (58). The convergences of all simulations were verified by the root mean square inner-product analysis, which are provided in Supporting Materials and Methods, Section S4.

All the trajectory analyses in this work, including the calculations of root mean square deviations (RMSDs), radius of gyration (R_g), distances (R), dihedral angles θ, Q values, and contact analysis, were based on the Cα coordinates. More analyzing details, including error analyses, are included in Supporting Materials and Methods, Sections S2 and S3. P-values from t-tests are calculated and provided along with all probability distributions.

RESULTS

Binding free energy of two dimers forming a tetramer

Motivated by the previous observation of CENP-A dimer flexibility (37) compared with its canonical counterpart, we first investigated the formation of tetramers from two canonical H3 and CENP-A dimers. Via a mixed enhanced sampling methodology that couples replica exchange with umbrella sampling, we mapped their corresponding binding-free-energy landscapes. The calculated free-energy profiles (FEPs) were projected into two reaction coordinates: the distance between centers of mass of the two dimers, R_{COM}, and another order parameter, $Q_{interface}$, that quantifies the nativeness of the binding interface between the dimers. $Q_{interface}$ is the fraction of native interface contacts, defined as, $(1/n)\sum_{i<j-2}\exp[-((r_{ij} - r_{ij}^{native})^2/2\sigma_{ij}^2)]$, 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. Here, the $Q_{interface}$ calculations were conducted with respect to the tetramer interface from the corresponding nucleosomes containing canonical H3 and variant CENP-A (PDB: 1KX5 (40) and 3AN2 (35)).

As seen in Fig. 1, the binding-free-energy landscape for H3/H4 dimers is relatively rugged, with multiple energy minima at $Q_{interface} = 0.4$, ~0.1–0.2, and 0.0 and R_{COM} of 29, ~32–33, and 38 Å in the other dimension (Fig. 1 A). These minima are relatively flat compared to that of CENP-A, occupying a large portion of configuration space described in terms of R_{COM} and $Q_{interface}$, indicating larger structural heterogeneity of $(H3/H4)_2$ with a broad ensemble of accessible conformations. This result is consistent with the experimental observation that histone H3 tetramer is unstable at moderate ionic strengths, determined by guanidinium-induced denaturation (48). On the other hand, the two CENP-A/H4 dimers present a deep, well-funneled association-free-energy landscape (Fig. 1 B), with the minimum at $R_{COM} = 29$ Å, $Q_{interface} = 0.4$, suggesting that there is a thermodynamically favorable binding state forming the tetramer $(CENP\text{-}A/H4)_2$.

To further quantitatively compare the binding energy differences between H3/H4 and CENP-A/H4, we projected both two computed FEPs along R_{COM} after aligning the far-ends of two converged curves to zero, at which (i.e., when $R_{COM} > 50$ Å) we assume there is no interaction between any two dimers. Fig. 2 presents the FEP as a function of the distance between the COMs of two H3 dimers or CENP-A dimers, demonstrating that the FEP minima for $(CENP\text{-}A/H4)_2$ and $(H3/H4)_2$ are correspondingly at −7 and −3 kcal/mol. Because 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 (Fig. 2), indicating that the thermodynamically favored CENP-A tetramer is more compact than the H3 tetramer. This result agrees quantitatively with previous SAXS measurements that found the CENP-A tetramer to be substantially more compact relative to its H3 counterpart (16). Also, Banks and Gloss used far-ultraviolet circular dichroism to measure the folding and unfolding kinetics of $(H3/H4)_2$ experimentally (49). The free energy of the dimer-tetramer reaction they obtained is −2.6 kcal/mol. Our computed minimum, at

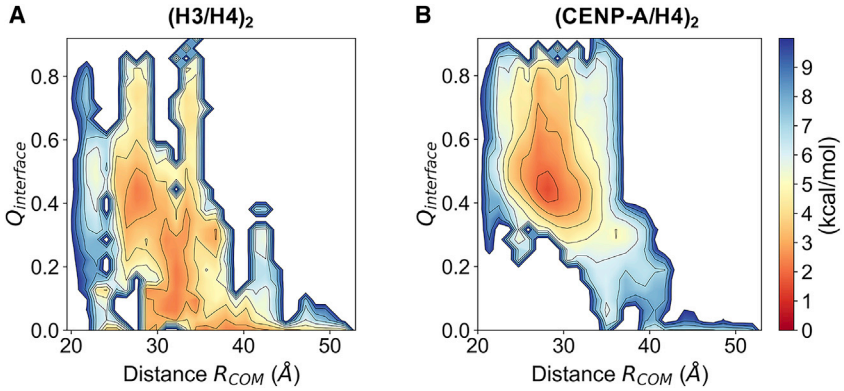

FIGURE 1 The binding-free-energy landscapes of two H3 dimers and that of two CENP-A dimers. Two-dimensional FEPs are mapped as a function of the distance between two interacting dimers R_{COM} and of the quantification of the nativeness of their binding interface $Q_{interface}$ for $(H3/H4)_2$ (*A*) and $(CENP\text{-}A/H4)_2$ (*B*). To see this figure in color, go online.

-3 ± 0.2 kcal/mol, is in qualitative agreement with their experimentally measured data. Overall, in agreement with experimental work (Black et al. (34)), we find that the CENP-A homotypic tetramer, on its own, is thermodynamically more stable and more compact than the tetramer of H3/H4. Additional FEPs projected on other reaction coordinates, both one-dimensional and two-dimensional, are provided in Figs. S4 and S5.

Tetramer geometries and the swiveling dynamics

To further explore the intrinsic dynamics of histone tetramers, we performed microsecond-scale continuous constant temperature CG-AWSEM simulations for CENP-A and H3 tetramers, starting from preassembled conformations taken from the central tetramers of the corresponding octameric nucleosome crystal structures (Fig. 3 A). Other structures from octamer or chaperone-tetramer complexes could have been used as well because the overall structures of the tetramer are nearly identical despite divergent crystallization conditions (Fig. S4). Overall, our results obtained from these continuous simulations were broadly consistent

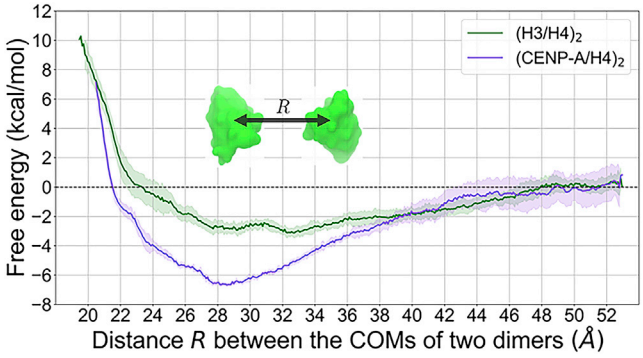

FIGURE 2 $(CENP\text{-}A/H4)_2$ has a deeper FEP than $(H3/H4)_2$. The potential of mean force along the distance *R* between histone dimers is deeper for $(CENP\text{-}A/H4)_2$ (*purple*) than for $(H3/H4)_2$ (*green*). *R* is measured from the center of mass of one dimer to the other. The shaded areas illustrate the standard deviations (SD) of the curves. To see this figure in color, go online.

with the above enhanced sampling simulations, and they provide important dynamics insights. We present here some of the most salient observations; additional analyses, including the principal component analysis and other structural quantities, including the RMSD, R_{COM}, and $Q_{interface}$, can be found in Figs. S7, *A–C* and S9.

To quantify the molecules' degree of compaction, we calculated the tetramer's radius of gyration, defined as $R_g = \sqrt{(1/N^2)\sum_{i=1,i<j}^{N}(\boldsymbol{r_i} - \boldsymbol{r_j})^2}$, where N is the total number of residues and $\boldsymbol{r}_i$ are the coordinates of the ith residue. Fig. 3 B shows that the average R_g for $(CENP\text{-}A/H4)_2$ is 21 ± 0.7 and 23 ± 1.4 Å for $(H3/H4)_2$, implying that $(CENP\text{-}A/H4)_2$ samples more compact geometries with fewer R_g fluctuations. The R_g distribution of $(CENP\text{-}A/H4)_2$ is unimodal, with a dominant central peak, whereas the H3 tetramer R_g samples a much broader distribution (Fig. 3 B), consistent with the above free-energy calculations (Fig. 2). Moreover, the mean value difference between the two systems in our simulation matches the previous experimental data, in which Black et al. measured that the CENP-A tetramer complex chromatographs as a single species with a Stokes radius of 28 Å, smaller than that of H3/H4, 30.5 Å (34). Together, these results suggest that the CENP-A tetramer is more closely packed and structurally more well-defined than the canonical H3 tetramer.

In recent magnetic tweezer experiments, the DNA of H3-containing tetrasomes were observed to flip between left- and right-handed superhelically wound states (5,6), which may be initiated by conformational changes of the proteins inside. To better compare with these experiments, we examined the overall orientation of the simulated tetramers by measuring the dihedral angle between the two composing dimers. We chose to measure the dihedral angle of the two H3 $\alpha 2$ helices (similarly for CENP-A) because they are the longest continuous structural component in each dimer molecule.

Our results demonstrate that compared to $(CENP\text{-}A/H4)_2$, the two H3 dimers in $(H3/H4)_2$ occupy a range of orientations, as indicated by the distribution of abovementioned dihedral angles that includes three populations

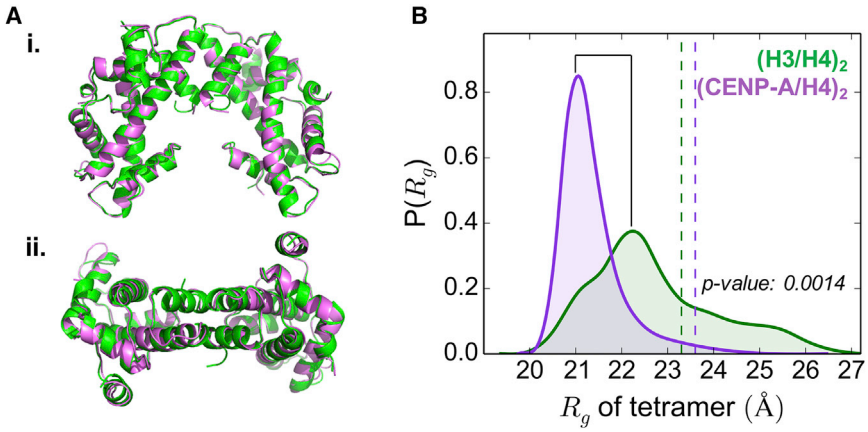

FIGURE 3 $(CENP\text{-}A/H4)_2$ is more compact than $(H3/H4)_2$. (*A*) The initial conformations of the H3 tetramer (*green*) and CENP-A tetramer (*purple*) were taken from their nucleosome crystal structures (PDB: 1KX5 and 3AN2). Lateral view (*i*) and top view (*ii*) of aligned structures are displayed. (*B*) The CENP-A tetramer has a smaller radius of gyration, R_g, than the H3 tetramer, with a narrower distribution. The vertical dashed lines mark the measured R_g values of the initial structures. To see this figure in color, go online.

(Fig. 4 B): one positive and two negative, three distinct states in total (Fig. 4, A and B *i*, *ii*, *iii*). Furthermore, $(H3/H4)_2$ frequently transits from one dihedral angle to another (15 major switches in the measurement of dihedral angle), undergoing swiveling motion around the binding interface (Fig. S10; Video S1). The range of orientations for two histone dimers and its dynamical transition found in our simulations can explain the spontaneous flipping behavior of DNA handedness in the tetrasome as revealed in magnetic tweezer experiments (6). A positive dihedral angle of the tetramer would correspond to left-handed superhelically wrapped DNA and vice versa (Fig. 4 C).

In contrast, $(CENP\text{-}A/H4)_2$ maintains a relatively fixed orientation, with no obvious rotational motions around the interface (one switch in the measurement of dihedral angle, Fig. S10; Video S2). The dihedral angle between the scaffold helices is about 90° (Fig. 4, A and B *iv*), less than the angle measured in crystal structures of the CENP-A tetramer from the nucleosome or with chaperones (110°), implying a twisted interface geometry. 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 occupy the same plane compared to the geometries of their respective nucleosome structures (Fig. 4 A). The $\alpha 2$ helices of CENP-A were found to be curved (Fig. S18), as was also revealed from experimental observation (16). The curvature of $\alpha 2$ helices could be a result of the absence of surrounding DNA and bracketing H2A/H2B, which provides necessary topological support to the central tetramer.

Distinct dynamics at the tetrameric interface

To uncover the mechanistic basis for the observed difference in behavior between CENP-A and H3 tetramers, we then assessed whether it arises from the tetrameric interface (i.e., the interface between two dimers). We calculated $Q_{interface}$ for the continual simulations, referring to the native interface contacts from the crystal structure (PDB: 1KX5). The distribution for the CENP-A tetramer is centered at 0.5, whereas the same distribution for the H3 tetramer contains three peaks, with an average value of 0.2 (Fig. S8 *B*). This result implies that $(CENP\text{-}A/H4)_2$ forms a tetrameric interface that is better defined and more native-like compared to $(H3/H4)_2$. In the context of the DNA-free tetramer, the four-helix bundle of $(CENP\text{-}A/H4)_2$ that comprises the tetrameric interface still maintains a well-connected and symmetric geometric arrangement (Fig. 5 B), despite some structural twisting. This is not found in the H3 tetramer case.

Furthermore, we performed contact analysis for the four-helix bundle region. It demonstrates that there are more contacts, on average, in the corresponding region of $(CENP\text{-}A/H4)_2$ (∼27) than in the same region of $(H3/H4)_2$ (∼17) (Fig. 5 A). Also, one dominant peak is found in the $(CENP\text{-}A/H4)_2$ contact histogram, but two peaks exist in that of $(H3/H4)_2$. Detailed residual pair interactions from AWSEM show that the CENP-A residues Leu111, Gln127, and Arg131 contribute strong hydrophobic interactions to the four-helix bundle tetramer interface (Table S1), which H3 lacks. $(CENP\text{-}A/H4)_2$ maintains a well-defined, native-like four-helix bundle throughout the simulation (Fig. 5 B), with the αN helices remaining outside the central interface. Note that the previously suggested CENP-A Leu112 residue (16), which is next to Leu111, is not found in the top strong interacting pairs of our simulations. The reason for this discrepancy is unclear.

Meanwhile, we observed structurally heterogeneous H3 αN helices, consistent with previous EPR experimental findings (51). Moreover, we notice that the αN sections of histone H3 may play an important role in disrupting the four-helix bundle at the H3 tetrameric interface (Fig. 5 B; Fig. S11). Indeed, the distances between the αN helices from each H3 copy shows that the H3 αN helices feature a considerably wide distribution, including two prominent peaks (at ∼20 and 32 Å apart) (Fig. S9 *D*). Further, this disruption is mainly mediated through the hydrophobic interactions between Val46, Ala47, and Leu48 from αN and

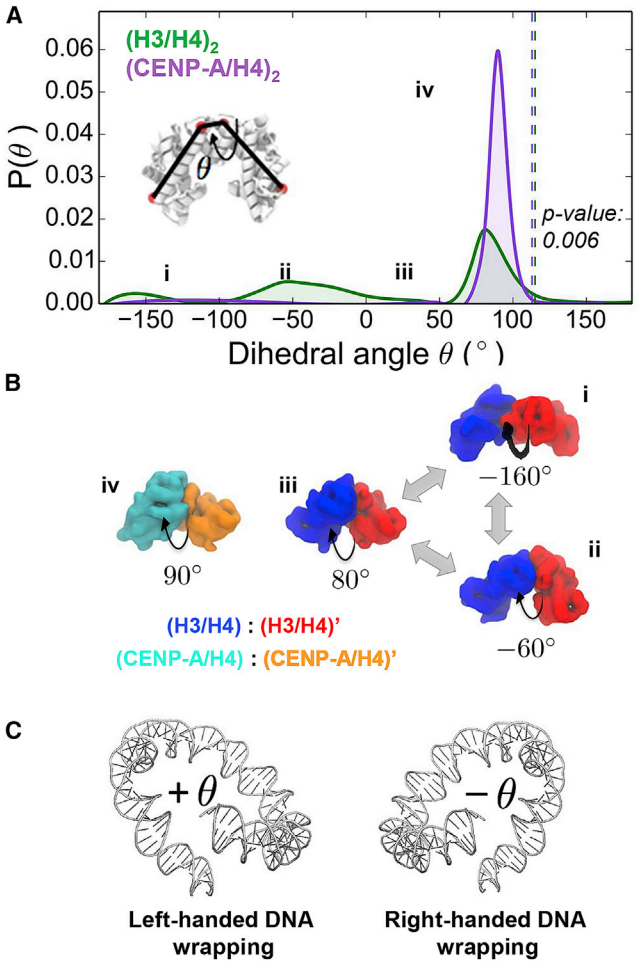

FIGURE 4 The H3 tetramer swivels around its binding interface, whereas the CENP-A tetramer remains relatively stable. (*A*) The distribution for the dihedral angle between $\alpha 2$ helices features one prominent peak for $(CENP\text{-}A/H4)_2$ and three smaller peaks for $(H3/H4)_2$, indicating $(CENP\text{-}A/H4)_2$ maintains a more fixed orientation than $(H3/H4)_2$. Vertical dashed lines are the corresponding dihedral angles of the initial structures. (*B*) Representative conformations from each population are displayed. (*C*) Positive (+) and negative (−) dihedral angles of the histone tetramer measured here correspond to the left-handed and right-handed DNA superhelical wrapping in the tetrasome, respectively. To see this figure in color, go online.

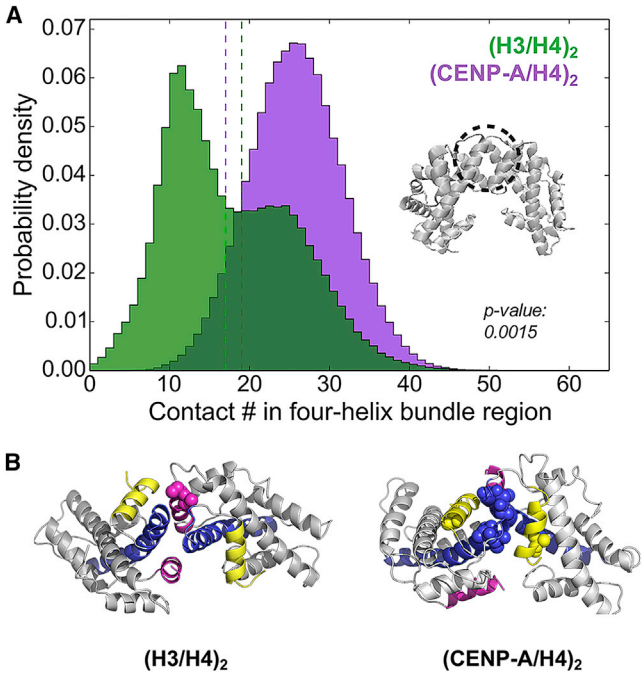

FIGURE 5 $(CENP\text{-}A/H4)_2$ has a more stable four-helix bundle than $(H3/H4)_2$. (*A*) $(H3/H4)_2$ (*green*) forms fewer contacts than $(CENP\text{-}A/H4)_2$ (*purple*) in the four-helix bundle region. The histogram of the number of contacts for $(H3/H4)_2$ has two peaks at 13 and 25, whereas $(CENP\text{-}A/H4)_2$ has a single peak at 27. The dash lines mark the four-helix contacts number in corresponding crystal structures. (*B*) Corresponding representative structures demonstrate that the $(H3/H4)_2$ four-helix bundle becomes broken or disrupted by αN helices (*pink*), whereas the four-helix bundle ($\alpha 2$ and $\alpha 3$, *blue* and *yellow*) remains stable in $(CENP\text{-}A/H4)_2$ throughout the simulation. $\alpha 2$ and $\alpha 3$ helices are marked in blue and yellow. CENP-A-specific residues L112, T113, L114, and V126 and H3-specific V46 and A47 are shown in coarse-grained spheres. To see this figure in color, go online.

Leu107 and Ala111 from $\alpha 2$. The αN helix of H3 has greater hydrophobicity than CENP-A does, which could explain, in part, 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. We tested this hypothesis by performing similar simulations for both systems but starting from structures excluding αN helices. The analyses (Supporting Materials and Methods, Section S11) confirmed our hypothesis that the flexible αN helices play a significant role in the swiveling motion of the H3 tetramer because switching between different H3 tetrameric dihedrals is significantly diminished when αN helices are excluded (Fig. S13 *D*). Interestingly, even without αN helices, CENP-A still forms more four-helix contacts (Fig. S13 *B*) and a more native-like binding interface (Fig. S13 *C*) than the H3 tetramer.

Hence, from this analysis, we suggest that 1) specifically in the DNA-free tetramer context, the unique hydrophobic residues (Leu112, Thr113, Leu114, Val126) at the CENP-A:CENP-A interface may help contribute an intrinsically stronger four-helix bundle than H3; and 2) the more hydrophobic H3 αN helix (Val46, Ala47, Leu48) tends to disrupt the relatively weak four-helix bundle formation and lead to the swiveling motion around the H3 tetramer interface.

Effects of H2A/H2B on histone tetramers

Finally, we wanted to examine the effects of histone dimer H2A/H2B on the dynamics of tetramers $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$. We investigated the DNA-free canonical H3 and variant CENP-A octamers using similar simulation procedures. Both the H2A/H2B dimers maintained well-native conformations throughout the simulations (Fig. S14 *D*). However, their distances to the central tetramer are diverse for H3 and CENP-A cases (Fig. S14 *C*), implying different effects of H2A/H2B on each tetramer.

As done for tetramers, similar analyses such as R_{COM}, R_g, and tetrameric dihedral θ were performed to explore the dynamics features of the histone octamers. For the H3 octamer, the distributions of both the tetrameric R_g and the distance R between H3/H4 pairs become more focused and Gaussian-like, compared to the solo tetramer situation ("solo" refers to the tetramer in isolation, without any other proteins; Fig. 6 A versus Fig. 3 B; Fig. S14 *B* versus Fig. S9 *B*). The SD decreases from 3.8 to 1.9 Å for R and from 1.4 to 0.7 Å for R_g, agreeing with previous EPR experimental data (51) showing the reduced H3 tetramer flexibility in an octamer. The distribution of the tetrameric dihedral angles of H3 features a dominant peak at 90° (Fig. 6 B), similar to that measured in the case of CENP-A, with the other two populations observed in solo H3 tetramer simulations diminished. 84% of H3 tetramer conformations in the octamer simulations have a positive tetrameric dihedral angle, significantly more than that in the solo tetramer simulations (58%). These data establish that the swiveling motion around the binding interface was reduced because of the bracketing histone dimers H2A/H2B on either side of the tetramer.

Nevertheless, 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 $(CENP\text{-}A/H4)_2$ 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 for the solo CENP-A tetramer (Fig. 6 A versus Fig. 3 B; Fig. S14 *B* versus Fig. S9 *B*). 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 that CENP-A-containing histone nucleosomes or octamers are structurally more flexible and heterogeneous than their canonical counterparts (38,39).

DISCUSSION

Maturation of the nucleosome stepping through dimers, tetramers, and octamers

In this work, we used coarse-grained modeling to study the thermodynamical and dynamical properties of canonical and variant CENP-A histone oligomers. We comprehensively compared the association energy of H3 dimers and CENP-A dimers forming their corresponding homotypic tetramers, inferring an energy difference of 4 kcal/mol between the two types. We also observed that the H3 αN helices enhance the lability of H3 tetramer, contributing to the overall swiveling motion around the less hydrophobic H3 tetrameric interface. The addition of (H2A/H2B)s restrains the flexibility of H3 and pushes CENP-A to adopt multiple conformations. Our results are largely in agreement with the prior experimental observations on these systems, including H3 tetramer (48,49) and octamer (51) and H3 tetrasome (5) and CENP-A tetramer (16,34).

We previously 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 (37). Here in the context of a tetramer, analyses of the monomer and dimer components yielded consistent results (see Supporting Materials and Methods, Section S13). For instance, the average $Q_{monomer}$ for H4 is larger than that of H3 or CENP-A (Fig. S16), implying its noticeable stability; Q_{dimer} and $Q_{dimer,\ interface}$ for H3 are larger, on average, than for CENP-A (Fig. S15), indicating that H3 dimers adopt more native-like conformations than CENP-A dimers. However, compared to the structural variabilities within the dimer level, the movements between dimers forming the tetramer are on a larger scale, with an RMSD of ~10–15 Å for the tetramer (Fig. S9 *A*) vs. ~3–4 Å for the dimer (Fig. S15 *B* and Fig. 2 in (37)). Therefore, the dynamics observed here by coarse-grained modeling are unlikely to be sampled ergodically using present-day atomistic simulations.

In our earlier study, the CENP-A nucleosome was shown to be more flexible than the H3 nucleosome, revealing a shearing motion at the tetramer interface

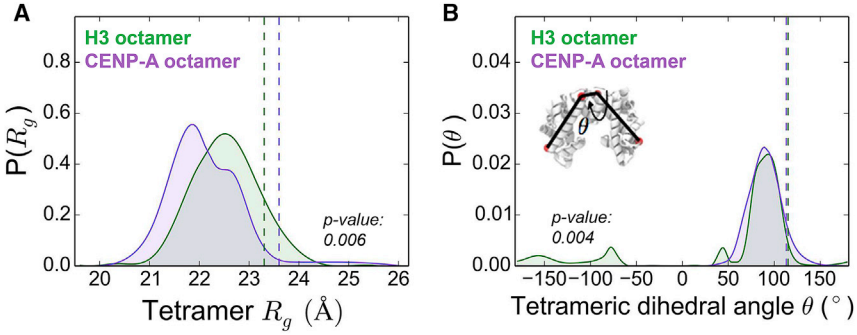

FIGURE 6 H2A/H2B stabilizes $(H3/H4)_2$ but not $(CENP\text{-}A/H4)_2$. (*A*) The probability distribution of H3 tetramer R_g features a more focused peak in the context of an octamer compared to that of the solo H3 tetramer (Fig. 3 C), whereas one peak and one shoulder exist in the same distribution 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°. Vertical dashed lines in both panels are the corresponding tetrameric R_g values and dihedral angles measured from the initial octamer. To see this figure in color, go online.

(38). Here, in the context of an octamer with H2A/H2B dimers, the CENP-A tetramer occupies two distinct conformational states: one that is similar to that of the isolated tetramer conformation and another less compact, structurally similar to the nucleosomal H3 (or CENP-A) tetramer. Therefore, disrupting the energetically stable interface of the CENP-A tetramer likely underpins the shearing motion observed in the octameric CENP-A nucleosome. The two-state memory of the CENP-A tetramer in the octamer may explain why the CENP-A nucleosome is more distortable and dynamic compared with the canonical one.

Decades of work in the chromatin field have demonstrated the crucial importance of not just histone variants but also the N-terminal tails and post-translational covalent modifications of histones. Together, all of these factors contribute to nucleosome dynamics (59–65) and alter not only the folded state of the chromatin fiber (66–71) but also the affinities of chromatin effector proteins. Additionally, the DNA sequence context is also crucial in determining nucleosome stability (72,73), nucleosome phasing (74), nucleosome positioning (75–77), and nucleosome spacing (78), all of which determine DNA accessibility (79). These critical epigenetic and genetic components will need to be studied rigorously in silico to arrive at a holistic representation of the epigenetic landscape of eukaryotic genomes.

Biological implications

We consider 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 the surrounding DNA, histone (H2A/H2B)s, or chaperone proteins. The canonical tetramer may have evolved to highly depend on other structural partners, which may be key to ensure the fidelity and stability of genetic material. On the other hand, CENP-A, as a functional variant histone, is intrinsically more stable in its tetramer form and is therefore less dependent on DNA or other proteins, which may fit better its unique assembly pathway and intricate regulation.

On the basis of our calculations, we speculate that forming the CENP-A tetramer may be nature's way to reduce the availability of individual CENP-A dimers. The stably formed $(CENP\text{-}A/H4)_2$ tetramer may serve as a sequestration channel, needed to maintain CENP-A homeostasis. One logical prediction is that histone modification in CENP-A, especially at the interface, that would either strengthen or weaken the rigidity/compactness of the tetramer might regulate the levels of dimer CENP-A/H4 available for chaperone-mediated assembly to a further extent. On the other hand, we previously found that the disordered terminal tail of CENP-A is very flexible and could easily encounter other proteins (37). The rigidity of the CENP-A tetramer may prevent CENP-A from associating with noncentromeric proteins so as to avoid the ectopic localization or promiscuous interactions that might occur more frequently in cancer cells when CENP-A is overexpressed (32).

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. 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 and disassembly pathways 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 (80).

The CENP-A-specific chaperone HJURP may block CENP-A dimers from self-associating into a tetramer by competing for the same binding site, the internal side of the CENP-A tetramer. It has been shown that two HJURP chaperones (81) and the dimerization of HJURP (29) is required for proper CENP-A nucleosome assembly. Therefore, an HJURP dimer may interact with two isolated CENP-A dimers instead of with a CENP-A tetramer (Fig. 7, *right*). On the contrary, as in the structure of H3 and its chaperone CAF-1 (82,83), CAF-1 binds with an H3 dimer at the external side, without the possibility of preventing it from forming a tetramer. Indeed, the kinetically less stable tetramer of H3 may need the enhanced stabilization via binding with CAF-1 chaperones at either side (Fig. 7, *left*). Taken all together, we propose two different chaperone models for CENP-A and H3 assembly, $CENP\text{-}A/H4\text{-}(HJURP)_2\text{-}CENP\text{-}A/H4$ versus $CAF\text{-}1\text{-}(H3/H4)_2\text{-}CAF\text{-}1$ (Fig. 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 preassembly complexes, whereas in the latter, CAF-1 proteins would stabilize a preformed H3 tetramer in preparation for

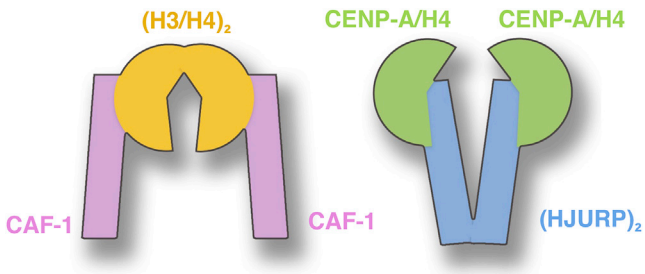

FIGURE 7 Suggested different models for histones and their chaperones during deposition. (*Left*) H3/H4 may be deposited in the form of a tetramer with each external side bracketed by a CAF-1 chaperone, which may stabilize the tetramer. (*Right*) CENP-A may be deposited as dimers; each dimer is loaded by one HJURP chaperone. To see this figure in color, go online.

subsequent nucleosome assembly. Our results support the possibility that canonical H3- and CENP-A-containing nucleosomes may have orthogonal assembly pathways: although H3 could be deposited as a tetramer, CENP-A may be loaded in the form of a dimer.

CONCLUSIONS

This work establishes that variant histone CENP-A thermodynamically favors a tetramer formation, whereas the canonical H3 presents remarkable swiveling dynamics about the tetramer interface contributing a rugged yet shallow binding free energy landscape. Moreover, H2A/H2B dimers restrain the internal rotational 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. These findings provide comprehensive molecular insights into the longstanding and recent experimental observations, offering new, to our knowledge, perspectives on the structural debates over CENP-A dynamics. Based on our results, we suggest two different assembly models for H3 and CENP-A. Lastly, we propose that the $(CENP\text{-}A/H4)_2$ tetramer may serve as a sequestration channel in vivo, which would provide another layer of regulation to ensure the proper homeostasis of CENP-A.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.03.021.

AUTHOR CONTRIBUTIONS

H.Z., Y.D., and G.A.P. designed the research. H.Z. carried out computational simulations. H.Z. and D.W. analyzed the data. All authors wrote the article.

ACKNOWLEDGMENTS

We thank Drs. Aram Davtyan and Bin Zhang for helpful discussions of AWSEM method, Drs. Minh Bui, Ignacia Echeverria, Longhua Hu for critical reading of the manuscript, and Sizhu Li for graphics help.

This work was supported by National Science Foundation Grant CHE-1800418, the intramural research program of Center for Cancer Research/National Cancer Institue, and the National Cancer Institute-University of Maryland Cancer Technology Partnership for Integrative Cancer Research. H.Z. was also supported by the Ann G. Wylie Dissertation Fellowship from University of Maryland, College Park.

REFERENCES

1. Luger, K., A. W. Mäder, …, T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature.* 389:251–260.
2. Bina-Stein, M., and R. T. Simpson. 1977. Specific folding and contraction of DNA by histones H3 and H4. *Cell.* 11:609–618.
3. Sollner-Webb, B., R. D. Camerini-Otero, and G. Felsenfeld. 1976. Chromatin structure as probed by nucleases and proteases: evidence for the central role of histones H3 and H4. *Cell.* 9:179–193.
4. Vlijm, R., M. Lee, …, C. Dekker. 2015. Comparing the assembly and handedness dynamics of (H3.3-H4)2 tetrasomes to canonical tetrasomes. *PLoS One.* 10:e0141267.
5. Vlijm, R., M. Lee, …, N. H. Dekker. 2015. Nucleosome assembly dynamics involve spontaneous fluctuations in the handedness of tetrasomes. *Cell Rep.* 10:216–225.
6. Ordu, O., L. Kremser, …, N. H. Dekker. 2018. Modification of the histone tetramer at the H3-H3 interface impacts tetrasome conformations and dynamics. *J. Chem. Phys.* 148:123323.
7. Henikoff, S., T. Furuyama, and K. Ahmad. 2004. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet.* 20:320–326.
8. Sarma, K., and D. Reinberg. 2005. Histone variants meet their match. *Nat. Rev. Mol. Cell Biol.* 6:139–149.
9. Melters, D. P., J. Nye, …, Y. Dalal. 2015. Chromatin dynamics in vivo: a game of musical chairs. *Genes (Basel).* 6:751–776.
10. Biterge, B., and R. Schneider. 2014. Histone variants: key players of chromatin. *Cell Tissue Res.* 356:457–466.
11. Sathyan, K. M., D. Fachinetti, and D. R. Foltz. 2017. α-amino trimethylation of CENP-A by NRMT is required for full recruitment of the centromere. *Nat. Commun.* 8:14678.
12. Bloom, K., and V. Costanzo. 2017. Centromere structure and function. *In* Centromeres and Kinetochores. Springer, pp. 515–539.
13. Stankovic, A., L. Y. Guo, …, L. E. Jansen. 2016. A dual inhibitory mechanism sufficient to maintain cell-cycle-restricted CENP-A assembly. *Mol. Cell.* 65:231–246, Published online December 22, 2016.
14. Nechemia-Arbely, Y., D. Fachinetti, …, D. W. Cleveland. 2017. Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *J. Cell Biol.* 216:607–621.
15. Wang, K., Z. Yu, …, G. Li. 2017. Ser68 phosphorylation ensures accurate cell-cycle-dependent CENP-A deposition at centromeres. *Dev. Cell.* 40:5–6.
16. Sekulic, N., E. A. Bassett, …, B. E. Black. 2010. The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature.* 467:347–351.
17. Shivaraju, M., J. R. Unruh, …, J. L. Gerton. 2012. Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell.* 150:304–316.
18. Bui, M., E. K. Dimitriadis, …, Y. Dalal. 2012. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell.* 150:317–326.
19. Dalal, Y., H. Wang, …, S. Henikoff. 2007. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS Biol.* 5:e218.
20. Hasson, D., T. Panchenko, …, B. E. Black. 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat. Struct. Mol. Biol.* 20:687–695.
21. Furuyama, T., C. A. Codomo, and S. Henikoff. 2013. Reconstitution of hemisomes on budding yeast centromeric DNA. *Nucleic Acids Res.* 41:5769–5783.
22. Tanaka, Y., H. Tachiwana, …, S. Yokoyama. 2005. Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. *J. Biol. Chem.* 280:41609–41618.
23. Williams, J. S., T. Hayashi, …, P. Russell. 2009. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol. Cell.* 33:287–298.
24. Zhang, W., S. U. Colmenares, and G. H. Karpen. 2012. Assembly of Drosophila centromeric nucleosomes requires CID dimerization. *Mol. Cell.* 45:263–269.
25. Kato, H., J. Jiang, …, Y. Bai. 2013. A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science.* 340:1110–1113.

26. Dechassa, M. L., K. Wyns, …, K. Luger. 2011. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat. Commun.* 2:313.

27. Foltz, D. R., L. E. Jansen, …, D. W. Cleveland. 2009. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell.* 137:472–484.

28. Dunleavy, E. M., D. Roche, …, G. Almouzni-Pettinotti. 2009. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell.* 137:485–497.

29. Zasadzińska, E., M. C. Barnhart-Dailey, …, D. R. Foltz. 2013. Dimerization of the CENP-A assembly factor HJURP is required for centromeric nucleosome deposition. *EMBO J.* 32:2113–2124.

30. Tachiwana, H., S. Müller, …, G. Almouzni. 2015. HJURP involvement in de novo CenH3(CENP-A) and CENP-C recruitment. *Cell Rep.* 11:22–32.

31. Barnhart-Dailey, M. C., P. Trivedi, …, D. R. Foltz. 2017. HJURP interaction with the condensin II complex during G1 promotes CENP-A deposition. *Mol. Biol. Cell.* 28:54–64.

32. Athwal, R. K., M. P. Walkiewicz, …, Y. Dalal. 2015. CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics Chromatin.* 8:2.

33. Lacoste, N., A. Woolfe, …, G. Almouzni. 2014. Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Mol. Cell.* 53:631–644.

34. Black, B. E., D. R. Foltz, …, D. W. Cleveland. 2004. Structural determinants for generating centromeric chromatin. *Nature.* 430:578–582.

35. Tachiwana, H., W. Kagawa, …, H. Kurumizaka. 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature.* 476:232–235.

36. Kurumizaka, H., N. Horikoshi, …, W. Kagawa. 2013. Current progress on structural studies of nucleosomes containing histone H3 variants. *Curr. Opin. Struct. Biol.* 23:109–115.

37. Zhao, H., D. Winogradoff, …, G. A. Papoian. 2016. Promiscuous histone mis-assembly is actively prevented by chaperones. *J. Am. Chem. Soc.* 138:13207–13218.

38. Winogradoff, D., H. Zhao, …, G. A. Papoian. 2015. Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome. *Sci. Rep.* 5:17038.

39. Falk, S. J., L. Y. Guo, …, B. E. Black. 2015. Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science.* 348:699–703.

40. Davey, C. A., D. F. Sargent, …, T. J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.* 319:1097–1113.

41. Wood, C. M., J. M. Nicholson, …, J. P. Baldwin. 2005. High-resolution structure of the native histone octamer. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 61:541–545.

42. Wood, C. M., S. Sodngam, …, J. P. Baldwin. 2006. The oxidised histone octamer does not form a H3 disulphide bond. *Biochim. Biophys. Acta.* 1764:1356–1362.

43. Arents, G., R. W. Burlingame, …, E. N. Moudrianakis. 1991. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. USA.* 88:10148–10152.

44. Tsunaka, Y., Y. Fujiwara, …, K. Morikawa. 2016. Integrated molecular mechanism directing nucleosome reorganization by human FACT. *Genes Dev.* 30:673–686.

45. Chen, S., A. Rufiange, …, D. J. Patel. 2015. Structure-function studies of histone H3/H4 tetramer maintenance during transcription by chaperone Spt2. *Genes Dev.* 29:1326–1340.

46. Saredi, G., H. Huang, …, A. Groth. 2016. H4K20me0 marks post-replicative chromatin and recruits the TONSL–MMS22L DNA repair complex. *Nature.* 534:714–718.

47. Huang, H., C. B. Strømme, …, D. J. Patel. 2015. A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. *Nat. Struct. Mol. Biol.* 22:618–626.

48. Banks, D. D., and L. M. Gloss. 2003. Equilibrium folding of the core histones: the H3-H4 tetramer is less stable than the H2A-H2B dimer. *Biochemistry.* 42:6827–6839.

49. Banks, D. D., and L. M. Gloss. 2004. Folding mechanism of the (H3-H4)2 histone tetramer of the core nucleosome. *Protein Sci.* 13:1304–1316.

50. Baxevanis, A. D., J. E. Godfrey, and E. N. Moudrianakis. 1991. Associative behavior of the histone (H3-H4)2 tetramer: dependence on ionic environment. *Biochemistry.* 30:8817–8823.

51. Bowman, A., R. Ward, …, D. G. Norman. 2010. Probing the (H3-H4)2 histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labelling. *Nucleic Acids Res.* 38:695–707.

52. Bui, M., M. Pitman, …, Y. Dalal. 2017. Internal modifications in the CENP-A nucleosome modulate centromeric dynamics. *Epigenetics Chromatin.* 10:17.

53. Wolynes, P. G., and G. A. Papoian. 2017. AWSEM-MD: from neural networks to protein structure prediction and functional dynamics of complex biomolecular assemblies. *In* Coarse-Grained Modeling of Biomolecules. CRC Press, pp. 121–190.

54. Davtyan, A., N. P. Schafer, …, G. A. Papoian. 2012. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B.* 116:8494–8503.

55. Zhang, B., W. Zheng, …, P. G. Wolynes. 2016. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* 138:8126–8133.

56. Kumar, S., J. M. Rosenberg, …, P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.

57. Boresch, S., and M. Karplus. 1996. The Jacobian factor in free energy simulations. *J. Chem. Phys.* 105:5145–5154.

58. Takada, S., R. Kanada, …, H. Kenzaki. 2015. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* 48:3026–3035.

59. Winogradoff, D., I. Echeverria, …, G. A. Papoian. 2015. The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects. *J. Am. Chem. Soc.* 137:6245–6253.

60. Materese, C. K., A. Savelyev, and G. A. Papoian. 2009. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J. Am. Chem. Soc.* 131:15005–15013.

61. Armstrong, R. L., T. J. R. Penke, …, R. J. Duronio. 2019. H3K9 promotes under-replication of pericentromeric heterochromatin in drosophila salivary gland polytene chromosomes. *Genes (Basel).* 10:E93.

62. Tencer, A. H., K. L. Cox, …, T. G. Kutateladze. 2017. Covalent modifications of histone H3K9 promote binding of CHD3. *Cell Rep.* 21:455–466.

63. Bhat, S., Y. Hwang, …, P. A. Cole. 2018. Hydrazide mimics for protein lysine acylation to assess nucleosome dynamics and deubiquitinase action. *J. Am. Chem. Soc.* 140:9478–9485.

64. Gatchalian, J., X. Wang, …, T. G. Kutateladze. 2017. Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat. Commun.* 8:1489.

65. Zhang, B., and T. Parsons. 2018. Critical role of histone tail entropy in nucleosome unwinding. *bioRxiv* https://doi.org/10.1101/494328.

66. Collepardo-Guevara, R., G. Portella, …, M. Orozco. 2015. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Am. Chem. Soc.* 137:10205–10215.

67. Arya, G., and T. Schlick. 2006. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. USA.* 103:16236–16241.

68. Potoyan, D. A., and G. A. Papoian. 2011. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* 133:7405–7415.

69. Potoyan, D. A., and G. A. Papoian. 2012. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl. Acad. Sci. USA*. 109:17857–17862.

70. Todolli, S., P. J. Perez, …, W. K. Olson. 2017. Contributions of sequence to the higher-order structures of DNA. *Biophys. J.* 112:416–426.

71. Bascom, G. D., and T. Schlick. 2018. Chromatin fiber folding directed by cooperative histone tail acetylation and linker histone binding. *Biophys. J.* 114:2376–2385.

72. Li, G., M. Levitus, …, J. Widom. 2005. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* 12:46–53.

73. Winogradoff, D., and A. Aksimentiev. 2019. Molecular mechanism of spontaneous nucleosome unraveling. *J. Mol. Biol.* 431:323–335.

74. Baldi, S., D. S. Jain, …, P. B. Becker. 2018. Genome-wide rules of nucleosome phasing in drosophila. *Mol. Cell.* 72:661–672.e4.

75. Struhl, K., and E. Segal. 2013. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20:267–273.

76. Dong, F., and K. E. van Holde. 1991. Nucleosome positioning is determined by the (H3-H4)2 tetramer. *Proc. Natl. Acad. Sci. USA*. 88:10596–10600.

77. Krietenstein, N., C. J. Wippo, …, P. Korber. 2012. Genome-wide in vitro reconstitution of yeast chromatin with in vivo-like nucleosome positioning. *Methods Enzymol.* 513:205–232.

78. Clark, D. J. 2010. Nucleosome positioning, nucleosome spacing and the nucleosome code. *J. Biomol. Struct. Dyn.* 27:781–793.

79. Vlijm, R., S. H. Kim, …, C. Dekker. 2017. The supercoiling state of DNA determines the handedness of both H3 and CENP-A nucleosomes. *Nanoscale.* 9:1862–1870.

80. Bassett, E. A., J. DeNizio, …, B. E. Black. 2012. HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. *Dev. Cell.* 22:749–762.

81. Bodor, D. L., and L. E. Jansen. 2013. How two become one: HJURP dimerization drives CENP-A assembly. *EMBO J.* 32:2090–2092.

82. Liu, W. H., S. C. Roemer, …, M. E. Churchill. 2016. The Cac1 subunit of histone chaperone CAF-1 organizes CAF-1-H3/H4 architecture and tetramerizes histones. *eLife*. 5:e18023.

83. Mattiroli, F., Y. Gu, …, K. Luger. 2017. DNA-mediated association of two histone-bound complexes of yeast chromatin assembly factor-1 (CAF-1) drives tetrasome assembly in the wake of DNA replication. *eLife*. 6:e22799.

Biophysical Journal, Volume *116*

Supplemental Information

The Oligomerization Landscape of Histones

Haiqing Zhao, David Winogradoff, Yamini Dalal, and Garegin A. Papoian

1 AWSEM Model Details

The Associative-memory, Water-mediated, Structure and Energy Model (AWSEM) was developed based on decades of efforts. History of related models can be referred to chapter 3 of a recent book by papoian *et al.*[1] It has been successfully applied to study protein folding,[2] binding,[3,4] aggregation,[5–7] intrinsically-disordered proteins,[8] membrane proteins,[9,10] and protein-DNA association.[11–13] Basically it is a transferable coarse-grained protein force field based on the energy landscape theory.

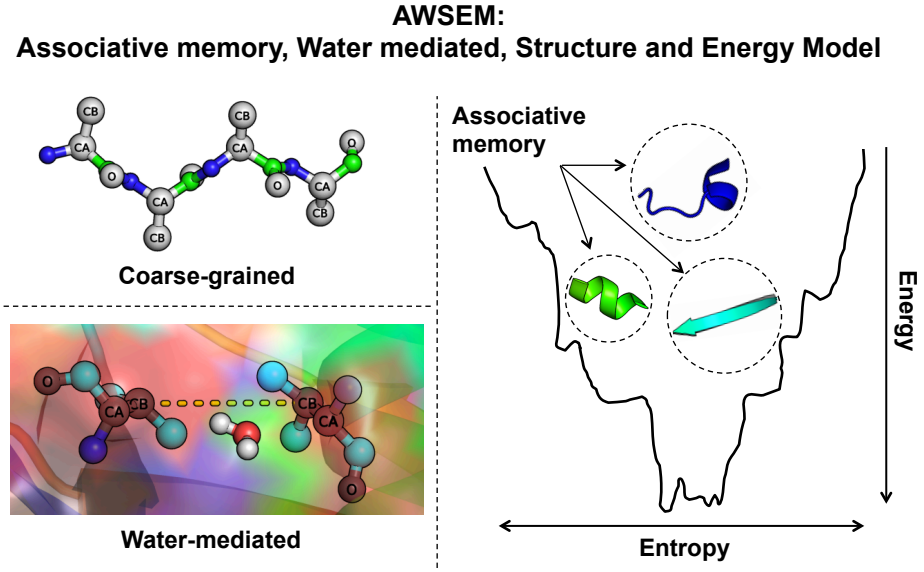

Figure S1: **A cartoon diagram of the Associative memory, Water mediated, Structure and Energy Model (AWSEM)** This diagram shows three main features of this model. Firstly, it uses three beads (C_α, C_β, O) to represent one amino acid. Residue information such as charge or radius is carried by C_β bead. Secondly, it features a mediated potential for water-residue interactions. Thirdly, it contains a local structure based biasing potential term, aligning the target sequence to short peptides with known structures in the pdb database. Figure adopted from ref. 14

Here, we used the AWSEM model as the force field to perform molecular dynamics simulations. In AWSEM, both physically-motivated potential terms, such as the 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 which are based on the local density of residues and their distances in between each. 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. Detailed function form for each potential term can be found in the SI of ref. 2.

$$V_{AWSEM} = V_{backbone} + V_{contact} + V_{burial} + V_{HB} + V_{AM} \tag{S1}$$

Parameters for this model were optimized self-consistently using energy landscape theory, basically maximizing the ratio of the folding temperature to the glass transition temperature T_f/T_g.[15] In the current simulations, single memories, typically 12-residue long, were set, exclusively from the histone monomers found in the corresponding nucleosomal crystal structures. Comparisons and analyses regarding different crystal resources are provided in the following section 6 (Figure S4). The weight of memory term V_{AM} is 0.2. Weights for $V_{contact}$ and V_{burial} are both 1.0. The overall potential weight ϵ is 1.0.

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.

2 Trajectory Analysis Details

We use the coordinates of Cα to calculate the root-mean-square deviations (RMSD), radius-of-gyration (R_g), distances (R), dihedral angles θ, Q values, and contact analysis. 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 $\alpha 2$ 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 was 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. Singular value decomposition (SVD)[16] 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:

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

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.

3 Error Analysis and P-value Calculation

Error analysis for the free energy profiles (FEPs) consisted of two parts (Eq. S2). The first part is more related to 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. S2, 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.[17,18] 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.

$$Error(FEP) = \sqrt{\frac{\sum_{i=1}^{N}(F_i - F_0)^2}{N-1}} + Error_{(WHAM)} \tag{S2}$$

We used the p-value from a t-test[19] 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 numbers of the data in each distribution. The same method was used for probability density function figures of the main text and SI.

4 Simulation Convergence

We calculated the root-mean-squared inner-product (RMSIP)[20] 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.

$$\text{RMSIP} = \left(\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}(\boldsymbol{\eta_i}\cdot\boldsymbol{\nu_j})^2\right)^{1/2} \tag{S3}$$

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.

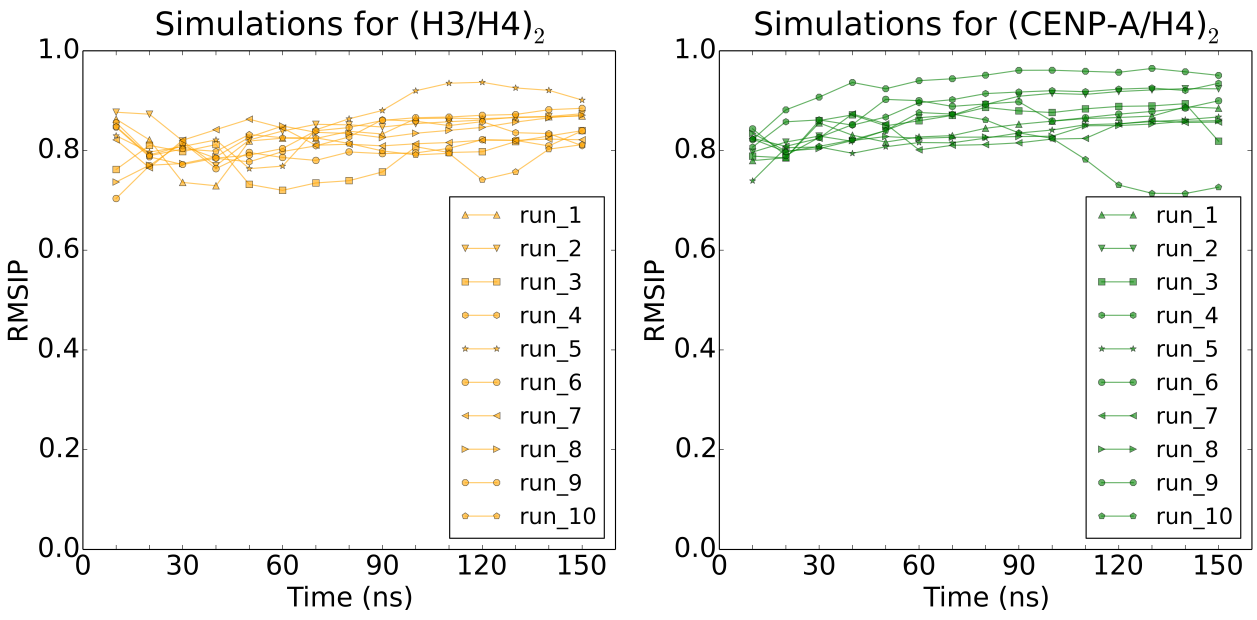

Figure S2: **RMSIP analysis shows the convergence of CG-AWSEM simulations.** RMSIP are calculated for every independent simulation of both H3/H4 and CENP-A/H4 tetramer. All calculated RMSIP are over 0.7, indicating adequate convergences.

5 Umbrella Sampling Histograms

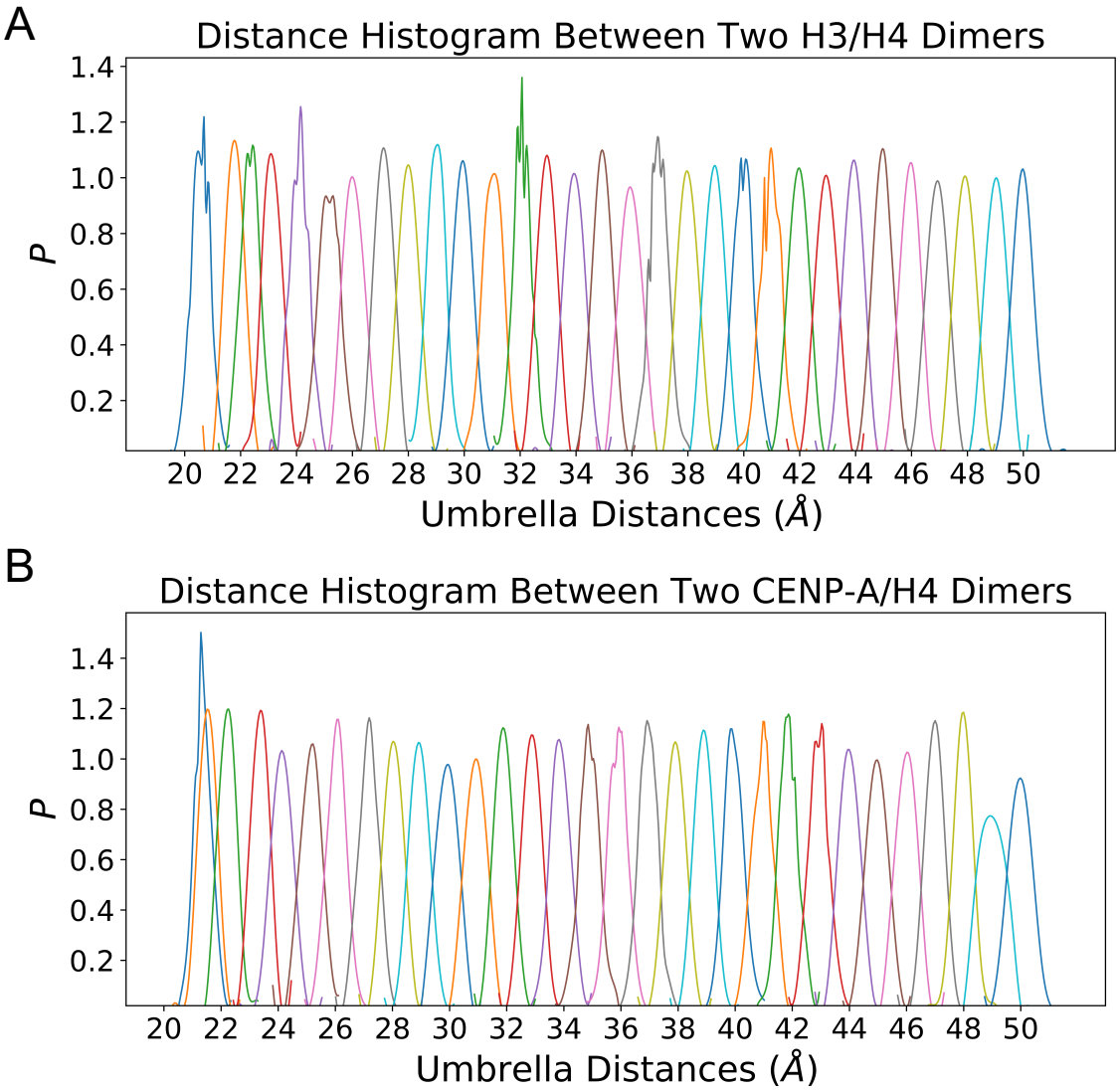

Figure S3: **Sufficient overlaps of reaction coordinate R between adjacent umbrella windows ensure the convergence of WHAM[17] calculation.** Distances from all umbrella windows at replica 300K are collected and histogramed for (A) two H3/H4 dimers and (B) two CENP-A/H4 dimers. PMFs were then calculated from these data using WHAM.

6 Crystal Structure Alignments

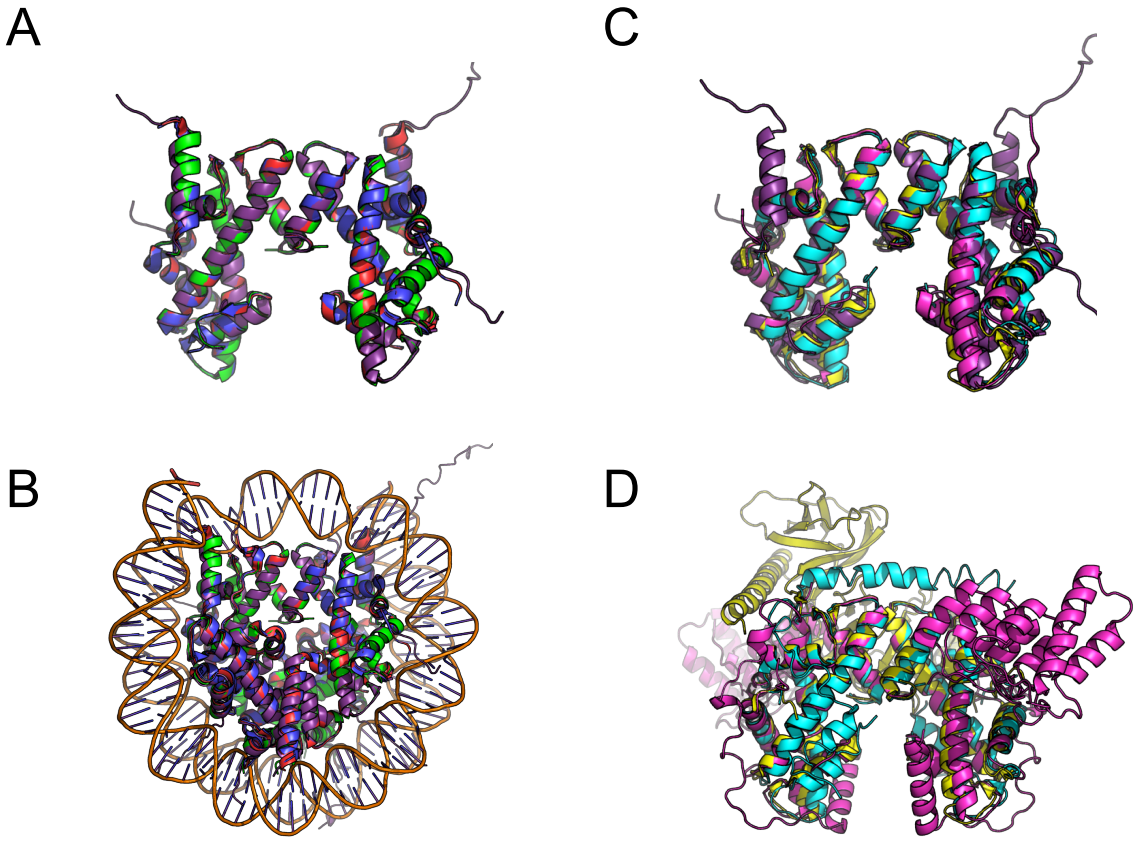

Figure S4: **$(H3/H4)_2$ tetramer presents virtually superposable conformations in the crystal structures of histone octamer, nucleosome and protein complex with chaperone protein(s).** (A) Tetramers from histone octamers and nucleosome are aligned. The average pairwise RMSD is 0.4. Red marks PDB structure 2ARO, blue 2ARO, green 2HIO, and purple is for tetramer from nucleosome structure 1KX5. (B) Structural alignments based on tetramer but shown in the original octamer and nucleosome contexts. Color schemes are the same as in (A). (C) Alignments of histone tetramer from protein complexes that involve tetramer and other proteins. Yellow is for tetramer with chaperone FACT (pdb: 4Z2M), cyan for tetramer with Spt2 (pdb: 5BSA), pink for tetramer with TONSL and MCM2 (pdb: 5JA4), purple refers to histone tetramer from nucleosome (pdb: 1KX5). (D) Same alignments/color schemes as in (C) are shown in their original protein complex environments.

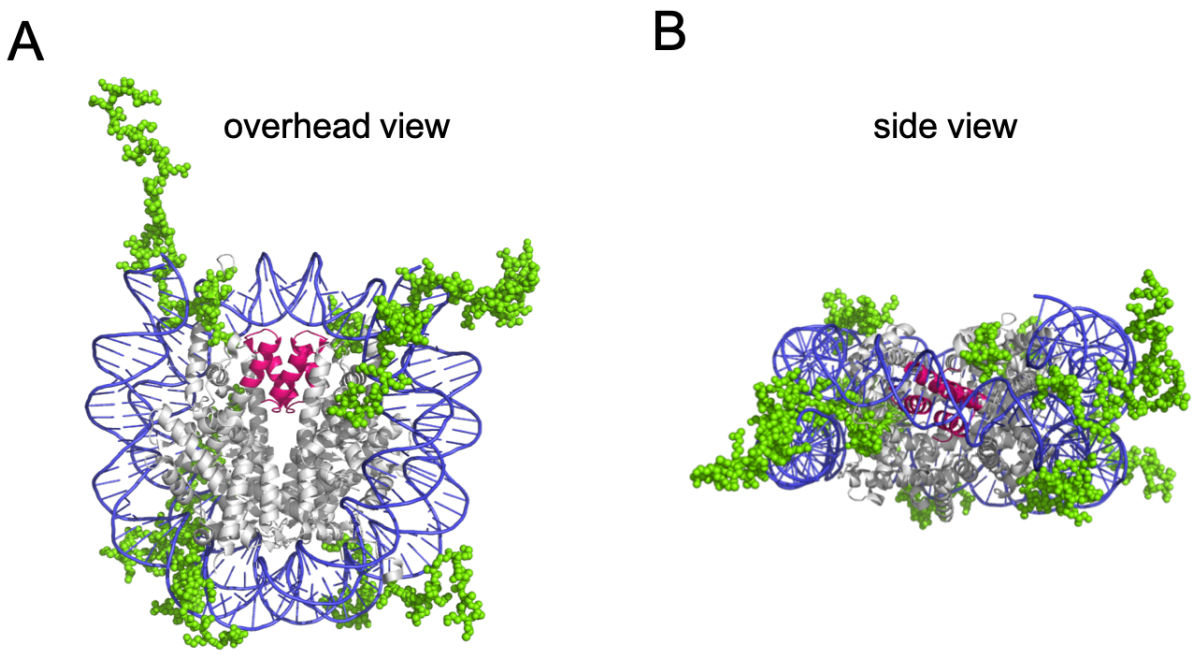

Figure S5: **Crystal structure of nucleosome highlights the position of histone tails and DNA.** Histone tails (green) generally extend out of the histone protein core (grey), closely interacting with DNA (blue) and other potential nucleosome binding partners. The tetramer interface of the histone protein core, four-helix bundle, are marked as red. Shown here is an example of canonical nucleosome structure (PDB: 1KX5), from the overhead view (A) and side view (B).

7 Amino Acid Sequences

H3 (seq # 46-134):

VALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLCAIHAKRVTIMPKDIQLARRIRGER

H4 (seq # 25-112):

NIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

H3' (seq # 48-135):

LREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

H4' (seq # 24-112):

DNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

CENP-A (seq # 46-134):

GWLKEIRKLQKSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRG

H4 (seq # 25-112):

NIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

CENP-A' (seq # 48-135):

LKEIRKLQKSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRGL

H4' (seq # 24-112):

DNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

Mark code: Four-helix Bundle H3 or CENP-A α2 helix

Figure S6: **The sequence number and sequence alignments for the histone proteins investigated in this study.** 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. Sequences of the four-helix bundle region are marked in red and sequences of the $\alpha 2$ helix are underlined.

8 Extended 2D/1D Free Energy Profiles from Enhanced Samplings

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 S9) 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.

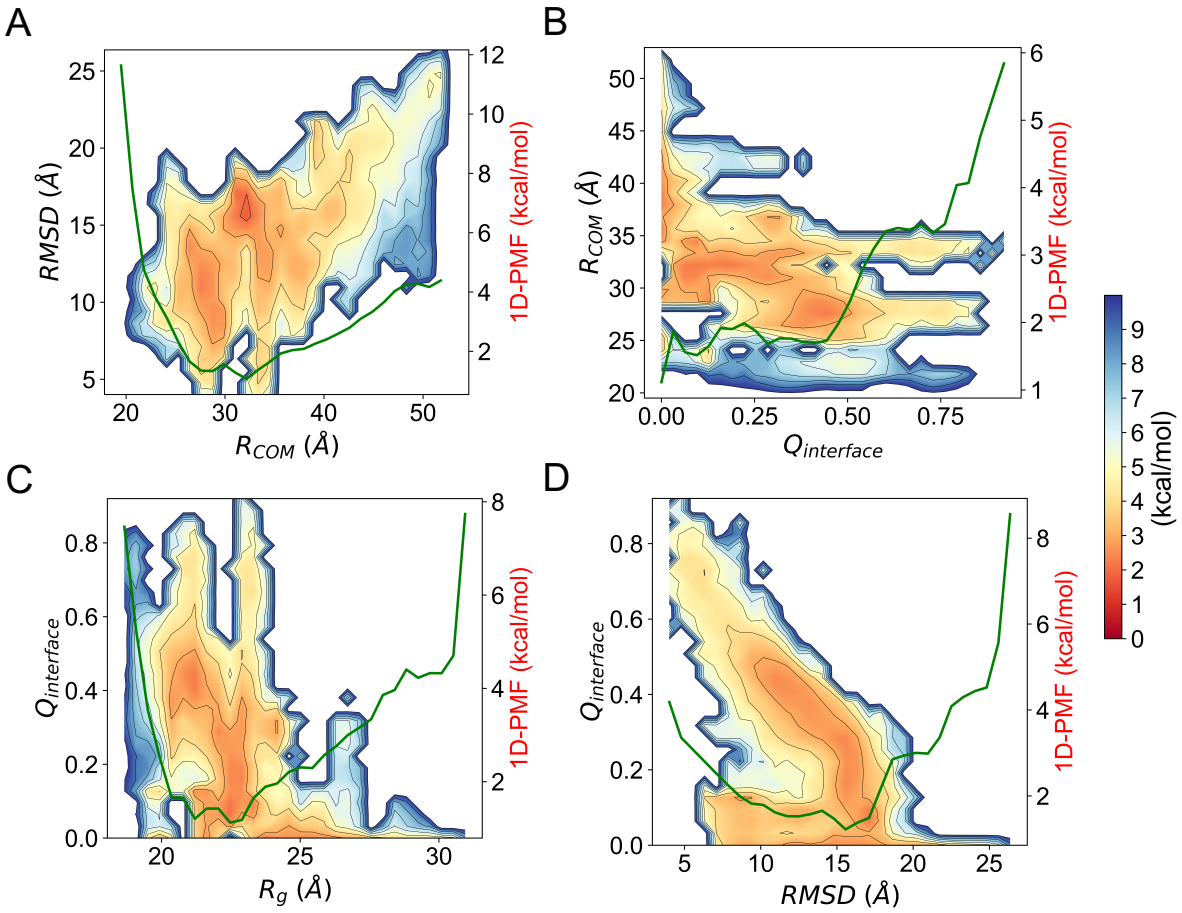

Figure S7: **Extended 2D and 1D free energy profiles for histone $(H3/H4)_2$.** Free energy profiles calculated from the enhanced sampling for $(H3/H4)_2$ tetramer are projected on the 2D reaction coordinates of the distance between centers of masses of each dimer R_{COM} and the measure of overall structural 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.

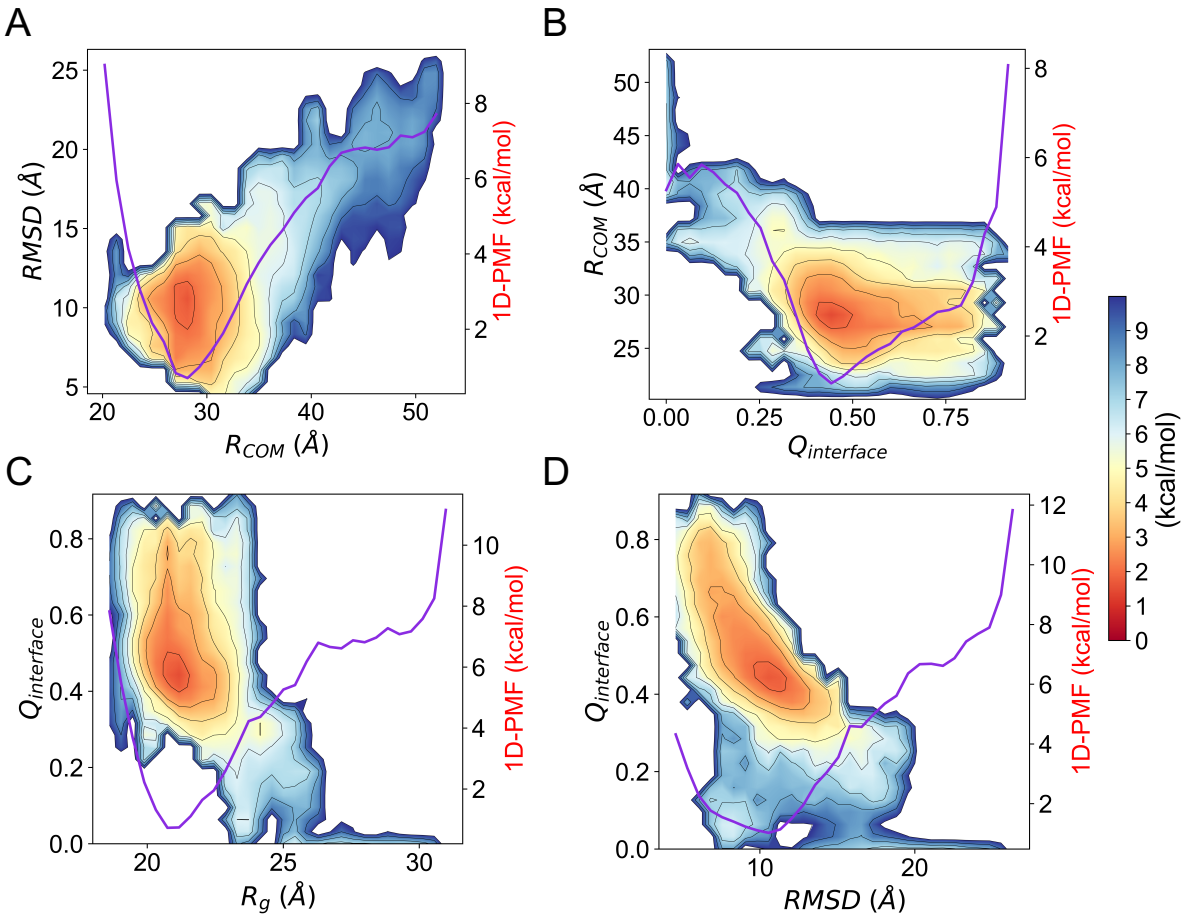

Figure S8: **Extended 2D and 1D free energy profiles for $(CENP\text{-}A/H4)_2$ histone tetramer.** Free energy profiles calculated from the enhanced sampling for CENP-A tetramer are projected on the 2D reaction coordinates of the distance between centers of masses of each dimer R_{COM} and the measure 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.

9 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 S9), demonstrating the convergence of both methods.

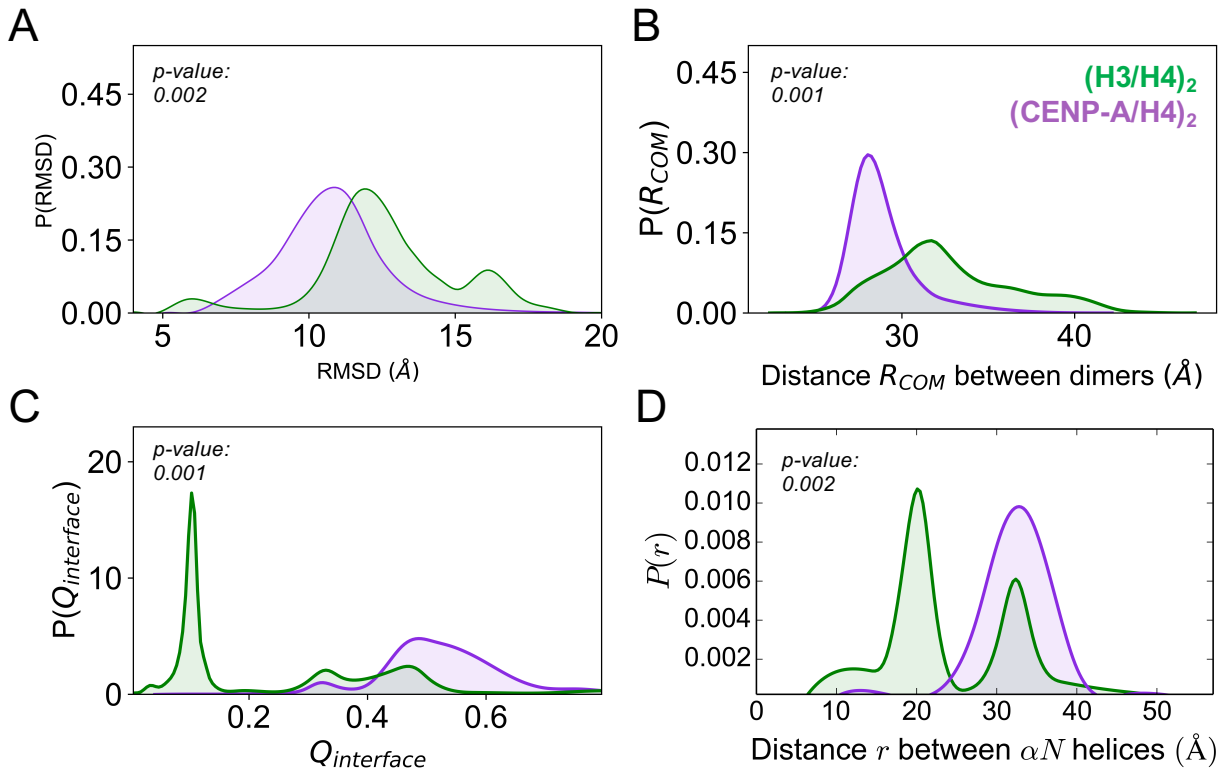

Figure S9: **Distributions of different structural measures confirm the conformational heterogeneity of the H3 tetramer (green) and the homogeneity of the CENP-A tetramer (purple).** (A) The RMSD distribution features multiple populations for the H3 tetramer and a single population for the CENP-A tetramer. (B) The distance between dimers R_{COM} is shorter on average for $(\text{CENP-A/H4})_2$, with a much narrower distribution, than that of $(\text{H3/H4})_2$. (C) $Q_{interface}$ distributions indicate that the interface of $(\text{CENP-A/H4})_2$ remains more stable and closer to its native state than $(\text{H3/H4})_2$. Locations of the peaks in these panels agree with the minima locations in free energy profiles calculated from the enhanced sampling simulations. (D) αN helices of H3 exhibits large heterogeneity while CENP-A αN helices stay stable. The distance measurement between H3 αN helices has two peaks (green) indicating their instability and relocation while the same measurement for CENP-A has one single peak (purple) implying a stable behavior.

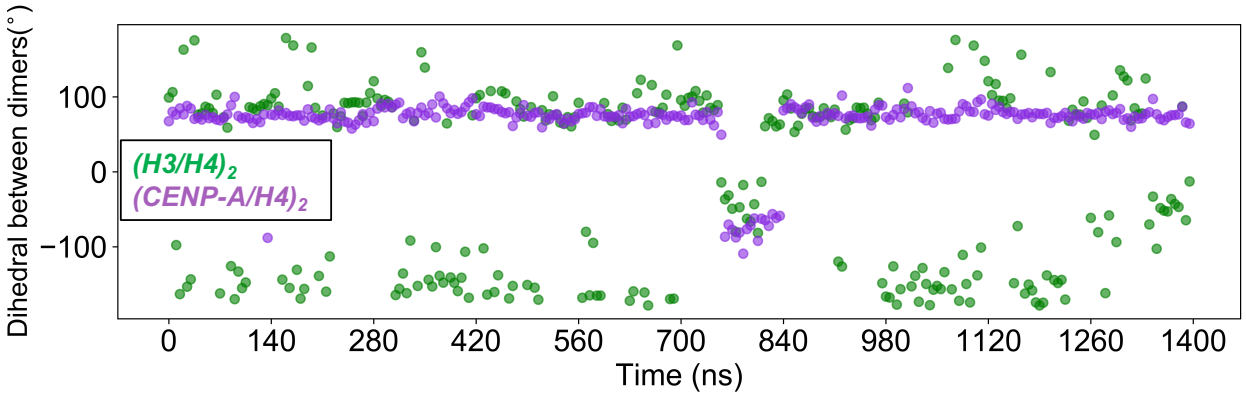

Figure S10: **The dihedral angle between $\alpha 2$ helices measured as a function of time emphasizes the rotational dynamics of the H3 tetramer.** The tetramer dihedral angle of H3 (green) frequently transitions between 90°, −150°, and −50°, while the dihedral of CENP-A (purple) remains constant throughout most of simulation, with only one dihedral angle transition observed.

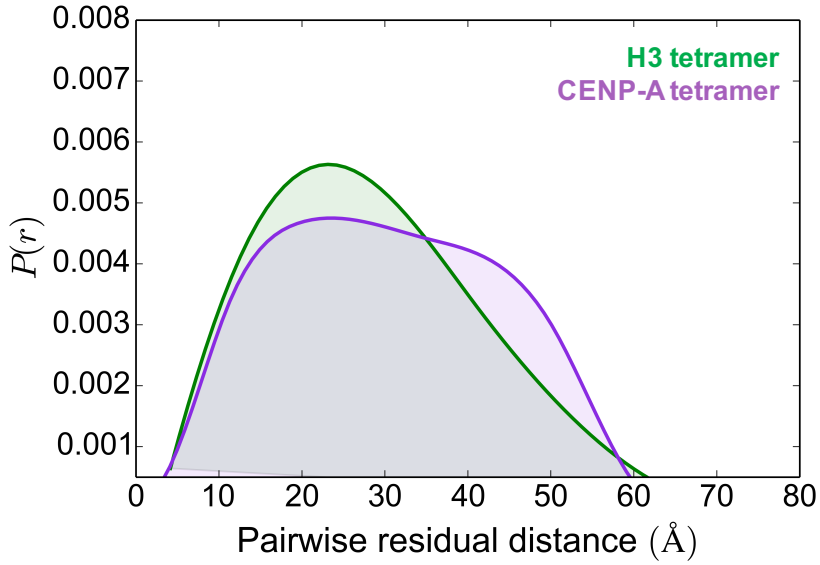

Figure S11: **The distribution of pairwise distance of C_α in representative structure of H3 and CENP-A tetramer.**

10 Principle Component Analysis in Constant T Simulations

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 were projected onto the first two principal components to illustrate the corresponding free energies. The result (Figure S12) is qualitatively consistent with the free energy profile computed from the enhanced the sampling (Figure 1, Figures S7&S8).

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 $\mathbf{Q}$,

$$q_i = (x_1, y_1, z_1, ..., x_n, y_n, z_n), \mathbf{Q} = (q(t_1), q(t_2), ..., q(t_N)) \tag{S4}$$

$$Q_{ij} = q_i(t_j) \tag{S5}$$

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. S6. We then diagonalize the covariance matrix $\mathbf{C}$,

$$C_{j,k} = \frac{1}{N-1} \sum_{i=1}^{N} (Q_{ji} - \langle Q_j \rangle)(Q_{ki} - \langle Q_k \rangle). \tag{S6}$$

$$\mathbf{CM} = \mathbf{M}\Lambda \tag{S7}$$

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:

$$\nu_1 = \mathbf{Q} \cdot m_1, \qquad \nu_2 = \mathbf{Q} \cdot m_2. \tag{S8}$$

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:

$$F(\nu_1, \nu_2) = -k_B T \ln P(\nu_1, \nu_2) - F_{\min}. \tag{S9}$$

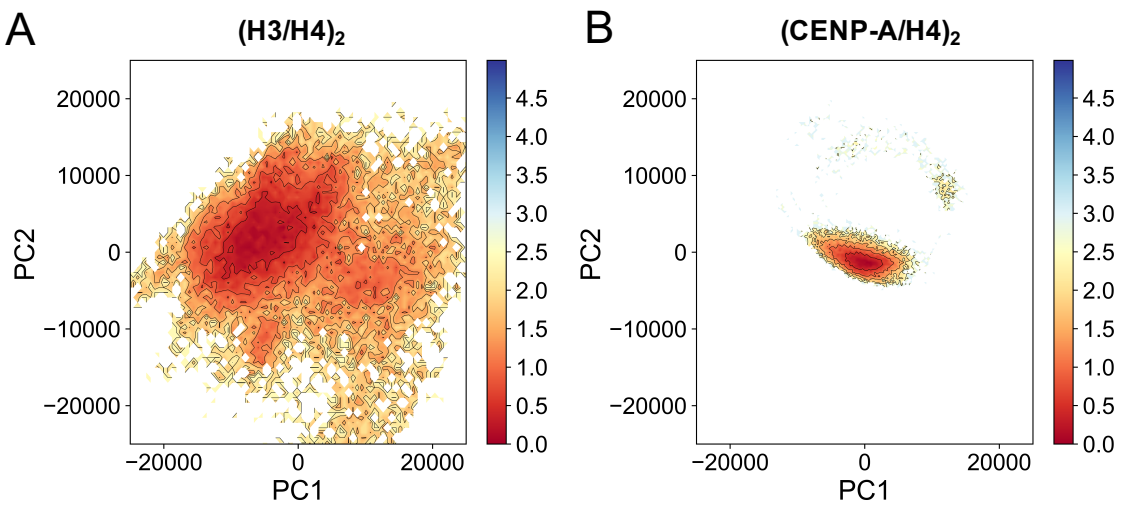

Figure S12: **Free energy projections along the first two PCA components show that $(H3/H4)_2$ occupies a more rugged free energy landscape than (CENP-A/H4)$_2$.** (A) Free energy projection of $(H3/H4)_2$ reveals a broad landscape with multiple conformations basins. (B) Free energy projection of (CENP-A/H4)$_2$ has only one single and deep basin.

11 Simulations of Tetramers Excluding αN Helices

Here we performed similar simulations for H3 and CENP-A tetramers excluding αN helices (Figure S13A). It shows that the CENP-A tetramer has a more closely interacting four-helix bundle (Figure S13B), a more native-like interface (Figure S13C), and a more compact formation (Figure S13D), than the H3 tetramer, even when αN helices are excluded, consistent with the SAXS experimental results measured in solution using the same length of proteins.[21] This set of simulations also confirmed our hypothesis that αN is playing a dominant role in leading the swiveling motion around the interface. It can be seen the swiveling between different tetrameric dihedral angles became much less when αN helices were removed (Figure S13D), compared to when αN helices were included (Figure 4A).

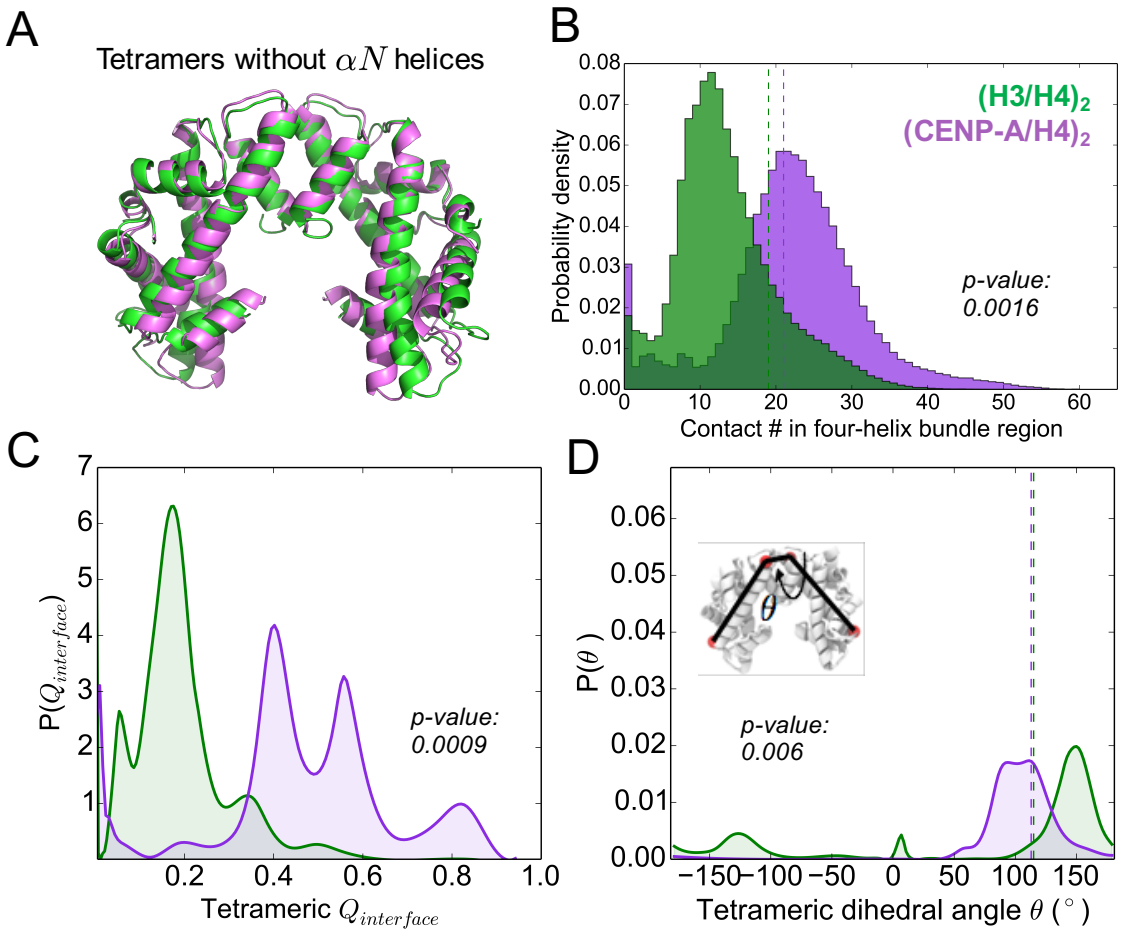

Figure S13: **Tetramer Simulations Excluding αN Helices.** (A) Initial conformations are from αN-helix-truncated CENP-A tetramer crystal structure (PDB: 3NQJ, purple) and the same length of H3 tetramer from the nucleosome (PDB: 1KX5, green). (B) More four-helix contacts are found in CENP-A tetramer. The dash lines mark the contacts number in PDB structures. (C) CENP-A has a larger tetrameric $Q_{interface}$, on average than that of H3, indicating a more native-like tetramer interface. The vertical dashed lines are the tetrameric dihedral angle measurements from the nucleosome crystal structures. (D) Tetrameric dihedral angle measurements for CENP-A and H3 tetramer display that CENP-A has a smaller dihedral angle than H3, implying a more compact conformation.

12 Extended histone octamer simulations results

Histone octamer simulations start from the octamer formations in the H3 and CENP-A nucleosomes (Figure S14A). Extended octamer simulations analyses show that, (H2A/H2B)s maintained well native-like conformations (Figure S14D) but the interactions between (H2A/H2B)s and the CENP-A tetramer appear to fluctuate more than H3 (Figure S14C).

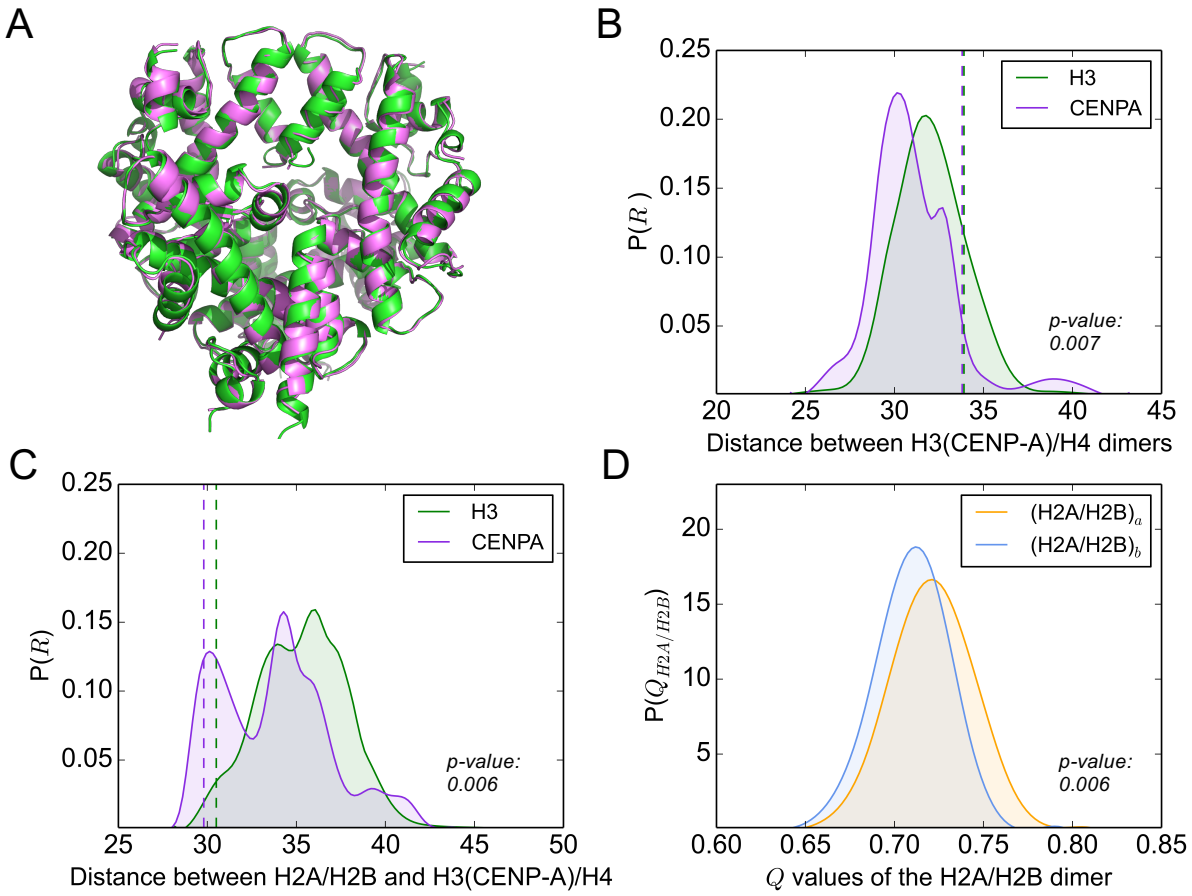

Figure S14: **Simulation results for H3 and CENP-A octamer.** (A) Initial configurations are respectively from the H3- and CENP-A-containing nucleosome structures (PDB: 1KX5 and 3AN2). (B) R distribution shows a single-peak distribution with less deviation for H3 tetramer while CENP-A features one peak and one shoulder in the same distribution. (C)The average distance between (H2A/H2B)s and CENP-A tetramer (purple) is more widely distributed than that of H3 (green). Vertical dashed lines in (B,C) panels are the corresponding distance measurements from the nucleosome crystal structures. (D) Q calculations of (H2A/H2B)s confirm that throughout the octamer simulations, (H2A/H2B)s maintained native-like formations as in their nucleosome structures as both the dimers has high Q values.

13 Analyses on the Level of Dimers and Monomers

To compare the results with our previously-published independent dimer study,[4] 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 the H3 dimer, and the H4 monomer is more native-like than its binding partner H3 or CENP-A.

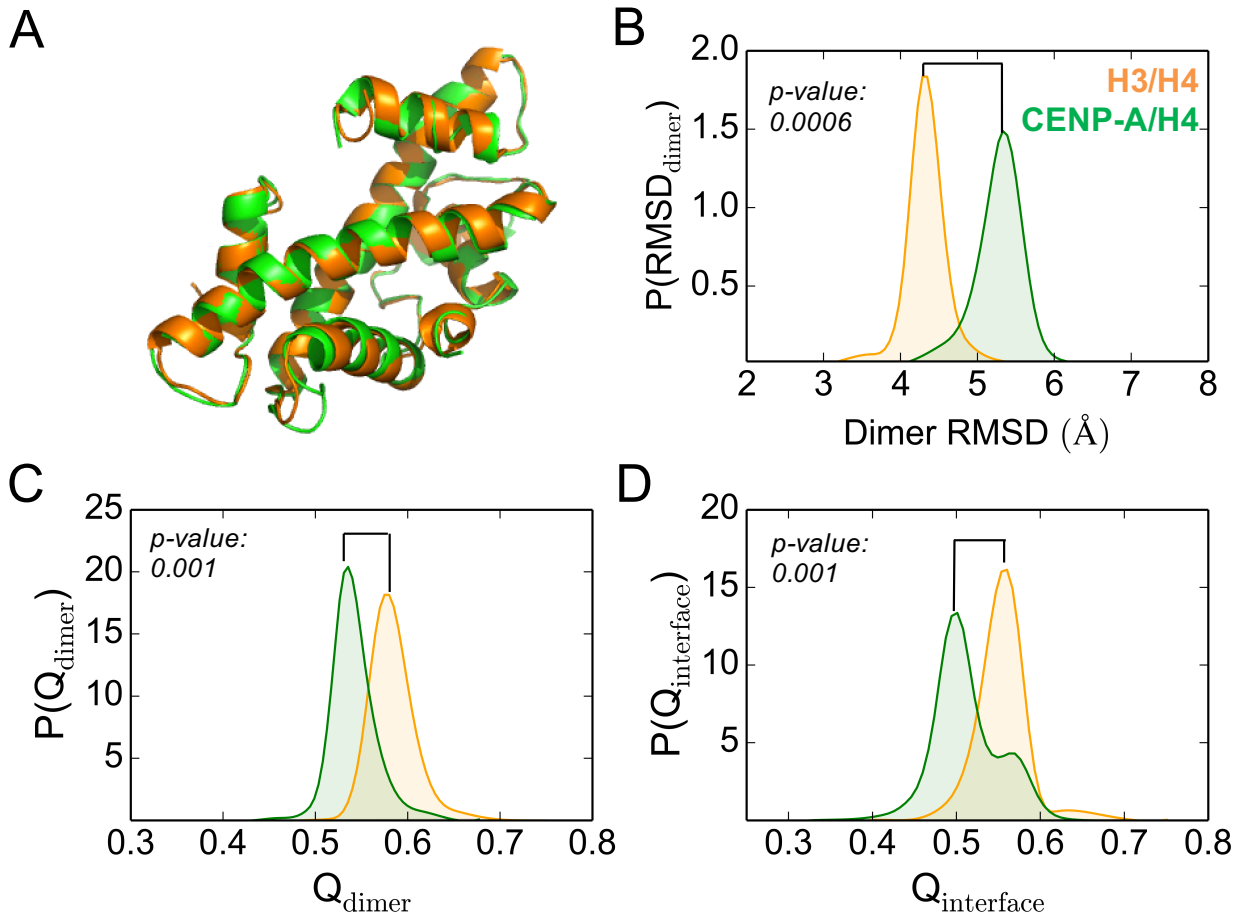

Figure S15: **The H3/H4 dimer is structurally more stable than the CENP-A/H4 dimer in tetramer simulations.** Q_{dimer} and $Q_{interface}$ of the dimer, characterize a dimer's overall structural resemblance or the resemblance of the monomeric interface to its native state (A) respectively. Analyses 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[22,23] and computational studies.[4,24]

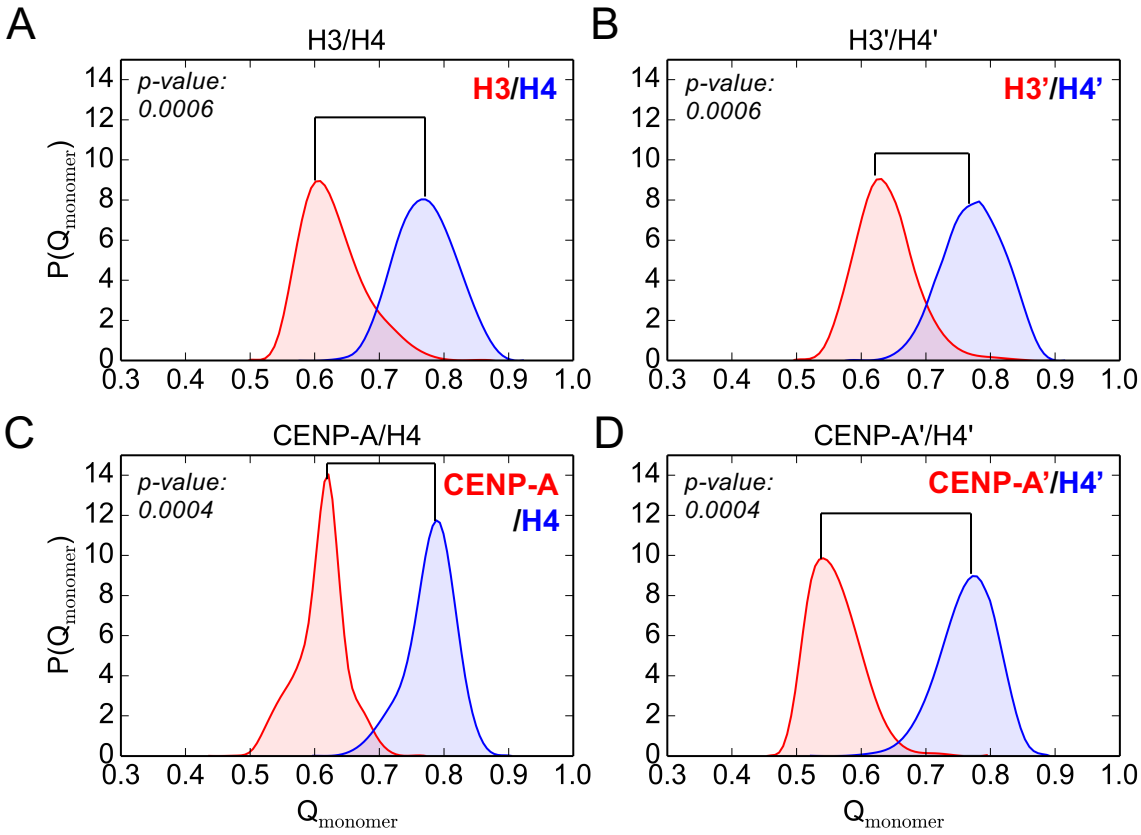

Figure S16: **The H4 monomer maintains a more native-like conformation than its binding partner, either H3 or CENP-A, in all tetramer AWSEM simulations.** $Q_{monomer}$ describes a histone monomer's overall structural similarity with respect to the crystal structures of the corresponding H3 nucleosome (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.[4]

14 AWSEM Energy Analysis of the Tetramer Interface

For a better understanding of the histone tetramer interface dynamics at the 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 tables S1 and S2. The cutoff was chosen as 0.65 kcal/mol for the absolute value of E_{pair}.

The shown two tables (tables 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 energies included 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' has more interacting residue pairs at the binding interface. Detailed residue positions are shown in the structure figures (Figure S17 for H3 and S18 for CENP-A).

Table S1: **Key residue interactions of the tetramer interface in AWSEM (representative conformations)**

H3	H3'	E_{pair} (kcal/mol)
A111	A111	-0.9720
N108	A111	-0.8886
R116	A111	-0.6897
A111	R116	-0.6885

CENP-A	CENP-A'	E_{pair} (kcal/mol)
Y110	L111	-1.0000
L111	T113	-1.0000
L111	R131	-1.0000
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 S2: **Key residual interactions of the tetramer interface in AWSEM (initial conformations)**

H3	H3'	E_{pair} (kcal/mol)
Q125	R129	-1.0000
R129	Q125	-1.0000
L126	R129	-0.9999
R129	L126	-0.9999
Q125	Q125	-0.9918
N108	R129	-0.9832
Q125	A111	-0.9549
A111	Q125	-0.9263
N108	Q125	-0.8416
H113	H133	-0.7130

CENP-A	CENP-A'	E_{pair} (kcal/mol)
Y110	R131	-1.0000
L111	R131	-1.0000
Q127	L111	-1.0000
Q127	R131	-1.0000
R131	L111	-0.9999
R131	Q127	-0.9994
L111	L111	-0.9991
L111	Q127	-0.9989
L128	R131	-0.9878
Q127	Y110	-0.9747
R131	R131	-0.9349
Y110	Q127	-0.8929
L111	Y110	-0.8346
R131	Y110	-0.6405

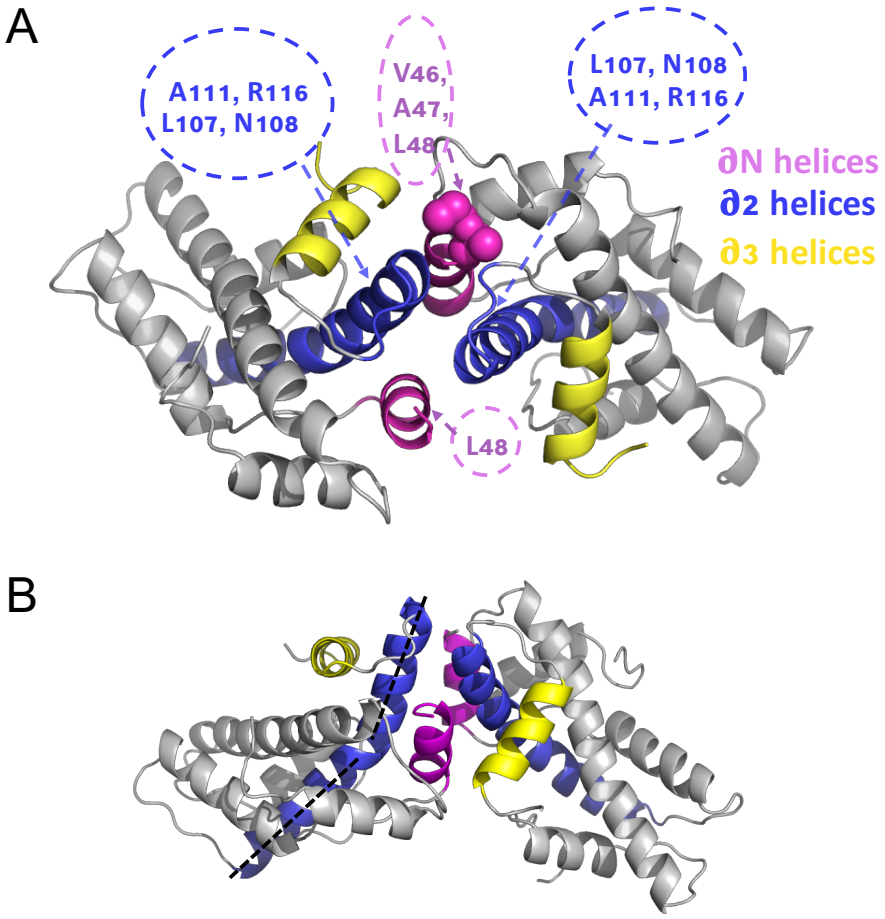

Figure S17: **Representative structure of H3 tetramer with interface interaction details.** (A) Top view of the structure highlights $(H3/H4)_2$ interface is a disrupted *four-helix* bundle region. αN helix competes with the $\alpha 3$ helix, to interact with $\alpha 2$ helix, forming hydrophobic interactions between αN V46, A47, L48 and $\alpha 2$ A111, L107, among which V46 and A47 are H3-specific residues, shown in coarse-grained spheres.(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.

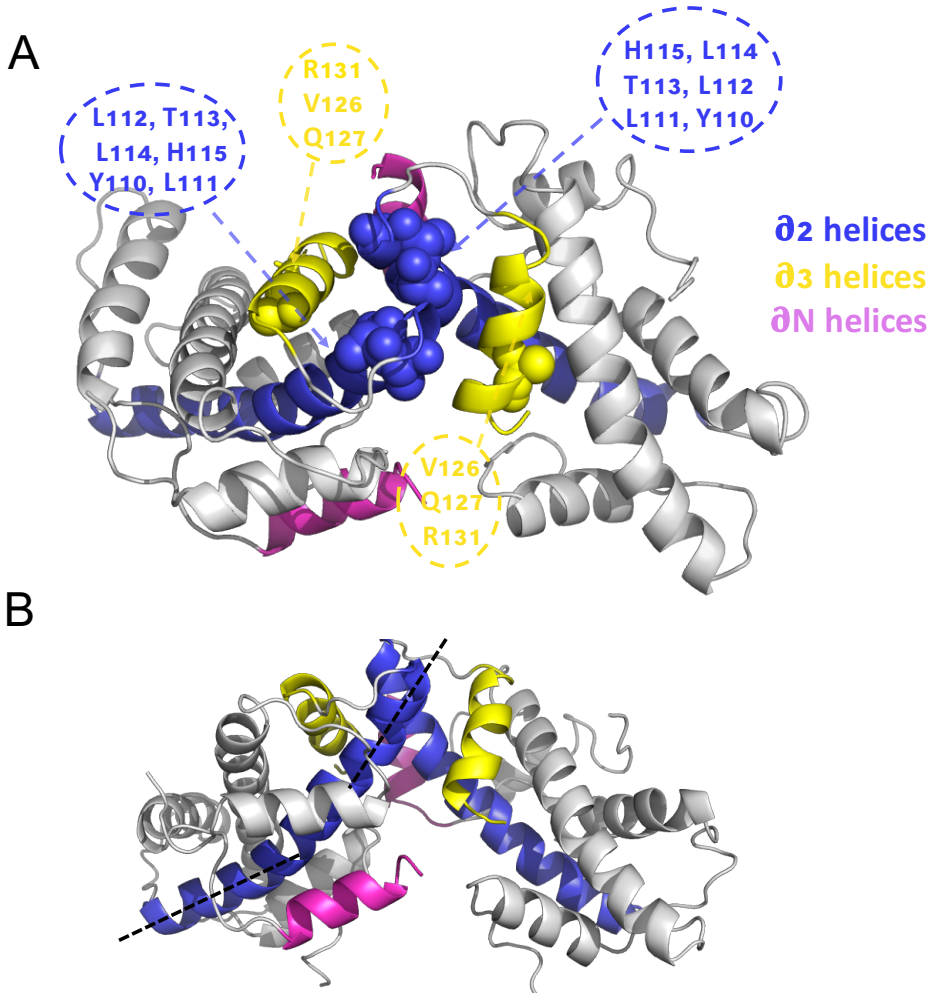

Figure S18: **Representative structure of CENP-A tetramer with interface interaction details.** (A) Top view of the interface highlights a well-formed *four-helix* bundle region, consisting of CENP-A's $\alpha 2$ and $\alpha 3$ helices. CENP-A-specific residues L112, T113, L114, V126 are shown in coarse-grained spheres. (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.

References

(1) Wolynes, P. G.; Papoian, G. A. *Coarse-Grained Modeling of Biomolecules*; CRC Press, 2017; pp 121–190.

(2) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.

(3) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 19244–9.

(4) Zhao, H.; Winogradoff, D.; Bui, M.; Dalal, Y.; Papoian, G. A. Promiscuous histone mis-assembly is actively prevented by chaperones. *Journal of the American Chemical Society* **2016**, *138*, 13207–13218.

(5) Zheng, W.; Schafer, N. P.; Wolynes, P. G. Free energy landscapes for initiation and branching of protein aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 20515–20.

(6) Chen, M.; Tsai, M.; Zheng, W.; Wolynes, P. G. The aggregation free energy landscapes of polyglutamine repeats. *Journal of the American Chemical Society* **2016**, *138*, 15197–15203.

(7) Chen, M.; Wolynes, P. G. 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* **2017**, *114*, 4406–4411.

(8) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* **2018**,

(9) Kim, B. L.; Schafer, N. P.; Wolynes, P. G. Predictive energy landscapes for folding -helical transmembrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 11031–6.

(10) Truong, H. H.; Kim, B. L.; Schafer, N. P.; Wolynes, P. G. Predictive energy landscapes for folding membrane protein assemblies. *J. Chem. Phys.* **2015**, *143*, 243101.

(11) Potoyan, D. A.; Zheng, W.; Komives, E. A.; Wolynes, P. G. Molecular stripping in the NF-κB/IκB/DNA genetic regulatory network. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 110–115.

(12) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* **2016**, *138*, 8126–8133.

(13) Tsai, M.; Zhang, B.; Zheng, W.; Wolynes, P. G. The Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *J. Am. Chem. Soc.* **2016**, *138*, 13497–13500.

(14) Zhao, H. UNCOVERING THE BIOPHYSICAL MECHANISMS OF HISTONE COMPLEX ASSEMBLY. Ph.D. thesis, 2018.

(15) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. Water in protein structure prediction. *Proceedings of the National Academy of Sciences* **2004**, *101*, 3352–3357.

(16) Hansen, P. C. The truncatedsvd as a method for regularization. *BIT Numerical Mathematics* **1987**, *27*, 534–553.

(17) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **1992**, *13*, 1011–1021.

(18) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical recipes. *Cambridge University, Cambridge* **1988**,

(19) Box, J. F. Guinness, Gosset, Fisher, and small samples. *Statistical science* **1987**, 45–52.

(20) Amadei, A.; Ceruso, M. A.; Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **1999**, *36*, 419–424.

(21) Sekulic, N.; Bassett, E. A.; Rogers, D. J.; Black, B. E. The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature* **2010**, *467*, 347–351.

(22) Bui, M.; Dimitriadis, E. K.; Hoischen, C.; An, E.; Quénet, D.; Giebe, S.; Nita-Lazar, A.; Diekmann, S.; Dalal, Y. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell* **2012**, *150*, 317–326.

(23) Falk, S. J.; Guo, L. Y.; Sekulic, N.; Smoak, E. M.; Mani, T.; Logsdon, G. A.; Gupta, K.; Jansen, L. E.; Van Duyne, G. D.; Vinogradov, S. A.; Lampson, M. A.; Black, B. E. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* **2015**, *348*, 699–703.

(24) Winogradoff, D.; Zhao, H.; Dalal, Y.; Papoian, G. A. Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome. *Sci. Rep.* **2015**, *5*, 17038.