

Supplementary Information: Promiscuous histone mis-assembly is actively prevented by chaperones

Haiqing Zhao,^{†,‡,||} David Winogradoff,^{¶,||} Minh Bui,[‡] Yamini Dalal,^{*,‡} and
Garegin A. Papoian^{*,†,¶,§}

Biophysics Program, University of Maryland, College Park, MD 20742, USA, Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, Chemical Physics Program, University of Maryland, College Park, MD 20742, USA, and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA

E-mail: dalaly@mail.nih.gov; gpapoian@umd.edu

*To whom correspondence should be addressed

[†]Biophysics Program, University of Maryland, College Park, MD 20742, USA

[‡]Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

[¶]Chemical Physics Program, University of Maryland, College Park, MD 20742, USA

[§]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA

^{||}Contributed equally to this work.

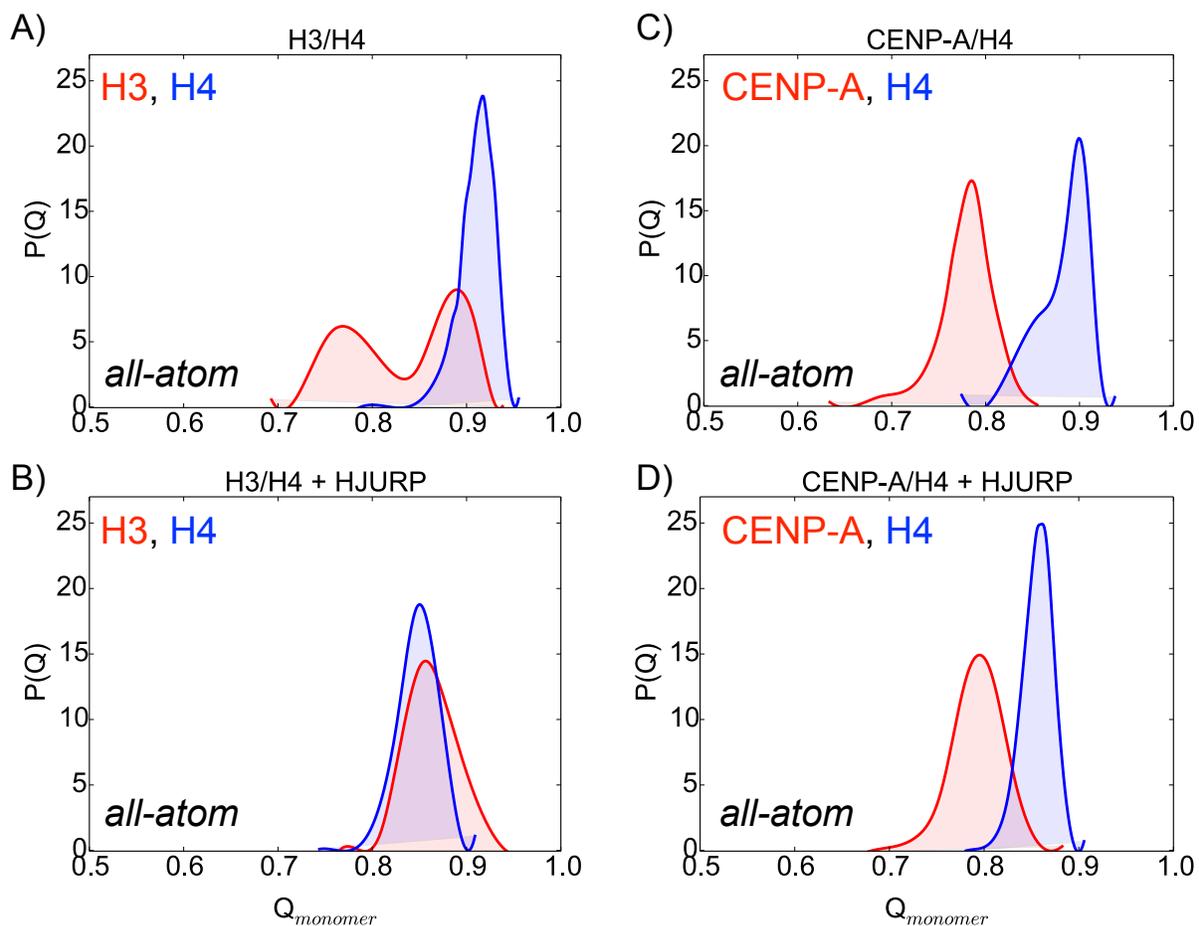


Figure S2: **H4 is structurally consistent in all-atom MD simulations.** $Q_{monomer}$ analysis of the all-atom MD trajectories of (A) isolated H3/H4, (B) H3/H4 in conjunction with HJURP, (C) isolated CENP-A/H4, and (D) CENP-A/H4 in a complex with chaperone HJURP reveals qualitative agreement with the AWSEM coarse-grained MD trajectories. H4 adopts conformations closer to the native state (*i.e.* the experimentally determined crystal structure) than CENP-A or canonical H3 for every all-atom system studied except for H3/H4 in conjunction with HJURP, where histones H3 and H4 are equally close to their respective native states.

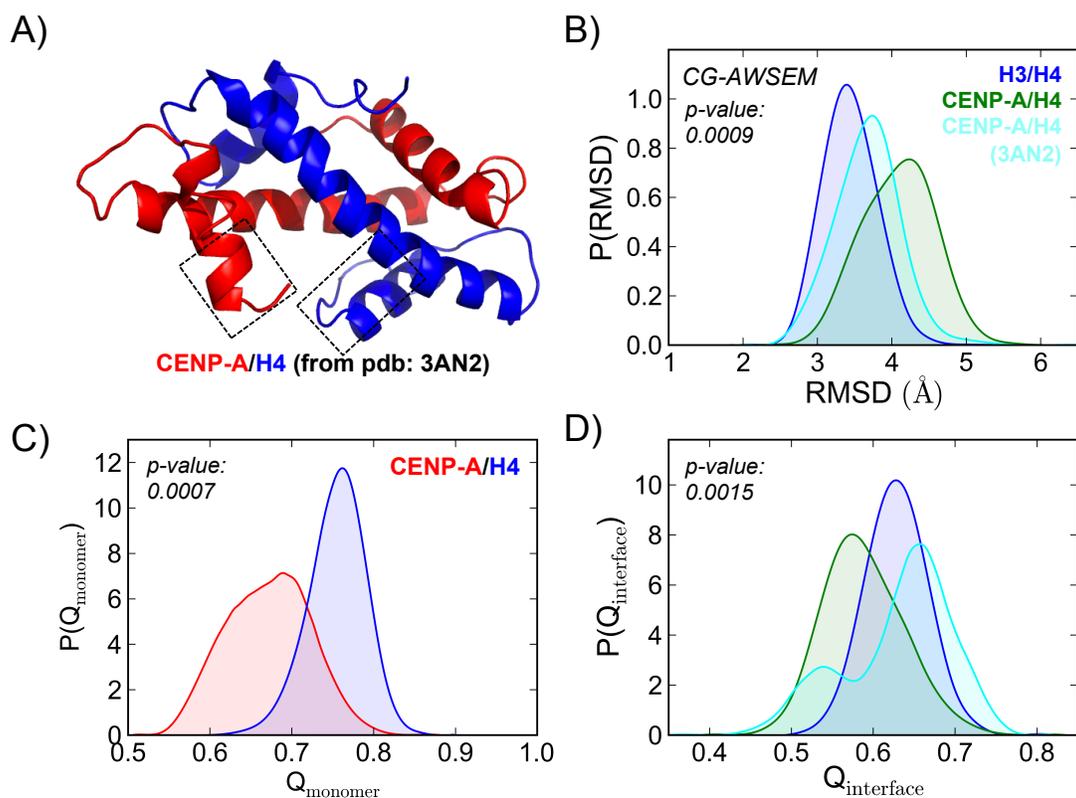


Figure S3: **CG simulation of the CENP-A/H4 dimer from the CENP-A nucleosome crystal structure.** (A) The CENP-A $\alpha 3$ helix (Box on red) is not fully resolved in CENP-A nucleosome crystal structure (PDB ID: 3AN2). However, this structure does include the H4 C-terminal tail (Box on blue). (B) Without the fully-extended CENP-A $\alpha 3$ helix (*i.e.* the CENP-A $\alpha 3$ helix resolved in the CENP-A/H4/HJURP structure, PDB ID: 3R45), the H4 C-terminal tail does not increase the RMSD of CENP-A/H4. (C) Q_{monomer} analysis illustrates that H4 still adopts more native-like conformations than CENP-A. (D) The binding interface of CENP-A/H4^{3AN2} (cyan) has two peaks, compared to one for CENP-A/H4^{3R45} (green), demonstrating that the H4 C-terminal tail is unstable and disrupts the binding interface of CENP-A/H4.

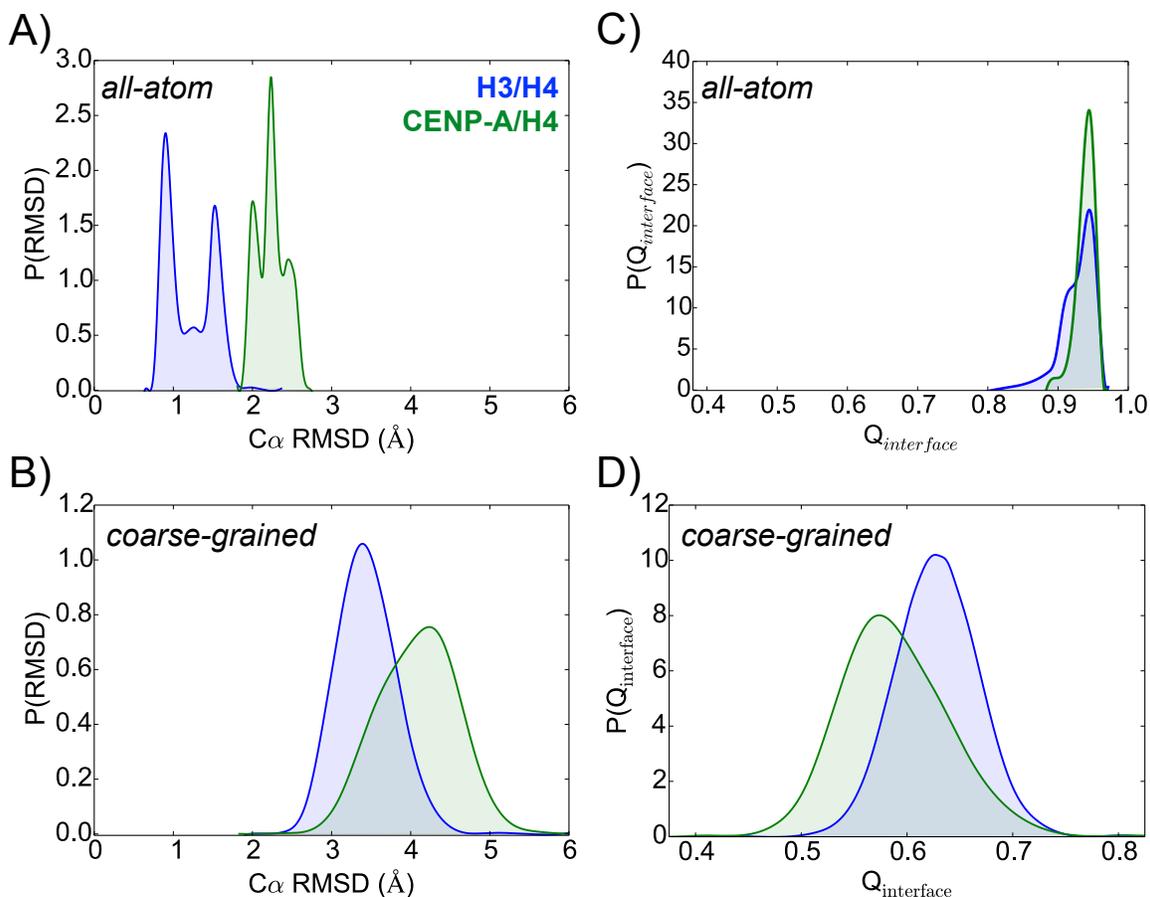


Figure S4: **All-atom and CG-AWSEM MD results qualitatively agree and play complementary roles in analysis.** (A) Probability distribution functions of the $C\alpha$ root-mean-square deviations (RMSD) for the all-atom MD simulation trajectories reveal that replacing H3 with CENP-A leads to greater structural variability in the heterodimer. (B) Probability distributions of the interface Q indicate that both the CENP-A and H3 dimers adopt conformations close to the native state (*i.e.* $Q=1$) in all-atom simulations. (C) RMSD probability distribution functions for the CG-AWSEM simulations demonstrate that CENP-A/H4 is more conformationally variable than H3/H4, an example of the overall qualitative agreement between all-atom and CG-AWSEM MD. (D) Centered at lower averages, with wider variances, compared to all-atom results, the Q interface probability distributions for CG-AWSEM illustrate that coarse-grained MD explores more conformational space further from the native state than all-atom MD.

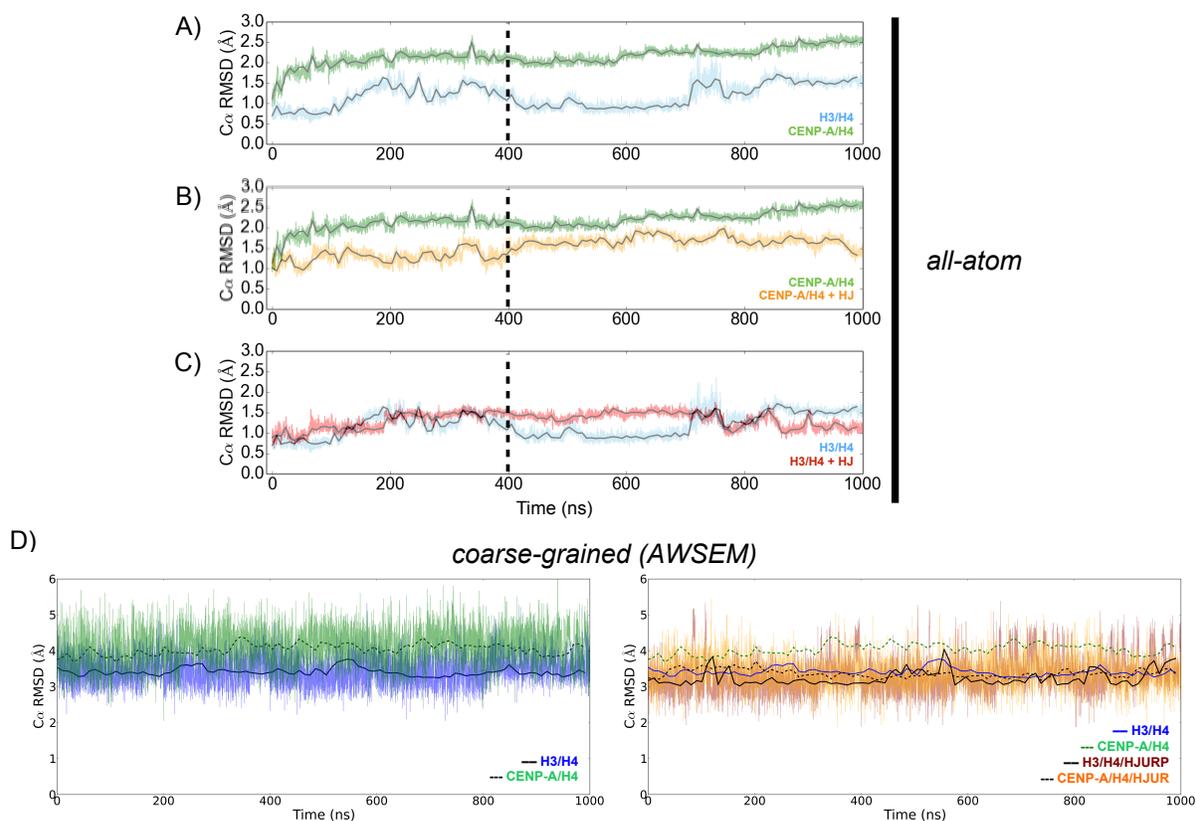


Figure S5: **RMSD illustrates that both all-atom and CG simulations reached convergence.** We first examined the convergence of the all-atom and AWSEM coarse-grained MD trajectories by calculating the C α RMSD of the simulation snapshots with respect to their positions in the experimentally-determined crystal structures as functions of simulation time. In the all-atom simulations, (A) the isolated CENP-A/H4 dimer is more structurally variable than H3/H4 in isolation; (B) the introduction of HJURP reduces the structural variation of the CENP-A/H4 dimer, bringing it closer to the native state, and (C) the presence of HJURP is not an important factor in determining the structural heterogeneity of the canonical H3/H4 dimer. Every all-atom system studied reaches convergence by 400 ns of simulation time (represented by the dashed, vertical lines), therefore only the final 600 ns are used for analysis. (D) Five independent 200 ns CG-AWSEM simulation trajectories were performed for each system, summing to 1000 ns of total CG simulation time. The CG-AWSEM simulations rapidly reach equilibration, therefore we combined those trajectories for further analysis after removing the first 10 ns from each 200 ns run.

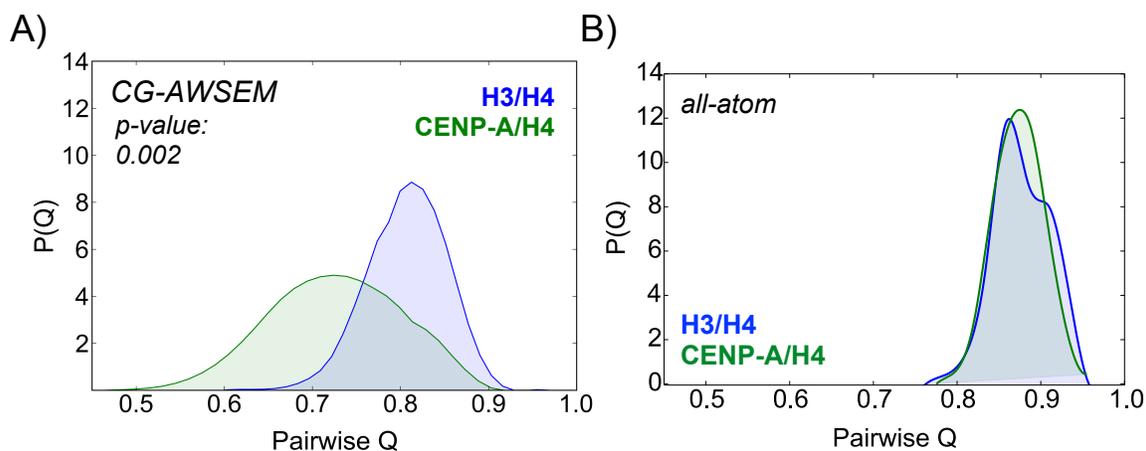


Figure S6: **Pairwise Q value demonstrates that CENP-A/H4 has greater conformational heterogeneity than H3/H4 in CG-AWSEM simulations.** Pairwise Q is when the Q value is calculated between every two conformations from the same simulation, Instead of comparing the simulation conformations to the experimentally determined crystal structure. For each simulation, pairwise Q is calculated pairs of 1000 snapshots, chosen every 500,000 timesteps, corresponding to 1ns. (A) In CG-AWSEM simulations, the pairwise Q distribution for CENP-A/H4 (green) is broader and lower on average than that of H3/H4 (blue), implying that CENP-A/H4 is more conformationally heterogeneous than H3/H4. (B) On the other hand, in all-atom simulations, pairwise Q for both H3/H4 and CENP-A/H4 are high and narrowly distributed, implying that all-atom simulation probes dynamics near the native-state and samples relatively limited conformational space compared to CG simulation.

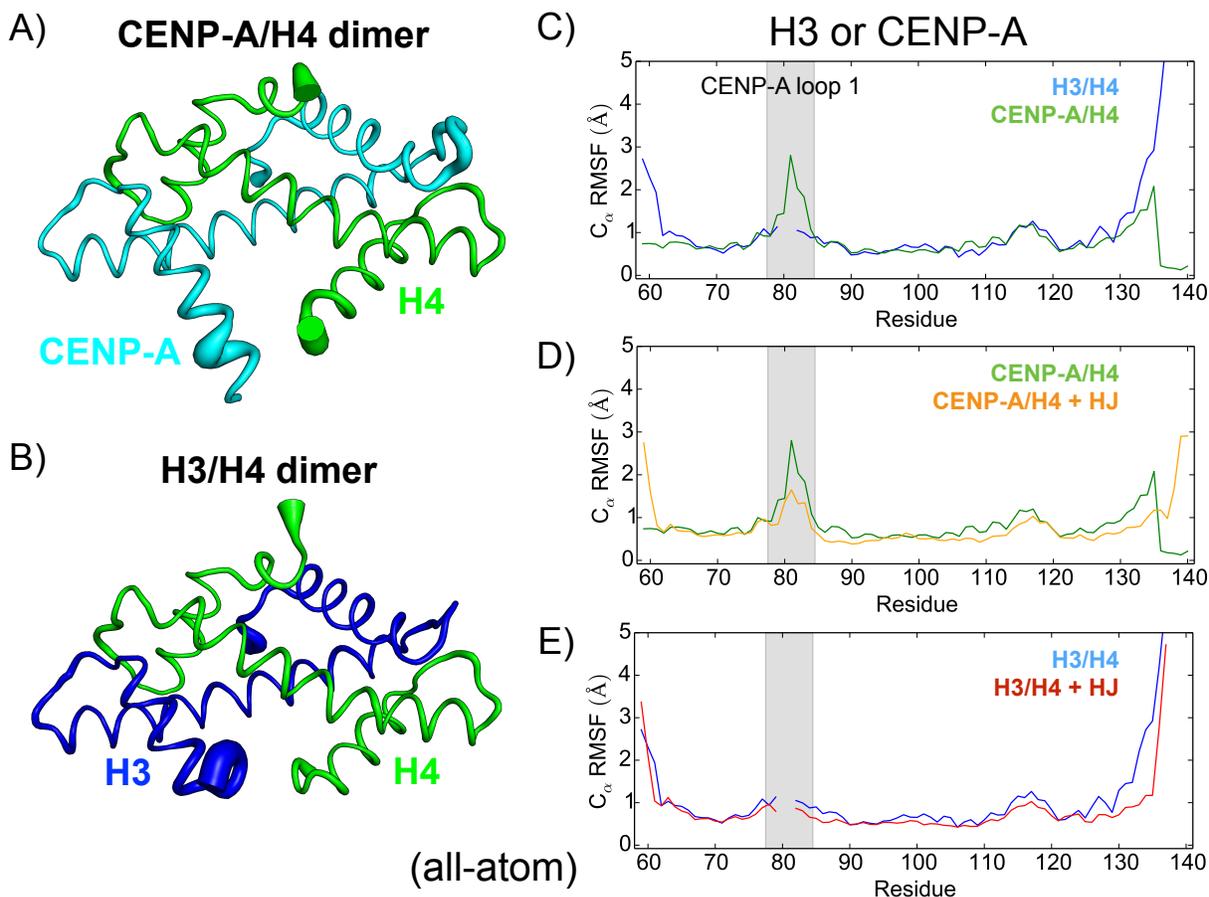


Figure S7: **All-atom local mobility by RMSF.** Root-mean-squared fluctuations (RMSF) are a measure of local mobility. $C\alpha$ RMSF, with respect to the geometric centers, of the all-atom MD simulation snapshots projected onto the crystal structures of (A) the CENP-A/H4 dimer, and (B) the H3/H4, where the tube width is proportional to RMSF, reveals that CENP-A loop 1 exhibits greater local mobility than the same region of canonical H3. (C) In isolated dimers with H4, CENP-A local mobility is only significantly greater than that of H3 at loop 1, except for the highly variable terminal regions. (D) The introduction of HJURP slightly reduces the local flexibility of CENP-A, stabilizing CENP-A loop 1. (E) The presence of HJURP has only a minimal effect on the local mobility of canonical H3.

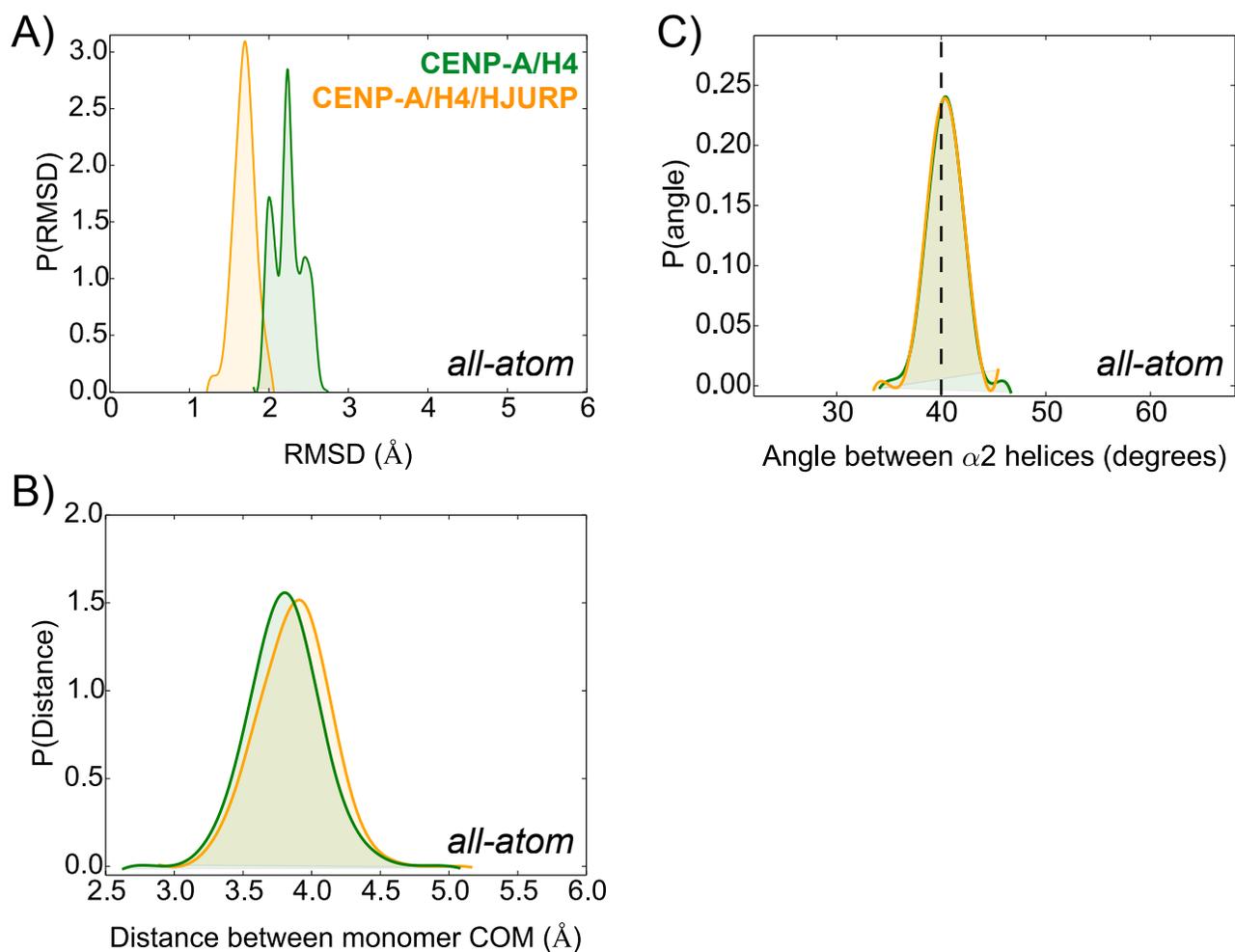


Figure S8: **Global preferences do not change significantly upon the introduction of HJURP to the CENP-A/H4 dimer in all-atom simulations.** (A) Upon the introduction of HJURP, the C α RMSD of CENP-A/H4 decreases, adopting a conformation closer to the 3R45 crystal structure conformation. However, for all-atom MD, the introduction of HJURP does not significantly influence (B) the distance between histone centers-of-mass or (C) the angle between the central α 2 helices. Namely, adding HJURP does not change the global preferences of CENP-A/H4 in all-atom MD simulations.

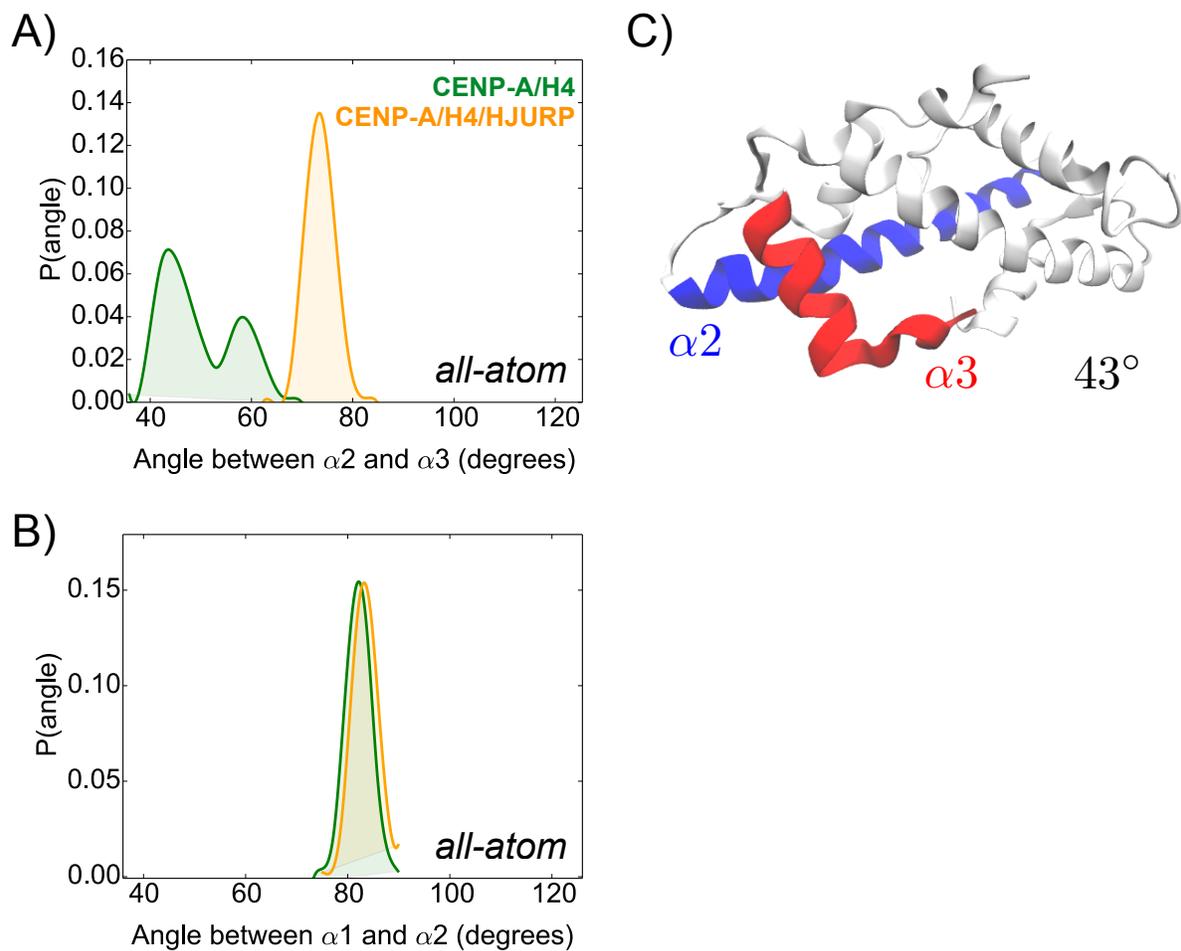


Figure S9: **The introduction of HJURP stabilizes CENP-A $\alpha 3$ in all-atom simulations.** (A) In the absence of HJURP, the angle between CENP-A helices $\alpha 2$ and $\alpha 3$ adopts a bimodal distribution, with two peaks at about 43 and 60 degrees. Upon the introduction of HJURP, this angle becomes relatively fixed, in qualitative agreement with CG-AWSEM MD results. (B) Furthermore, the angle between CENP-A helices $\alpha 1$ and $\alpha 2$ remain the same whether HJURP is present or not, also agreeing with the results from CG simulation. (C) A representative all-atom simulation snapshot of the first peak in the $\alpha 2$ - $\alpha 3$ angle distribution reveals that CENP-A $\alpha 3$ becomes partially unraveled in the absence of HJURP.

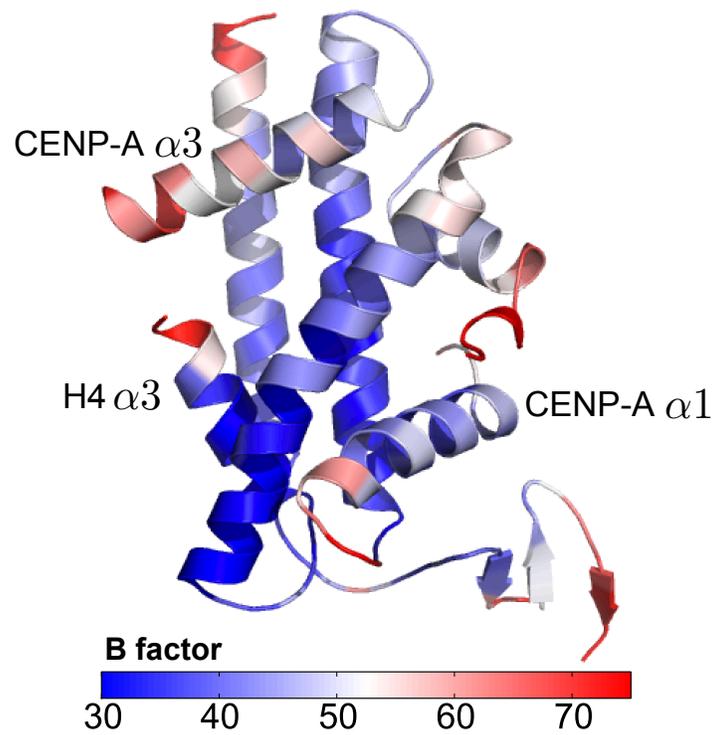


Figure S10: B-factor-colored crystal structure highlights CENP-A $\alpha 3$ and H4 C-terminal residues as regions of high local mobility.

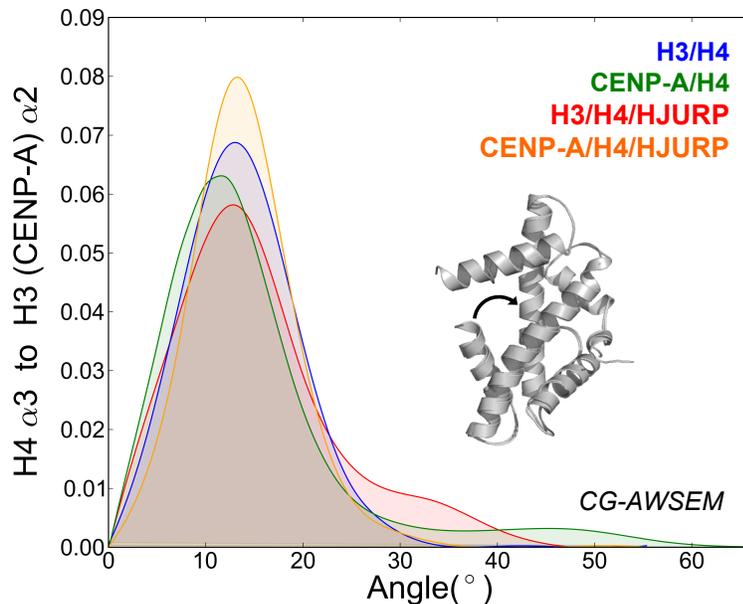


Figure S11: **The angle between H4 $\alpha 3$ and H3 (CENP-A) $\alpha 2$ helices is mostly stable in the absence of H4 C-terminal tail in CG simulations.** For all CG simulations of H3/H4 (blue), CENP-A/H4 (green), H3/H4/HJURP (red), CENP-A/H4/HJURP (orange), the angle between the H4 $\alpha 3$ and H3 (CENP-A) $\alpha 2$ helices is mostly stable. Notice that, due to the flexible C-terminal, the angle distribution for CENP-A/H4 has a slight shoulder based on the interactions between CENP-A C-terminal and the C-terminal end of H4 $\alpha 3$, consistent with all-atom contact analysis (Figure 5). Furthermore, upon the introduction of HJURP, this shoulder disappears, in agreement with the role of HJURP revealed in this paper: stabilizing and regulating the CENP-A C-terminal. Lastly, when introducing HJURP to canonical H3/H4, the angle between H4 $\alpha 3$ and H3 $\alpha 2$ adopts a broader probability distribution, suggesting that HJURP may disrupt the binding interface between H3 and H4. The curved arrow shown with the structure identifies the angle measured.

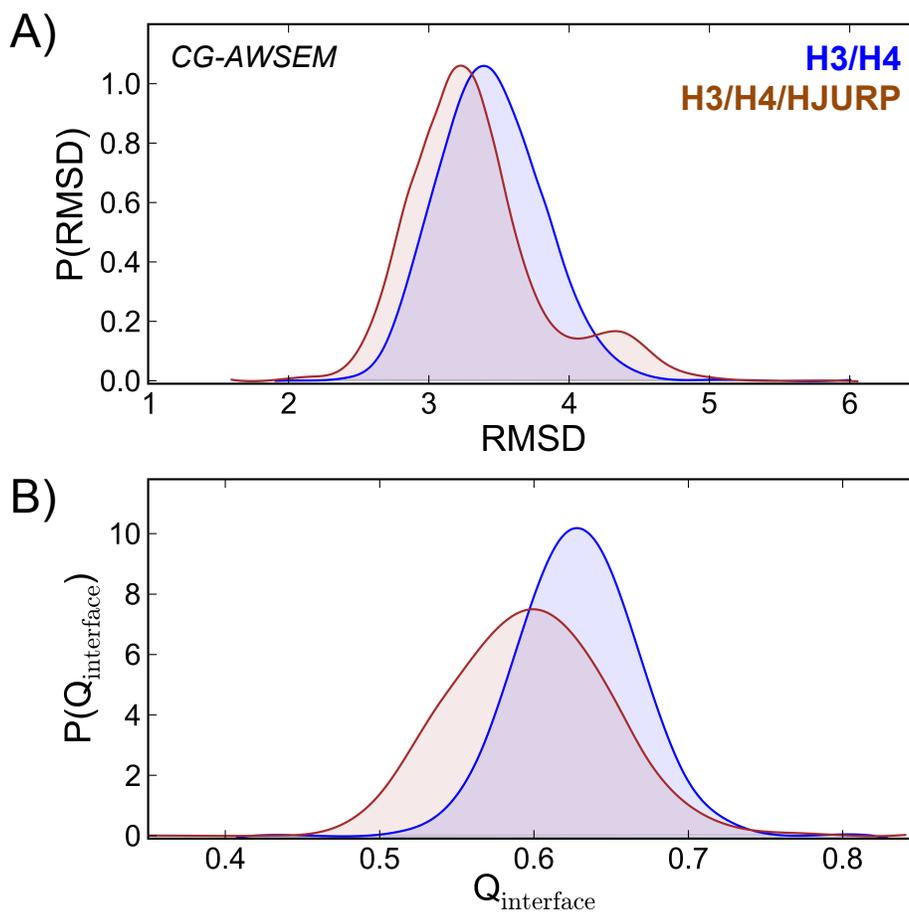


Figure S12: **HJURP disrupts the binding stability of H3 and H4 in CG simulations.** (A) RMSD probability distributions demonstrate that the introduction of HJURP slightly increases the average overall deviation of the canonical H3/H4 dimer from the experimentally determined crystal structure, and leads to a subpopulation of conformations further from the native state (at ~ 4.5 Å RMSD). (B) Furthermore, upon the introduction of HJURP, the binding interface between H3 and H4 becomes less native-like, adopting lower $Q_{interface}$ values on average than when HJURP is absent.

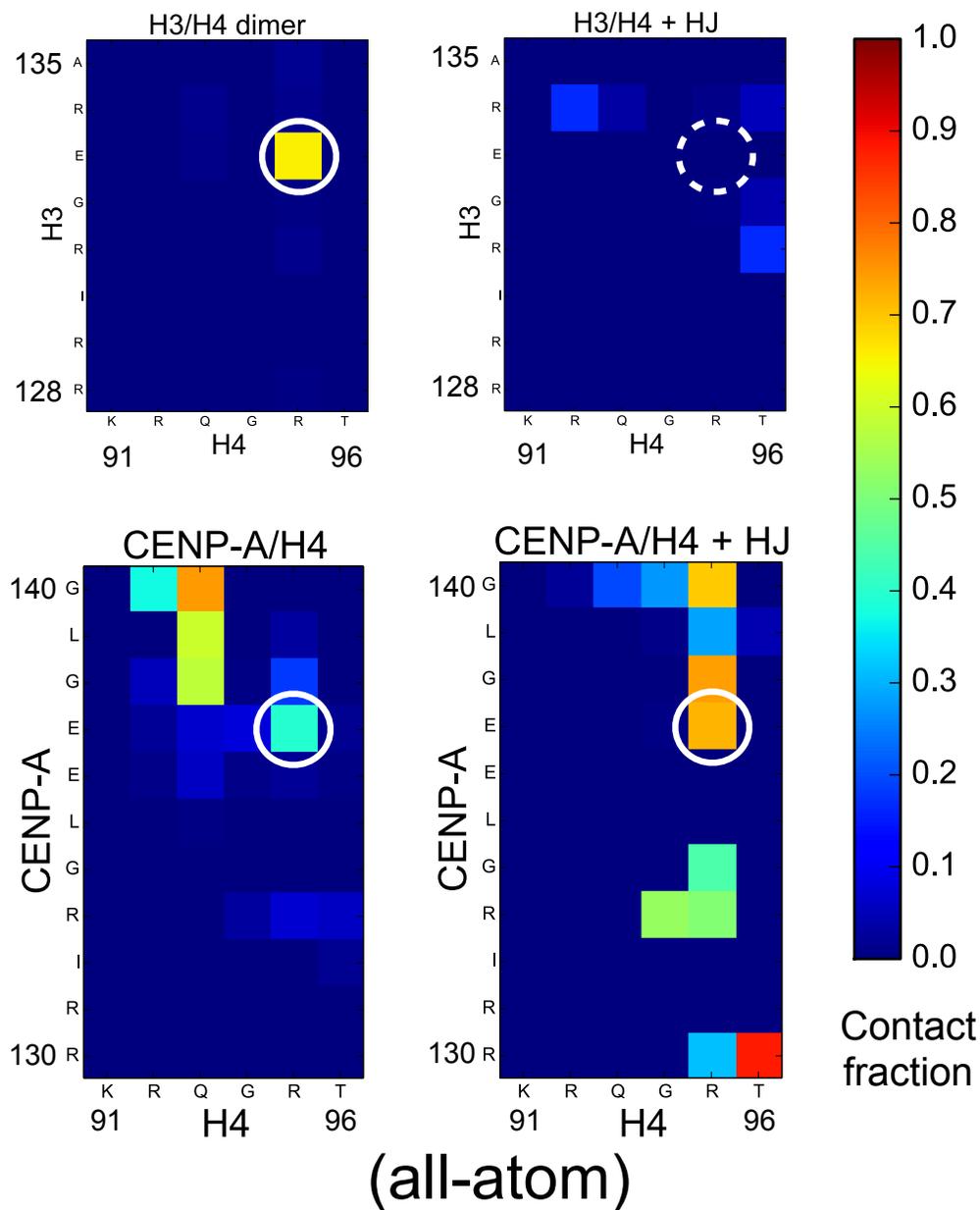


Figure S13: **HJURP stabilizes interactions between the C-termini of CENP-A and H4, but not between H3 and H4, in all-atom simulations.** In isolation, one salt-bridge dominates the interactions between the C-termini of H3 and H4, H3 E133 to H4 R95, whereas the C-termini of CENP-A and H4 form several different contacts, including a salt-bridge between CENP-A E137 and H4 R95. Upon the introduction of HJURP, the C-termini of CENP-A and H4 form even more interactions, while the contacts between the C-termini of H3 and H4 become disrupted.

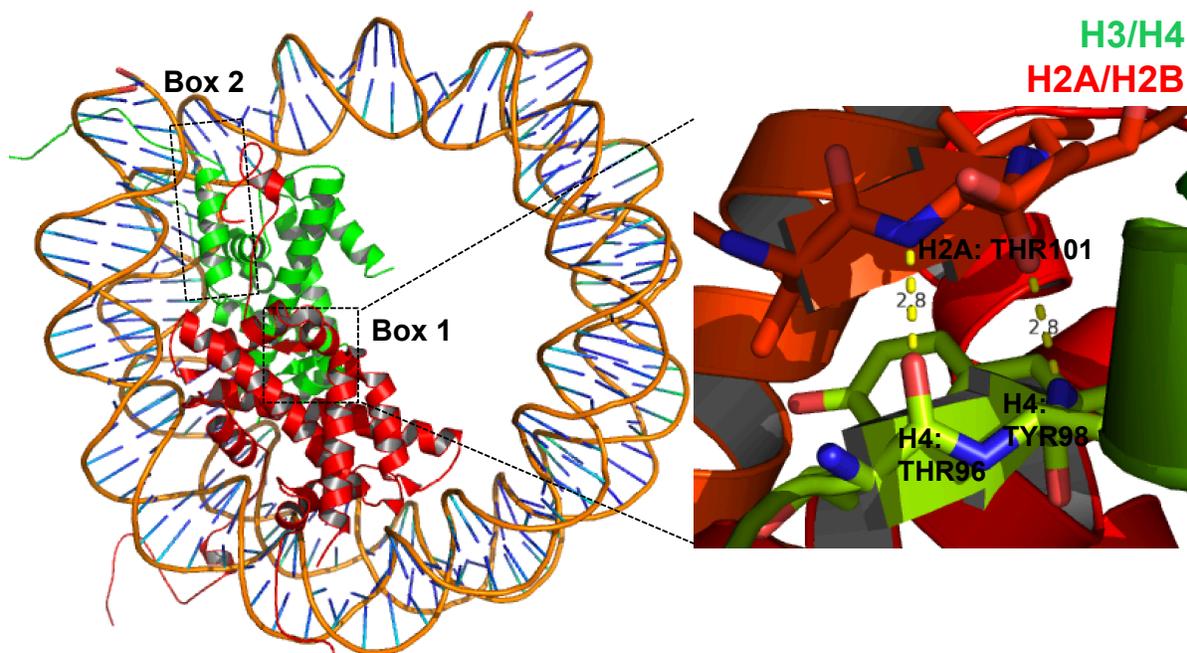


Figure S14: **Histone tails and H3 (CENP-A) α N helix primarily interact with DNA and other histones.** Colors identify histone dimers H3/H4 (green) and H2A/H2B (red) in a typical nucleosome structure (PDB ID: 1KX5). In this structure, the H3 α N helix (Box 2) largely interacts with DNA and the H2A histone tail. Additionally, in the nucleosome context, the H4 C-terminal tail region forms a β strand, between H4 THR96/TYR98 and H2A' THR101, shown in Box 1 and in the zoomed-in view.

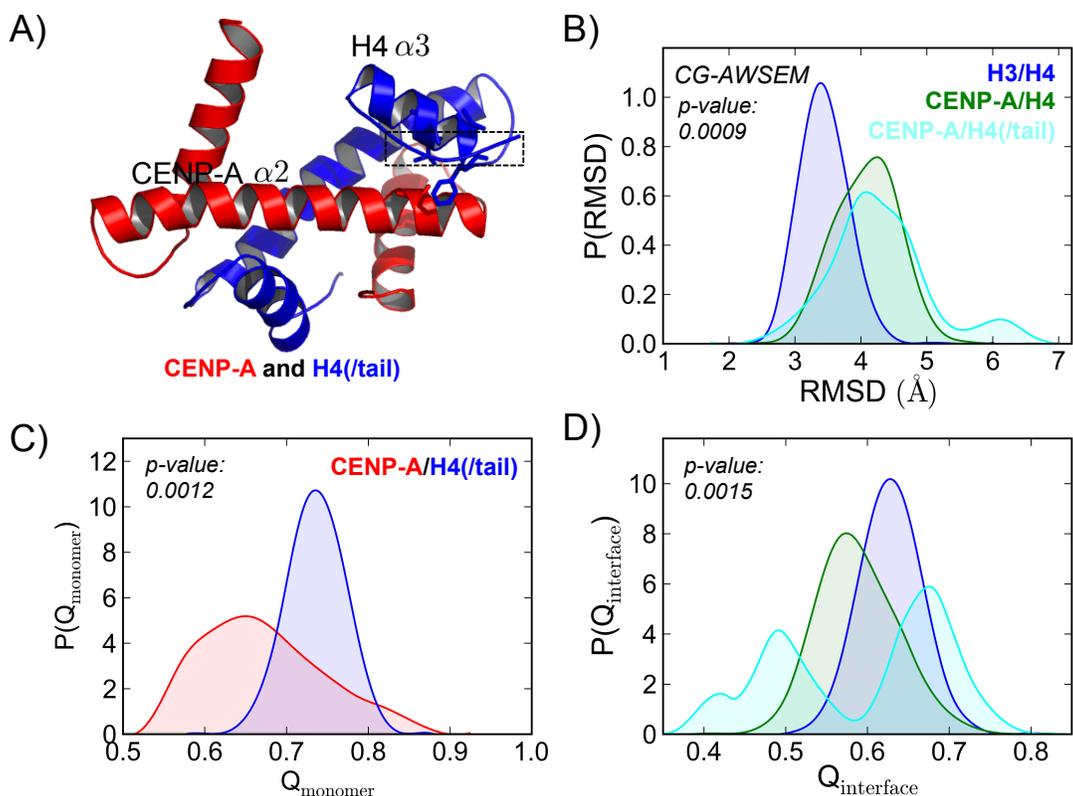


Figure S15: **Including the H4 C-terminal tail increases the structural flexibility of the CENP-A/H4 dimer.** The initial conformation of the “CENP-A/H4(w/tail)” simulation is composed of CENP-A from the CENP-A/H4/HJURP structure (PDB ID: 3R45) and H4 from the CENP-A nucleosome structure (PDB ID: 3AN2) after structural alignment. (A) The boxed area in the structure figure illustrates where the H4 C-terminal tail has hydrophobic interactions with H4 $\alpha 3$ and CENP-A $\alpha 2$. (B) CENP-A/H4 with the H4 C-terminal tail included has a larger average RMSD with respect to the crystal structure than that of the CENP-A/H4 dimer structure excluding the H4 C-terminal tail. (C) Q_{monomer} analysis demonstrates that, even with the H4 C-terminal tail included, H4 still remains more native-like than CENP-A. (D) Lastly, the $Q_{\text{interface}}$ probability distributions show that, when adding the H4 C-terminal tail, the binding interface of CENP-A/H4 is no longer stable, clearly adopting multiple different conformational states.

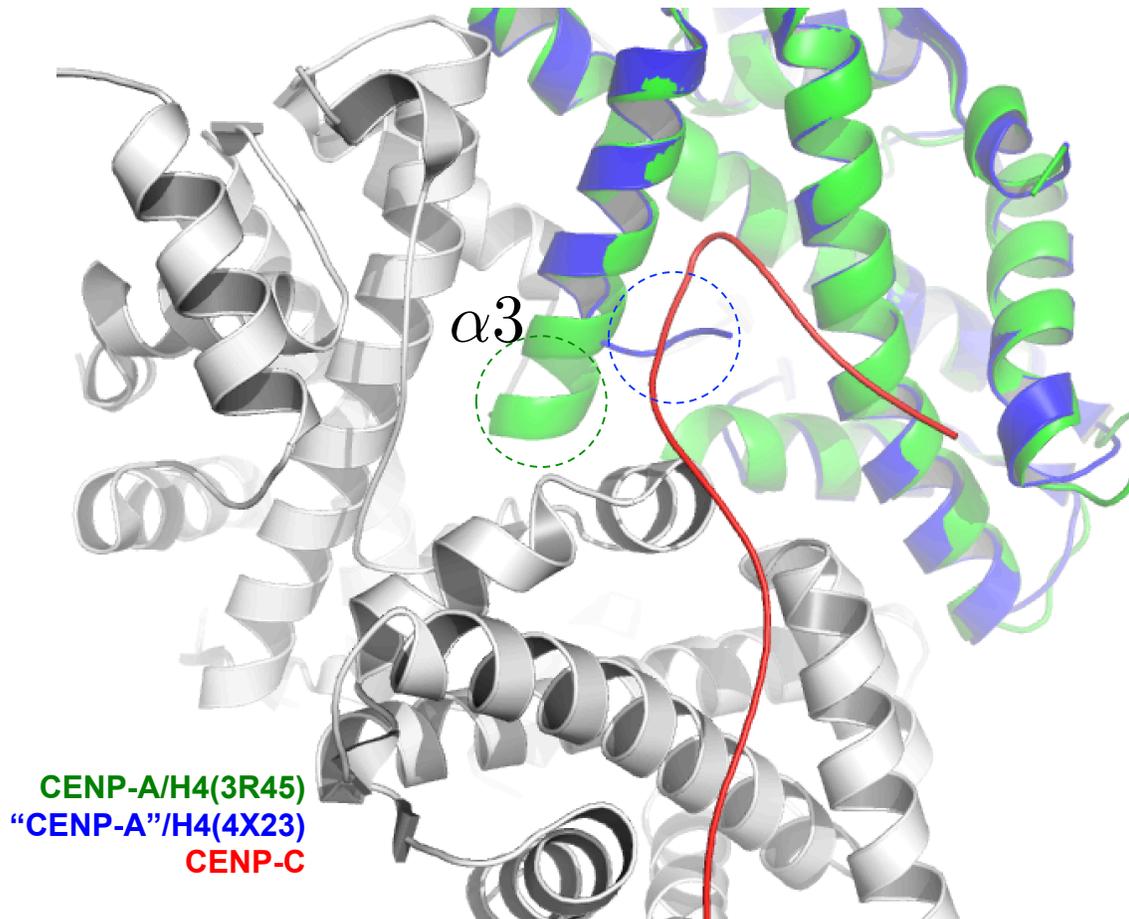


Figure S16: **Alignment of CENP-A structures from different contexts show the “on” and “off” states of its C-terminal tail.** The CENP-A/H4 structure from CENP-A/H4/HJURP crystallography (PDB ID: 3R45) is shown in green, featuring an ordered C-terminal tail (circled in green) at the end of $\alpha 3$, corresponding to the “off” binding state. The blue, gray and red structures are all from the chimeric-nucleosome/CENP-C complex (PDB ID: 4X23). One chimeric “CENP-A/H4” is shown in blue, containing the C-terminal residues of CENP-A (circled in blue) and the remainder of H3. The rest of the histone core is colored gray. The C-terminal tail of CENP-A, at the end of $\alpha 3$, is disordered in the nucleosomal context, in the “on” binding state, interacting with CENP-C (red).

RMSIP calculation

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

$$RMSIP = \left(\frac{1}{10} \sum_{i=1}^{10} \sum_{j=1}^{10} (\vec{\eta}_i \cdot \vec{\nu}_j) \right)^{\frac{1}{2}}, \quad (S1)$$

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

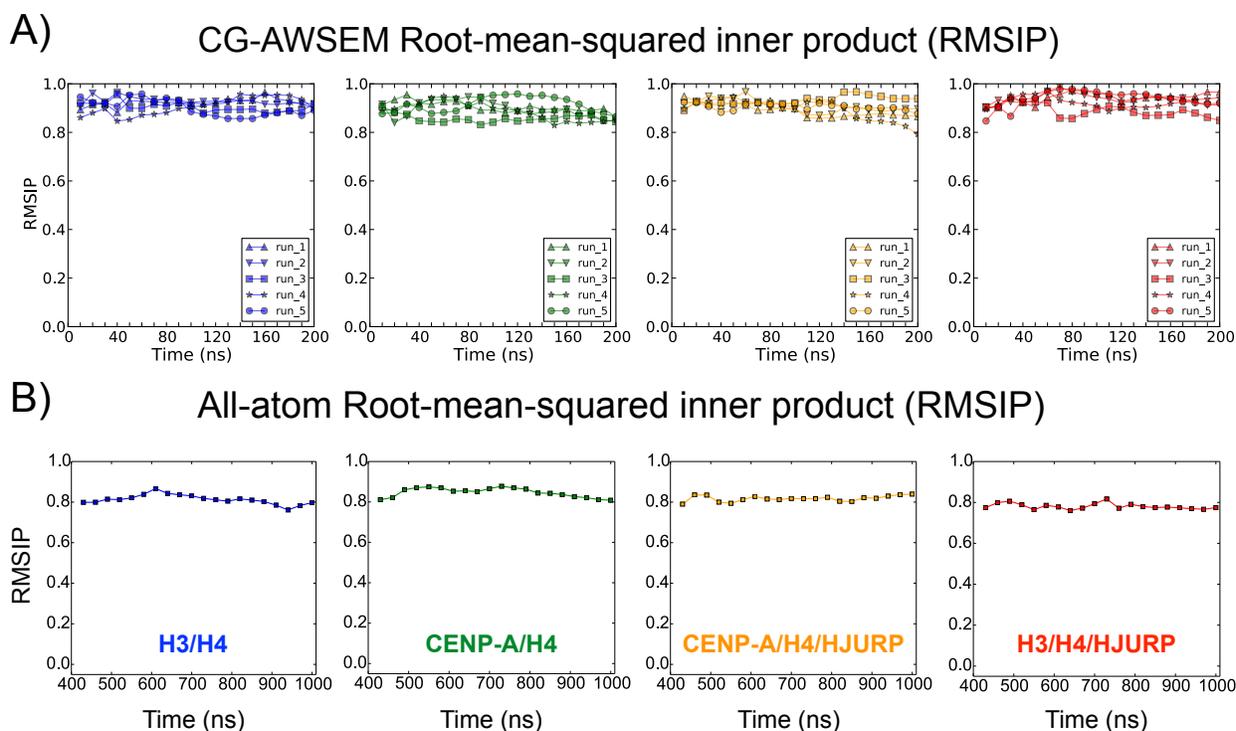


Figure S17: **RMSIP demonstrates the convergence of CG and all-atom simulations.** To extend our evaluation of convergence, we calculated the root-mean-squared inner product (RMSIP). RMSIP is a sum of every dot product between the first ten eigenvectors of the first half of the trajectory and first ten eigenvectors of the second half. It is a normalized measure, where 1 indicates identical eigenvectors, and the simulation reaches convergence when RMSIP is close to 1. (A) CG simulations reached convergence (RMSIP > 0.8) after 10 ns, and, (B) convergence was achieved for the final 600 ns of all-atom MD simulation which are considered for analysis (RMSIP > 0.75).

Angle Analysis

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

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

where (x_i, y_i, z_i) represents the position of the i th C_α , and (x_0, y_0, z_0) is the coordinates of the geometric center of the selected helix. Then we use singular value decomposition (SVD) to determine all the eigenvalues of matrix V . The eigenvector corresponding to the biggest eigenvalue provides the orientation vector. A diagrammatic sketch is shown in Figure S18.

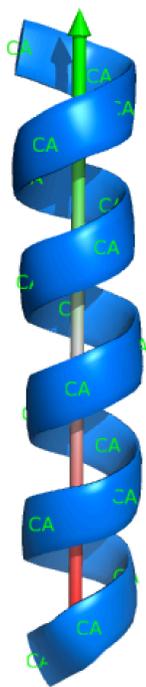


Figure S18: **Orientation vector sketch for one α helix.**