Effects of H2A.B incorporation on nucleosome structures and dynamics

Havva Kohestani[1] and Jeff Wereszczynski [2,*]

[1]Department of Biology and Center for Molecular Study of Condensed Soft Matter, Illinois Institute of Technology, Chicago, Illinois

[2]Department of Physics and Center for Molecular Study of Condensed Soft Matter, Illinois Institute of Technology, Chicago, Illinois

[*]Correspondence: jwereszc@iit.edu

ABSTRACT

The H2A.B histone variant is an epigenetic regulator involved in transcriptional upregulation, DNA synthesis, and splicing that functions by replacing the canonical H2A histone in the nucleosome core particle. Introduction of H2A.B results in less compact nucleosome states with increased DNA unwinding and accessibility at the nucleosomal entry and exit sites. Despite being well characterized experimentally, the molecular mechanisms by which H2A.B incorporation alters nucleosome stability and dynamics remain poorly understood. To study the molecular mechanisms of H2A.B, we have performed a series of conventional and enhanced sampling molecular dynamics simulation of H2A.B and canonical H2A containing nucleosomes. Results of conventional simulations show that H2A.B weakens protein/protein and protein/DNA interactions at specific locations throughout the nucleosome. These weakened interactions result in significantly more DNA opening from both the entry and exit sites in enhanced sampling simulations. Furthermore, free energy profiles show that H2A.B containing nucleosomes have significantly broader free wells, and that H2A.B allows for sampling of states with increased DNA breathing, which are shown to be stable on the hundreds of nanoseconds timescale with further conventional simulations. Together, our results show the molecular mechanisms by which H2A.B creates less compacted nucleosome states as a means of increasing genetic accessibility and gene transcription.

SIGNIFICANCE Nature has evolved a plethora of mechanisms for altering the physical and chemical properties of chromatin fibers as a means of controlling gene expression. These epigenetic processes may serve to increase or decrease DNA accessibility, manage the recruitment of remodeling factors, or tune the stability of the nucleosomes that make up chromatin. Here, we have used both conventional and enhanced sampling molecular dynamics simulations to understand how one of these epigenetic mechanisms, the substitution of canonical H2A proteins with the H2A.B variant, exerts its influence on the structures and dynamics of the nucleosome. Results show at the molecular level how this variant alters inter-molecular interactions to increase DNA accessibility as a means of increasing genetic accessibility and gene transcription.

INTRODUCTION

The capability to store, transfer, and modify genetic information is one of the distinctive features of living organisms. Eukaryotes have solved this problem by evolving multiple molecular mechanisms to control the packing and accessibility of their genetic code in their confined cellular environments. Central to these processes is the nucleosome core particle (NCP), a complex composed of eight histone proteins that wraps ~147 DNA basepairs (1–4). It consists of eight proteins, including two copies each of the H2A, H2B, H3, and H4 histones. Each core histone is comprised of three α-helices (α_1, α_2, and α_3) connected by two short loops (L1 and L2), with each histone also having various lengths of unstructured N- and C-terminal "tail" regions. The histones are arranged into a $(H3\text{-}H4)_2$ tetramer that is flanked by two H2A-H2B dimers (Figure 1) (1–7). Nucleosomes assemble on long DNA molecules to form chromatin fibers which condense into higher-order structures that are stabilized by significant electrostatic forces between the DNA and histone protein core (4, 8). Several mechanisms exist to tune the nucleosome stability, and thus DNA accessibility and chromatin structure, which include post-translational modifications to select histone residues, the recruitment of ATP-dependent chromatin remodeling complexes, and the substitution of canonical histones with histone variant proteins (9–19).

The H2A histone family has the most diverse and populated group of variants (15, 20–23). This group includes the canonical H2A histone as well as the H2A.B, H2A.L, H2A.P, H2A.W, H2A.Z, H2A.X, and macroH2A variants(24). Of particular note, H2A.B (which was formerly referred to as H2A.Bbd for Barr-body-deficient) is associated with DNA synthesis, splicing, and transcription upregulation (23–29). It was first described by Chadwick and Willard as a histone variant that is excluded from the inactive X chromosome of females and is localized to H4 hyperacetylated regions (30). It is only 48-50% identical to canonical H2A, lacks both the C-terminal tail and parts of the "docking domain," and does not contain the "acidic patch" which promotes chromatin condensation (31–33).

Substitution of canonical H2A with H2A.B results in large-scale structural changes within the nucleosome. *In vitro*, H2A.B containing nucleosomes protect only 118-130 basepairs from micrococcal nuclease digestion, suggesting that the entry and exit DNA segments are more exposed to solution in H2A.B nucleosomes (34, 35). Fluorescence resonance energy transfer (FRET) experiments have also shown this, with H2A.B substitution increasing the distance between DNA ends in reconstituted nucleosomes (34). Both atomic force and cryo-electron microscopy experiments have revealed coarse representations of these extended states, with the most common conformations observed to have ~10 basepairs of DNA unwrapped from each DNA end (35). These extended states were further confirmed with small angle X-ray scattering experiments, which showed that in solution H2A.B substitution increases the nucleosome's radius of gyration from 43.4±0.3 Å to 49.4±0.4 Å (29).

Together, these results demonstrate that H2A.B has a significant affect on nucleosome structures which directly influences their functions. However, to date there has been a lack of details regarding the molecular origins of these effects. Indeed, the most comprehensive atomic-scale study of H2A.B's effect is a recently published crystal structure of the H2A.B-H2B dimer at 2.6 Å resolution which demonstrates that variant substitution drastically alters the surface electrostatic potential of

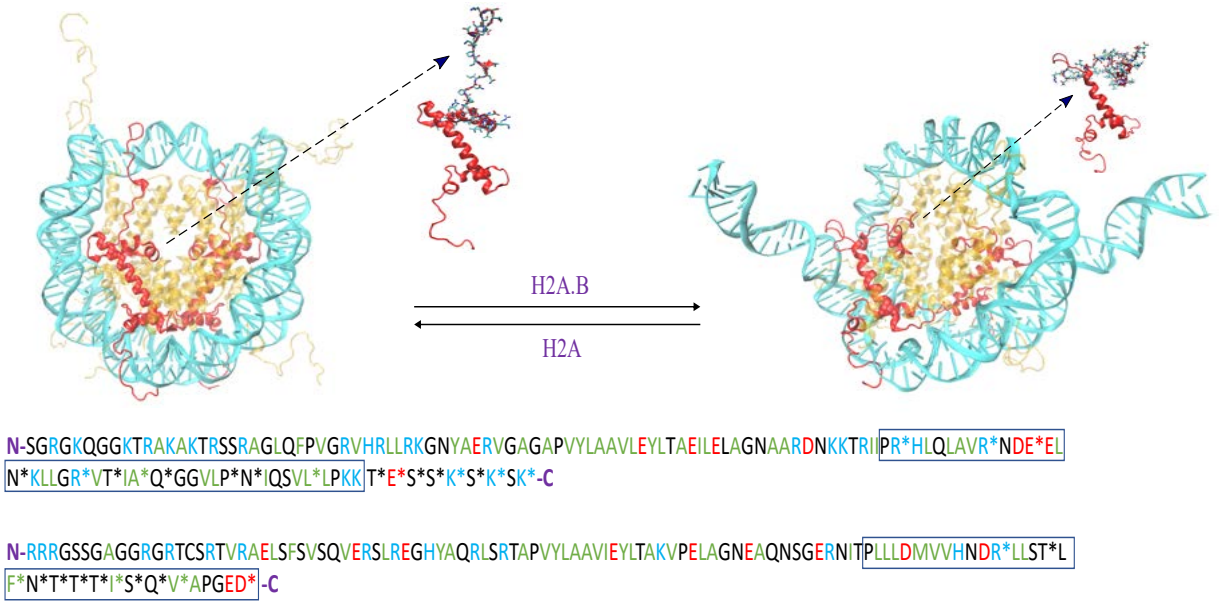

Figure 1: Substitution of canonical H2A with H2A.B results in more open nucleosome structures (top). Sequence alignment of H2A and H2A.B histones (bottom) show H2A.B has a shorter docking domain (boxed), alterations to the acidic patch, and no C-terminal tails(11). Basic residues are in blue, acidic in red, hydrophobic in green, polar in black.

the dimer by eliminating the overall positive charge in the DNA binding site (36). However, there has yet to be a structure solved of a complete H2A.B containing nucleosome, likely due to the absence of the C-terminal, shorter docking domain, and increased DNA flexibility hindering crystallization. This has precluded further studies on the effects of H2A.B substitution on nucleosome dynamics, and how those effects propagate into large-scale structural rearrangements and increased DNA motions. To address these challenges, we have constructed an atomic-scale model of an H2A.B containing nucleosome. Using comparative conventional and free energy molecular dynamics simulations, our results demonstrate that H2A.B substitution alters interactions at the DNA-dimer and dimer-tetramer interfaces. These changes show distinct energetic characteristics, with modifications in the L2 loops leading to increased DNA opening and changes in the docking domains likely resulting in destabilization of the dimer/tetramer interface on timescales significantly longer than those sampled here. Taken together, our results show how histone substitution creates less compacted nucleosomes which are responsible for H2A.B's biological functions.

MATERIALS AND METHODS

Molecular Dynamics Simulations

A model of the canonical nucleosome was constructed based on the 1KX5 crystal structure (3). The H2A.B containing nucleosome was modelled by modifying the H2A sequences in the 1KX5 structure to the H2A.B sequence (11). After initiation of this work, a structure of the H2A.B/H2B dimer was released, which showed a good agreement to our model, with a heavy

atom root-mean square deviation (RMSD) of 1.8 Å between the two structures, indicating our model was of sufficient accuracy for molecular dynamics simulations (36).

All systems were built in the tleap tool of the Amber molecular dynamics package (37). The Amber ff14SB, OL15, and TIP3P force fields were used for the protein, DNA, and water force fields, while ions parameters were based on the Joung and Cheatham model (38–41). Systems were neutralized and solvated in a water box with a 12 Å buffer of solvent and an additional a 0.15 M NaCl concentration. All systems were minimized twice for 10000 steps, once with and once without heavy atom restraints. Systems were gradually heated in the canonical ensemble, from 0 K to 300 K over 5000 steps with restraints on solute heavy atoms. The restraints were released by relaxing the systems in the isothermal-isobaric ensemble. The Langevin piston was used for a barostat and Langevin dynamics were performed for temperature control with a collision frequency of 2 ps^{-1} (42). Systems of the compact and open states were each performed twice for 550 ns, resulting in 8.8 µs of explicit solvent simulations. All simulations was performed by with the GPU accelerated version of PMEMD (43).

Enhanced sampling simulations followed a similar protocol with three major exceptions. First, the H3 N-terminal tails were truncated. Second, they were performed in implicit solvent while the default PBRadii of the Generalized Born Surface Area model was set to *mbondi2* with *igb=5* and a salt concentration of 0.15 M (44, 45). We used the default value of rgbmax=25 for the maximum distance between atom pairs to calculate the Born radii and a nonbonded cutoff of effectively infinity (999.0 Å). The ParmEd (parameter file editor) program was used to repartition the hydrogen atom masses to reduce the system high frequency motions, which allowed the use of a 4 ps MD timestep (46). Finally, Adaptively Biased Molecular Dynamics (ABMD) were performed (47). For ABMD, two reaction coordinates were used. The first was the radius of gyration of the heavy atoms of the first 73 DNA basepairs, and the second was the radius of gyration of heavy atoms of the last 73 DNA basepairs. A flooding time of 100 ps was used along with a spatial resolution of 1 Å. For each system, three simulations of 1.6 µs were run, resulting in 4.8 µs of sampling for each. Simulations were initiated from the closed DNA states, with an equilibration time of 50 ns.

Simulation Analyses

We used the cpptraj clustering tool to extract the dominant structural clusters. The optimal number of clusters was based on Davis-Bouldin Index (DBI) and seudo-F statistic (pSF) metrics (48, 49). The maximum DBI and minimum pSF numbers gives the optimal number of structural clusters. This number is six for both H2A and H2A.B NCPs (Figure S1). UCSF Chimera modeller was used to reattach the truncated H3 histone tails for the final explicit solvent simulations of open states (50). The *MMPBSA.py* script was used to perform an MM/GBSA analysis (51). This was performed on the final 500 ns of the compact state simulations. The *mbondi3* radii set were used with the *igb*=8 implicit solvent model and a salt concentration of 0.15 M (45, 52). Binding energies were calculated for DNA/dimer, dimers-tetramer and H2A/H2A.B-tetramer units. The cpptraj analysis tool was used for RMSD, Radius of gyration, RMSF, and structural clustering analyses (53). Visual Molecular Dynamics (VMD) was used for figures (54).

RESULTS

Simulations of Compact States Show Weaker Interfaces in H2A.B Nucleosomes

Two, 550 ns conventional simulations of canonical and H2A.B containing nucleosomes were performed based upon the compact state in the 1KX5 crystal structure. Large scale dynamics were similar between systems, which were primarily characterized by H3 tail collapse onto the nucleosomal DNA, as had been observed in other full-size nucleosome simulations (55–60). This is exemplified by root-mean-square-deviation (RMSD) calculations, which show changes on the order of 1.0 and 0.5 Å in the DNA and core protein structures, respectively, in these systems (Figures S2 & S3). Given that experiments have suggested that open states of H2A.B nucleosomes dominate in solution, the fact that these simulations remained in compact states is likely an artifact of the limited sampling inherent to conventional MD simulations.

Although large-scale structural rearrangements were not observed in these simulations, energetic analysis shows that incorporation of H2A.B may lead to destabilization of the compact nucleosome structure. An MM/GBSA analysis shows that dimer-tetramer interactions in canonical nucleosomes are largely stabilized by favorable van der Waals (vdW) interactions, which are counterbalanced by non-favorable electrostatic interactions between the two cationic species (Table 1). In H2A.B containing nucleosomes, there are reduced dimer-tetramer contacts, which results in a weaker interface. Although the electrostatic repulsive energies are reduced, the attractive vdW forces are much more significantly reduced, resulting in a less favorable ΔE_{total}. In both systems we note that nucleosome stability is enhanced by strong DNA-dimer interactions, which are roughly equal in strength in H2A and H2A.B containing systems. This is in line with experiments which have shown that DNA is required for nucleosome stability at physiological salt concentrations, and that the histone core will only assemble at high salt concentrations without DNA (61–63). We note that the MM/GBSA analysis presented here is relatively crude and does not include important thermodynamic terms such as conformational entropy and explicit solvation, therefore although the trends are of qualitative value, care should be taken in quantitative interpretation of the energetic values (64–68).

To determine the location of residues that contribute to destabilization of H2A.B nucleosomes, an energetic decomposition was performed for residues in the histone core. Residues in the long histone tails were excluded from this analyses, as their sampling is highly stochastic and subject to significant noise (69). The change in energy upon H2A.B substitution, $\Delta\Delta E_{total} = \Delta E_{Total,H2A.B} - \Delta E_{Total,H2A}$, was computed, and the most significant values are reported in Table 2. Residues that contribute the most to canonical dimer-tetramer stability are primarily located in three regions: the H2B α_2/H4, the H2A/H3-H4, and the H2A docking domain/H3 interfaces, which we denote as regions one, two, and three in Figure 2 (Tables 2, S1 & S2). Upon H2A.B substitution, these interaction energies are altered (Table S3 & S4), and residues in these three areas primarily lead to destabilization of the histone core (Table 2). The first region of this interface corresponds largely to residues Asp68, Asp85, Lys91, Arg92, and Tyr98 from the H4 α_3 helix and Glu73 and Leu97 from the H2B α_2 helix. Together, substitution destabilizes this complex through a $\Delta\Delta E_1 = 6.6 \pm 2.6$ kcal/mol of energy. The second region corresponds primarily to residues Glu94 and Glu97 of the H3 α_2 helix, and Arg134 of the H3 C-terminal, and Thr101 and Ala103 of the H2A docking

System	Interacting Segments	ΔE_{vdW}	ΔE_{elec}	ΔE_{total}
HA2 NCP	Dimer-Tetramer	-238.2±6.1	113.3±4.8	-124.9±5.1
	H2A-Tetramer	-137.9±4.7	55.2±3.6	-82.7±4.4
	Dimer-DNA	-369.5±7.2	-27.9±10.6	-397.4±11.2
H2A.B NCP	Dimer-Tetramer	-173.9±4.6	83.9±4.8	-90±19.1
	H2A.B-Tetramer	-79.8±3.1	32.1±3.7	-47.7±3.4
	Dimer-DNA	-357.7±8.8	-22.7±9.2	-380.4±10.6

Table 1: Energy differences (kcal/mol) of association between the dimer/tetramer, H2A/tetramer, and dimer/DNA constituents as estimated by an MM/GBSA analysis. A negative value indicates more favorable interactions in the complex relative to in solution. The total interaction energy (ΔE_{total}) is a sum of the van der Waals (ΔE_{vdW}) and electrostatic (ΔE_{elec}) interactions.

Nucleosome Region (Figure 2)	Histone	Location	Residue	Mutated residues in H2A.B NCP	$\Delta\Delta E_{total}$ (kcal/mol)
1	H2B	α_2 Helix	GLU 73		2.0±0.6
	H2B	α_3 Helix	LEU 97		0.4±1.1
	H4	α_2 Helix	ASP 68		2.2±0.3
	H4	α_3 Helix	ASP 85		1.1±0.2
	H4	α_3 Helix	LYS 91		0.0±1.2
	H4	α_3 Helix	ARG 92		1.1±1.9
	H4	α_3 Helix	TYR 98		-0.2±1.3
2	H2A	Docking domain	THR 101		0.5±0.7
	H2A	Docking domain	ALA 103	SER 103	-2.1±0.8
	H3	α_2 Helix	GLU 94		-0.1±0.7
	H3	α_2 Helix	GLU 97		0.8±0.3
	H3	C-terminal tail	ARG 134		2.0±3.9
3	H2A	Docking domain	ARG 99	THR 99	4.7±1.7
	H2A	Docking domain	LEU 115	Absent in H2A.B	6.3±0.6
	H2A	Docking domain	LYS 118	Absent in H2A.B	-0.9±3.8
	H2A	Docking domain	LYS 119	Absent in H2A.B	-0.6±2.2
	H2A	C-terminal tail	GLU 121	Absent in H2A.B	3.8±4.1
	H3	α_2 Helix	ASN 108		0.0±0.2
	H4	L1-Loop	LYS 44		0.0±0.5

Table 2: Residues with the most significant differences in interaction energies in the three regions identified in Figure 2 at the dimer/tetramer interface ($\Delta\Delta E$) between H2A and H2A.B containing nucleosomes. Negative values correspond to locations that are more favorable in H2A.B systems, whereas positive values are more favorable in H2A nucleosomes.

domain. It has an energy difference of $\Delta\Delta E_2 = 1.1 \pm 1.2$ kcal/mol, which is not statistically significant but does suggest there may be a relatively minor destabilization in this region. The most significant energy difference is in region three, which includes Arg99, Leu115, Lys118, and Lys119 of the H2A docking domain, Glu121 of the H2A C-terminal, Asn108 of the H3 α_2 helix, and Lys44 of the H4 L1-loop. Residues Leu115 through Glu121 are absent in H2A.B while Arg99 of H2A is mutated to a threonine. This results in a significant destabilization of the dimer-tetramer interface in this region, with an energy difference of $\Delta\Delta E_3 = 13.3 \pm 3.8$ kcal/mol.

A similar energetic decomposition was performed for residues at the dimer core/DNA interface. In canonical nucleosomes, the major points of contacts between H2A and DNA occur in the α_1 helix, the L1, and the L2 loops, along with Arg17 in the αN helix near the N-terminal tail (Figure 3). In each of these regions there are strong stabilizing contacts formed, particularly between Arg42, Thr76, and Arg77 with the DNA (Table 3). Substitution of H2A.B abolishes both the α_1 and L2 loop contacts while maintaining the L1 contacts (Table 4). This results in roughly similar interactions at the base of the nucleosome opposite of the dyad, but reduced contacts closer to the entry and exit DNA sites (Figure 3). Notably, by changing the L2 loops from

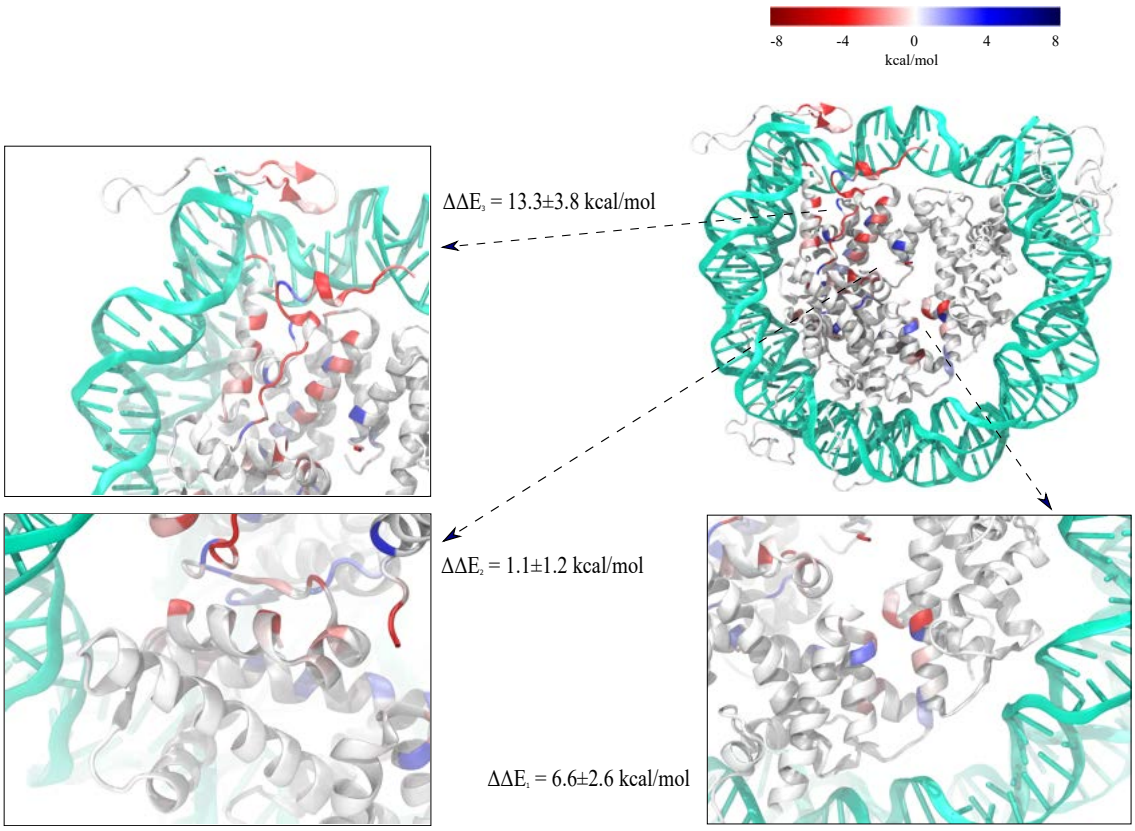

Figure 2: Nucleosome regions that have the most significant change in interaction energies upon H2A.B incorporation at the dimer/tetramer interface, along with their energy changes (kcal/mol). These regions correspond to (1) the H2B α_2/H4, (2) the H2A/H3-H4, and (3) the H2A docking domain/H3 interfaces. Positive values indicate unfavorable energy changes upon H2A.B substitution.

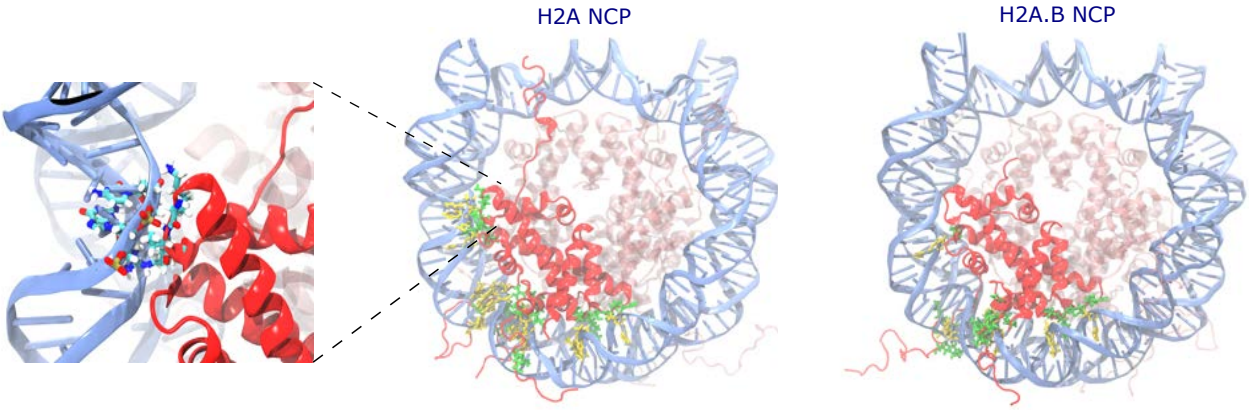

Figure 3: Stabilizing DNA/dimer contacts in H2A and H2A.B containing systems. Protein residues and DNA basepairs forming contacts are shown in green and yellow, respectively. Residues at the H2A/H2A.B L2 loops (see zoomed-in figure on the left) have the largest difference in stabilizing contacts upon variant substitution. Interaction energies for specific residues tables 3 and 4.

having a net charge of +2 to neutral upon H2A.B substitution, the DNA interaction energy of these loops is significantly destabilized from -10.7±1.5 kcal/mol to -3.1±1.9 kcal/mol. Contacts between H2B and DNA were similar in both systems, with

H2A NCP					
H2A Histone			H2B Histone		
Residue	Location	ΔE_{total} (kcal/mol)	Residue	Location	ΔE_{total} (kcal/mol)
ARG 17	αN Helix	-7.4±1.1	SER 52	L1 Loop	-1.7±0.6
GLY 28	α_1 Helix	-0.6±0.5	SER 53	α_2 Helix	-3.2±1.1
ARG 29	α_1 Helix	-3.1±1.2	ARG 83	L2 Loop	-4.7±1.2
ARG 42	L1 Loop	-4.9±1.9	SER 84	L2 Loop	-3.4±1.3
VAL 43	L1 Loop	-0.6±0.6	THR 85	L2 loop	-6.3±1.7
GLY 44	L1 Loop	-1±0.3			
LYS 75	L2 Loop	-1.8±1.4			
THR 76	L2 Loop	-4.5±1.5			
ARG 77	L2 Loop	-4.4±1.6			

Table 3: Residues with the most significant interaction energies at the dimer/DNA interface in H2A systems (kcal/mol). Negative values correspond to interactions that stabilize DNA binding.

H2A.B NCP					
H2A.B Histone			H2B Histone		
Residue	Location	ΔE_{total} (kcal/mol)	Residue	Location	ΔE_{total} (kcal/mol)
ARG 17	αN Helix	-7.7±0.9	SER 52	L1 Loop	-1.3±0.8
THR 18	αN Helix	-0.2±0.2	SER 53	α_2 Helix	-2.9±1.2
ARG 42	L1 Loop	-4.7±1.2	ARG 83	L2 Loop	-4.4±1.1
LEU 43	L1 Loop	0.1±0.5	SER 84	L2 Loop	-3.7±1.6
SER 44	L1 Loop	-3±0.8	THR 85	L2 loop	-6±1.1

Table 4: Residues with the most significant interaction energies at the dimer/DNA interface in H2A.B systems (kcal/mol). Negative values correspond to interactions that stabilize DNA binding.

stabilizing contacts largely formed between the L2 loop and DNA.

Enhanced Sampling Simulations Show Increased DNA Opening in H2A.B Nucleosomes

Adaptively biased MD (ABMD) simulations were performed to overcome the limited timescales of conventional MD simulations which precluded sampling of open nucleosomal states. For each system, three ABMD simulations of 1.6 µs each were performed. The reaction coordinates for these calculations were the radius of gyration (Rg) of the entry and exit DNA (see methods for details). To obtain sufficient sampling, simulations were performed in an implicit solvent environment, and the H3 N-terminal tails were truncated, as they tend to collapse on the DNA and prevent DNA breathing (69). Convergence of these simulations was assessed by Kullback–Leibler divergence calculations to the final probability states which demonstrated that in most simulations there was little change over the last 600 ns (Figures (S4)) (70, 71). In the case of canonical nucleosomes, the maximum observed Rgs were 55 and 50 Å for the entry and exit DNA, indicating a moderate amount of DNA breathing (Figure 4). The potential of mean force (PMF) shows a single free-energy well located at 44.5 and 41.3 Å (Figure 5). This corresponds to states with slightly opened DNA structures, which we attribute to the implicit solvent environment and the lack of H3 tails.

In contrast, ABMD simulations with H2A.B nucleosomes experienced significantly more DNA opening. The maximum Rgs in these simulations were 66 and 57 Å for the entry and exit DNA, indicating that states with more DNA breathing from both the entry and exit sites were sampled (Figure 4). Although the absolute location of the free energy minima in the PMFs is

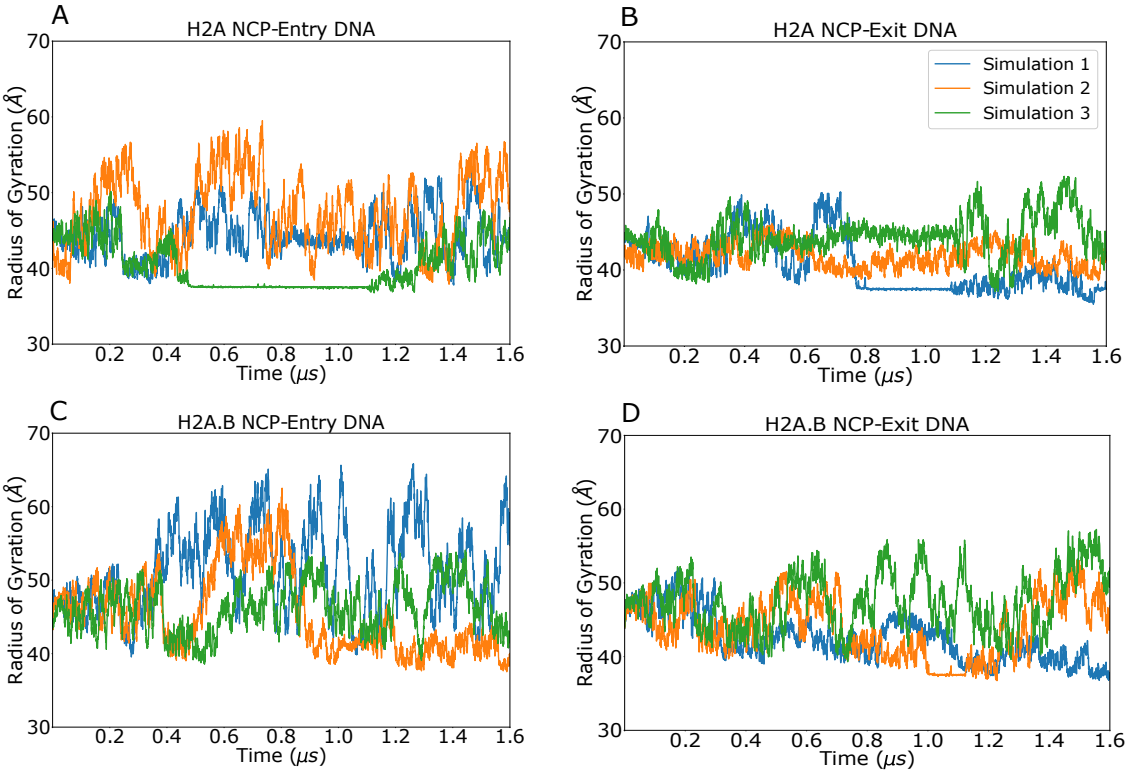

Figure 4: Radius of gyration for the entry and exit DNA segments in H2A and H2A.B nucleosomes during ABMD simulations. Both the entry and exit DNA segments in H2A.B systems have increased sampling and extensions relative to canonical nucleosomes which corresponds to more DNA breathing and increased DNA exposure.

nearly identical to the canonical nucleosome (46.5 and 44.2 Å for entry and exit DNA) the free energy wells are significantly broader and extend to higher Rg values, with low free energy states extending from 40 to 50 Å in each dimension.

Energetic analyses of the closed states simulations indicated that DNA-dimer interactions were significantly reduced in the H2A.B L2 loops upon variant substitution. In ABMD simulations, this destabilization resulted in increased separation between the DNA and H2A.B L2 loops with DNA. Indeed, in each of the three canonical simulations, the distance between the DNA and L2 loops remained at its initial value of ~13 Å, with the exception being the second ABMD simulation in which there were three detachment events where this distance increased (Figure 6). In contrast, in each of the H2A.B ABMD simulations the DNA/L2 Loop distance experienced multiple opening events for both the entry and exit DNA segments.This effect directly resulted in the frequent opening of up to 20 DNA basepairs from each side of the nucleosome.

To visualize the sampled states, a clustering analysis was performed on the trajectories for each system. Both Davis-Bouldin Index (DBI) and pseudo-F statistics (pSF) indicate that six clusters are optimal for describing each system (Figure S1). The representative structures from these clusters show that canonical nucleosomes adopt predominantly compact and semi-open states (Figure 7). These states are located primarily in the PMF free energy well, and have DNA Rg values between 37.5 and 49.8 Å (see Figures S5). H2A.B containing nucleosome structures represent more open states, with Rg values ranging from 41.4 to 56.2 Å (Figure 8 & S6). In both systems the histone core was relatively unchanged with RMSD values not exceeding 2

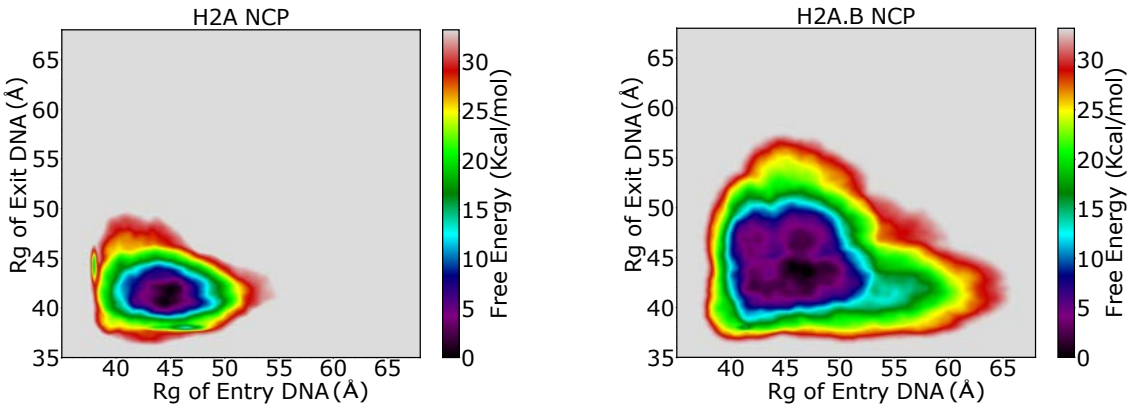

Figure 5: Free energy landscapes of canonical and H2A.B containing nucleosomes. Both systems have a single energy basin, however replacement of H2A with H2A.B results in a broader well and increased sampling of higher energy states.

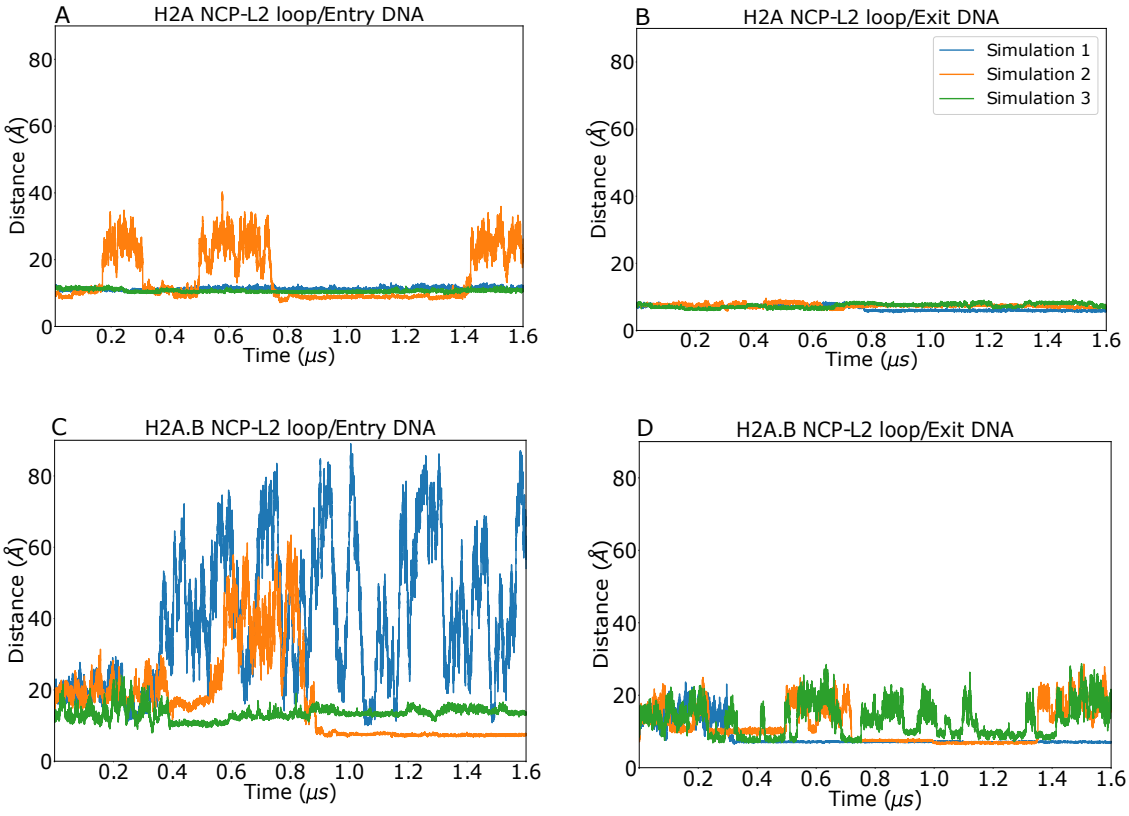

Figure 6: Separation distance between H2A (a&b) and H2A.B (c&d) L2 loops with their adjacent DNA basepairs in ABMD simulations. In canonical systems these distances are stable, indicating that the L2 loops remain in contact with the DNA throughout the majority of the simulations. In H2A.B systems there is a higher degree of detachment as the distances between the L2 loops and DNA sample a wide array of states throughout the simulations.

Å , demonstrating the histone core's stability in these simulations (Figure S7).

ABMD simulations revealed the presence of open canonical and H2A.B nucleosome states. However, to achieve the required level of sampling, simulations had to be performed in implicit solvent and without the N-terminal H3 tails. To analyze the

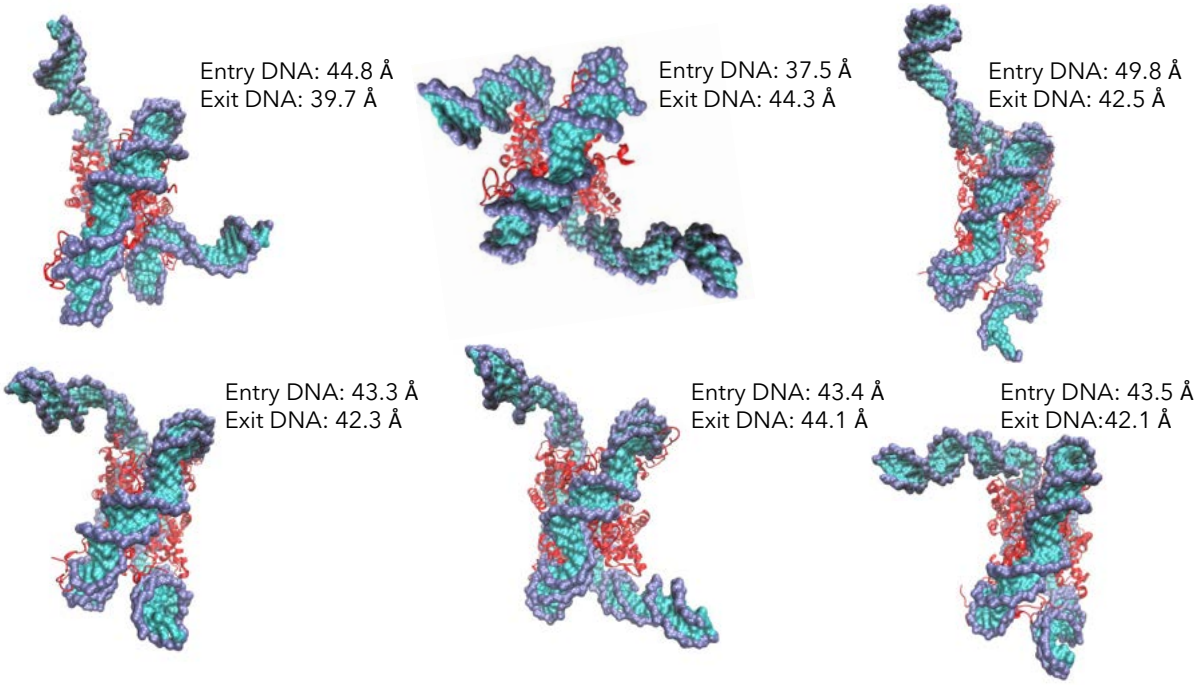

Figure 7: Structural clusters of H2A nucleosomes extracted from ABMD simulations. Structures represent close and semi-close states with raddi of gyration between 37.5 Å to 49.8 Å for entry DNA and from 39.7 Å to 44.3 Å for exit DNA.

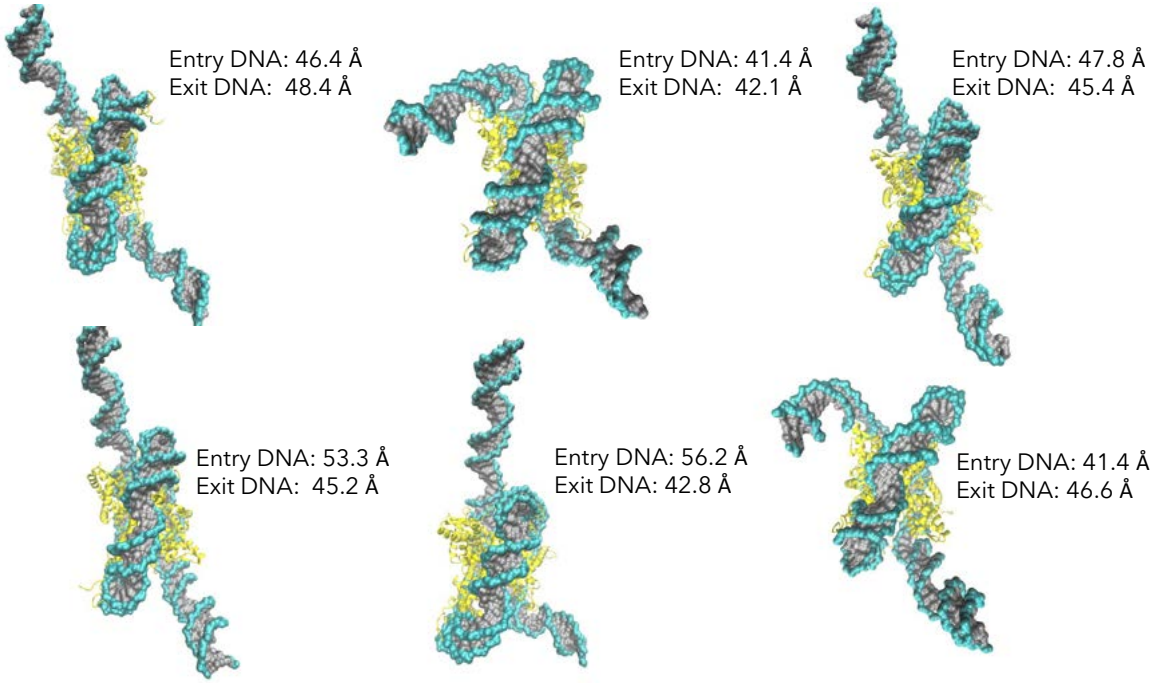

Figure 8: Structural clusters of H2A.B nucleosomes extracted from ABMD simulations. Structures represent semi-open and open states with raddi of gyration between 41.4 Å to 56.2 Å for entry DNA and from 42.1 Å to 48.4 Å for exit DNA.

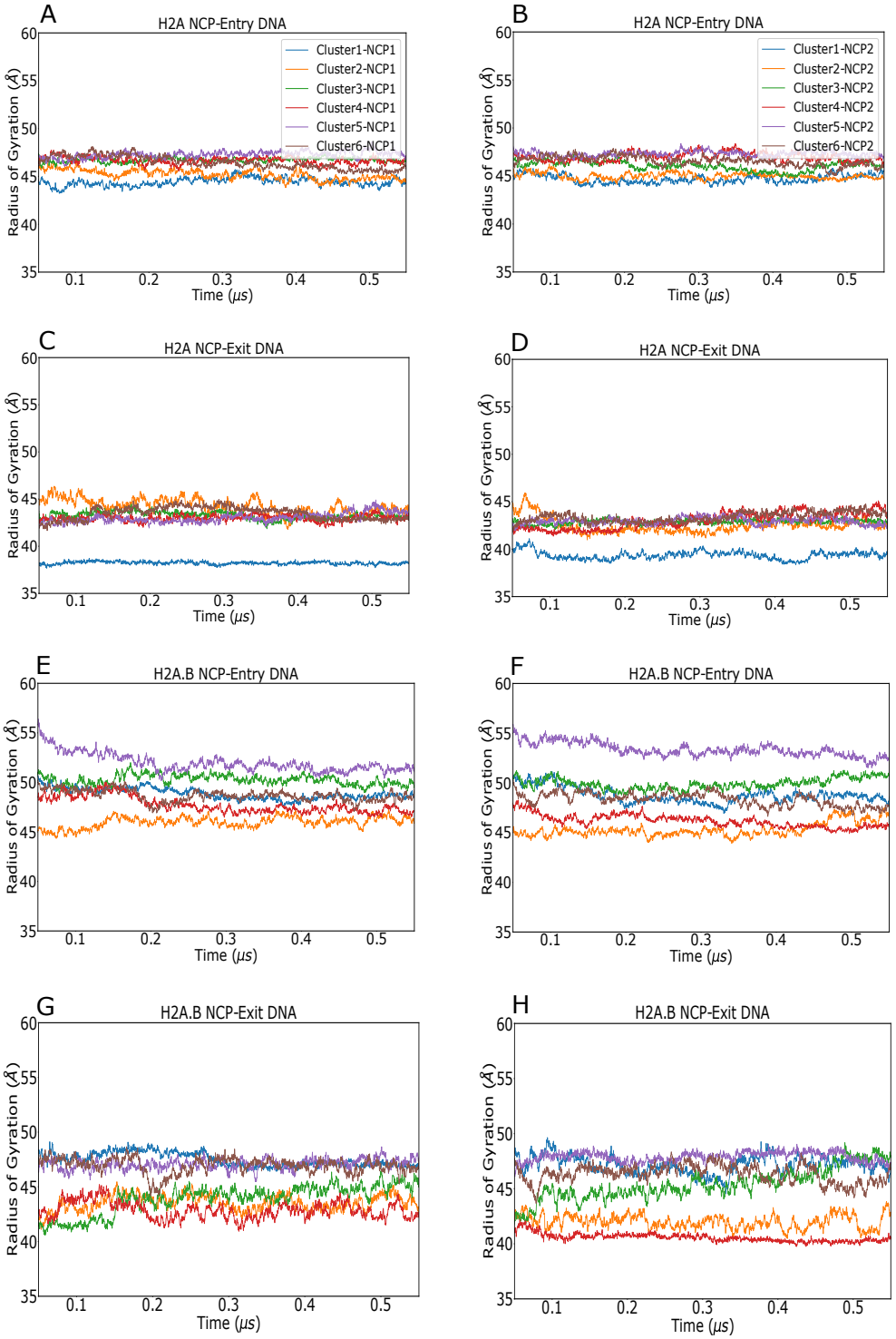

Figure 9: Radius of gyration for entry and exit DNA segments in explicit solvent/open state simulations for H2A (a-d) and H2A.B (e-h) systems from ABMD structural clustering (Figures 7 & 8). Two copies of each simulation were run (denoted as NCP1 and NCP2). In all simulations the radii of gyration were relatively constant, indicating the sampled open states are stable on the hundreds of nanonseconds timescale.

stability of these open states in more physiological conditions, the H3 tails were modelled onto the representative cluster centers from ABMD calculations and these structures were resolvated. For each structure, two 550 ns explicit MD simulations were performed resulting in 6.6 μs of sampling. In each simulation the histone core remained relatively stable, with RMSD values that did not exceed 4 Å (Figures S8). In all open simulations the DNA remained in their initial open states as evidenced by stable radii of gyration for the entry and exit DNA segments (Figure 9). Taken together, this confirms the stability of these open states in these more physiological conditions on the hundreds of nanoseconds timescale.

DISCUSSION

In this work, we have used a combination of conventional and enhanced sampling simulations to discern the effects of H2A.B substitution on nucleosome structures and dynamics. Our results show that H2A.B primarily alters intra-nucleosome interactions in two regions: the DNA-dimer and dimer-tetramer interfaces. At the DNA-dimer interface, residues in the H2A.B L2 loops appear to have the most significant influence on DNA stability, as the canonical Lys-Thr-Arg sequence is replaced with Glu-Arg-Asn in H2A.B. This substitution reduces the L2 loop's net charge from +2 to 0 and its ability to form stabilizing hydrogen bonds with the DNA. As we observed in enhanced sampling simulations, this point of contact was routinely broken in H2A.B nucleosomes, resulting in sampling of more elongated states with DNA opening by up to 20 basepairs on each side. These open state are stable on the hundreds of nanonseconds timescales, as exemplified by our conventional simulations of the dominant structural clusters. This change to the L2 loops appears to be unique to the H2A.B variant, as macroH2A, H2A.X, and H2A.Z all have two basic and no acidic residues in this sequence (11). Therefore, mutations targeting the basic residues in the L2 loops are likely to have significant effects on the free energy difference between open and closed nucleosome states.

The effects of H2A.B modifications at the dimer/tetramer interface on nucleosome structures are less clear. Although an energetic analysis showed a significantly weaker interface at the docking domain, which is shorter and less basic in H2A.B, all conventional and enhanced sampling simulations performed here showed the nucleosome core remaining intact with relatively minor structural rearrangements. It has been proposed that dimers can exhibit a butterfly opening motion which would occur on timescales that are far beyond what are sampled here (72). Furthermore, fluorescence recovery after photobleaching (FRAP) experiments have shown that H2A.B increases the rate of dimer loss and exchange *in vivo* (73). The weaker dimer/tetramer interactions we noted may results in enhancement of these butterfly motions and are likely one of the driving forces of this increased dimer exchange, which may further enhance transcription and nucleosome disassembly processes which utilize subnucleosomal particles (74). However, at equilibrium, our simulations, as well as previously performed SAXS experiments(28), suggest that dominant conformation of H2A.B systems *in vitro* are intact nucleosome cores with DNA existing in a dynamic equilibrium of open states. This weaker interface may also make H2A.B systems more susceptible to disassembly by post-translational modifications or changes in solution conditions (61, 75, 76).

AUTHOR CONTRIBUTIONS

H.K. performed the simulations. H.K. and J.W. designed the experiments, analyzed data, and wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank Dr. Samuel Bowerman, Mr. Joseph Clayton, and Mr. Dustin woods for valuable discussions concerning this work. This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health (Grants 1R15GM114758 & 1R35GM119647). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Kornberg, R. D., 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* 184:868–871.

2. McGinty, R. K., and S. Tan, 2015. Nucleosome structure and function. *Chem. Rev.* 115:2255–2273.

3. Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder, and 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.

4. Olins, A. L., and D. E. Olins, 1974. Spheroid chromatin units (v bodies). *Science* 183:330–332.

5. Baldi, S., P. Korber, and P. B. Becker, 2020. Beads on a string-nucleosome array arrangements and folding of the chromatin fiber. *Nat. Struct. Mol. Biol.* 27:109–118.

6. Wagner, F. R., C. Dienemann, H. Wang, A. Stutzer, D. Tegunov, H. Urlaub, and P. Cramer, 2020. Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature* 579:448–451.

7. Moller, J., and J. J. de Pablo, 2020. Bottom-Up Meets Top-Down: The Crossroads of Multiscale Chromatin Modeling. *Biophys. J.* 118:2057–2065.

8. Koellner, C. M., K. A. Mensink, and W. E. Highsmith, 2018. Chapter 5 - Basic Concepts in Human Molecular Genetics. *In* W. B. Coleman, and G. J. Tsongalis, editors, Molecular Pathology (Second Edition), Academic Press, 99 – 120. Second edition.

9. Talbert, P. B., K. Ahmad, G. Almouzni, J. Ausio, F. Berger, P. L. Bhalla, W. M. Bonner, W. Z. Cande, B. P. Chadwick, S. W. Chan, G. A. Cross, L. Cui, S. I. Dimitrov, D. Doenecke, J. M. Eirin-Lopez, M. A. Gorovsky, S. B. Hake, B. A. Hamkalo, S. Holec, S. E. Jacobsen, K. Kamieniarz, S. Khochbin, A. G. Ladurner, D. Landsman, J. A. Latham, B. Loppin, H. S. Malik, W. F. Marzluff, J. R. Pehrson, J. Postberg, R. Schneider, M. B. Singh, M. M. Smith, E. Thompson, M. E.

Torres-Padilla, D. J. Tremethick, B. M. Turner, J. H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu, and S. Henikoff, 2012. A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* 5:7.

10. Brown, D. T., 2001. Histone variants: are they functionally heterogeneous? *Genome Biol.* 2:REVIEWS0006.

11. Bönisch, C., and S. B. Hake, 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res.* 40:10719–10741.

12. Kamakaka, R. T., and S. Biggins, 2005. Histone variants: deviants? *Genes Dev.* 19:295–310.

13. Henikoff, S., and M. M. Smith, 2015. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* 7:a019364.

14. Musselman, C. A., M. E. Lalonde, J. Cote, and T. G. Kutateladze, 2012. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* 19:1218–1227.

15. Probst, A. V., B. Desvoyes, and C. Gutierrez, 2020. Similar yet Critically Different: The distribution, dynamics and function of Histone Variants. *J. Exp. Bot.* .

16. Bowman, G. D., and M. G. Poirier, 2015. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* 115:2274–2295.

17. Tessarz, P., and T. Kouzarides, 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* 15:703–708.

18. Kebede, A. F., R. Schneider, and S. Daujat, 2015. Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *FEBS J.* 282:1658–1674.

19. Jack, A. P., and S. B. Hake, 2014. Getting down to the core of histone modifications. *Chromosoma* 123:355–371.

20. Jiang, X., T. A. Soboleva, and D. J. Tremethick, 2020. Short Histone H2A Variants: Small in Stature but not in Function. *Cells* 9.

21. Shah, S., T. Verma, M. Rashid, N. Gadewal, and S. Gupta, 2020. Histone H2A isoforms: Potential implications in epigenome plasticity and diseases in eukaryotes. *J. Biosci.* 45.

22. Osakabe, A., Z. J. Lorkovic, W. Kobayashi, H. Tachiwana, R. Yelagandula, H. Kurumizaka, and F. Berger, 2018. Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Res.* 46:7675–7685.

23. Molaro, A., J. M. Young, and H. S. Malik, 2018. Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. *Genome Res.* 28:460–473.

24. Draizen, E. J., A. K. Shaytan, L. Marino-Ramirez, P. B. Talbert, D. Landsman, and A. R. Panchenko, 2016. HistoneDB 2.0: a histone database with variants–an integrated resource to explore histones and their variants. *Database (Oxford)* 2016.

25. Anuar, N. D., S. Kurscheid, M. Field, L. Zhang, E. Rebar, P. Gregory, T. Buchou, J. Bowles, P. Koopman, D. J. Tremethick, and T. A. Soboleva, 2019. Gene editing of the multi-copy H2A.B gene and its importance for fertility. *Genome Biol.* 20:23.

26. Soboleva, T. A., B. J. Parker, M. Nekrasov, G. Hart-Smith, Y. J. Tay, W. Q. Tng, M. Wilkins, D. Ryan, and D. J. Tremethick, 2017. A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS Genet.* 13:e1006633.

27. Arimura, Y., H. Kimura, T. Oda, K. Sato, A. Osakabe, H. Tachiwana, Y. Sato, Y. Kinugasa, T. Ikura, M. Sugiyama, M. Sato, and H. Kurumizaka, 2015. Corrigendum: Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* 5:9628.

28. Sugiyama, M., Y. Arimura, K. Shirayama, R. Fujita, Y. Oba, N. Sato, R. Inoue, T. Oda, M. Sato, R. K. Heenan, and H. Kurumizaka, 2014. Distinct features of the histone core structure in nucleosomes containing the histone H2A.B variant. *Biophys. J.* 106:2206–2213.

29. Arimura, Y., H. Kimura, T. Oda, K. Sato, A. Osakabe, H. Tachiwana, Y. Sato, Y. Kinugasa, T. Ikura, M. Sugiyama, M. Sato, and H. Kurumizaka, 2013. Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* 3:3510.

30. Chadwick, B. P., and H. F. Willard, 2001. Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum. Mol. Genet.* 10:1101–1113.

31. Tolstorukov, M. Y., J. A. Goldman, C. Gilbert, V. Ogryzko, R. E. Kingston, and P. J. Park, 2012. Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol. Cell* 47:596–607.

32. Gonzalez-Romero, R., J. Mendez, J. Ausio, and J. M. Eirin-Lopez, 2008. Quickly evolving histones, nucleosome stability and chromatin folding: all about histone H2A.Bbd. *Gene* 413:1–7.

33. Sansoni, V., C. S. Casas-Delucchi, M. Rajan, A. Schmidt, C. Bonisch, A. W. Thomae, M. S. Staege, S. B. Hake, M. C. Cardoso, and A. Imhof, 2014. The histone variant H2A.Bbd is enriched at sites of DNA synthesis. *Nucleic Acids Res.* 42:6405–6420.

34. Bao, Y., K. Konesky, Y. J. Park, S. Rosu, P. N. Dyer, D. Rangasamy, D. J. Tremethick, P. J. Laybourn, and K. Luger, 2004. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J.* 23:3314–3324.

35. Doyen, C. M., F. Montel, T. Gautier, H. Menoni, C. Claudet, M. Delacour-Larose, D. Angelov, A. Hamiche, J. Bednar, C. Faivre-Moskalenko, P. Bouvet, and S. Dimitrov, 2006. Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J.* 25:4234–4244.

36. Dai, L., X. Xie, and Z. Zhou, 2018. Crystal structure of the histone heterodimer containing histone variant H2A.Bbd. *Biochemical and Biophysical Research Communications* 503:1786 – 1791.

37. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* 117:5179–5197.

38. Joung, I. S., and T. E. Cheatham, 2009. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J Phys Chem B* 113:13279–13290.

39. Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713.

40. Cheatham, T. E., and D. A. Case, 2013. Twenty-five years of nucleic acid simulations. *Biopolymers* 99:969–977.

41. Case, D. A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, 2005. The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.

42. Feller, S. E., Y. Zhang, R. W. Pastor, and B. R. Brooks, 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* 103:4613–4621.

43. Salomon-Ferrer, R., A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker, 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* 9:3878–3888.

44. Feig, M., A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, 2004. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry* 25:265–284.

45. Onufriev, A., D. Bashford, and D. A. Case, 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55:383–394.

46. Hopkins, C. W., S. Le Grand, R. C. Walker, and A. E. Roitberg, 2015. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* 11:1864–1874.

47. Moradi, M., V. Babin, C. Roland, and C. Sagui, 2015. The Adaptively Biased Molecular Dynamics method revisited: New capabilities and an application. *Journal of Physics: Conference Series* 640:012020.

48. Davies, D. L., and D. W. Bouldin, 1979. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1:224–227.

49. Scott, A. J., and M. Knott, 1974. A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics* 30:507–512.

50. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.

51. Miller, B. R., T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, 2012. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* 8:3314–3321.

52. Sathyapriya, R., M. S. Vijayabaskar, and S. Vishveshwara, 2008. Insights into Protein DNA-Interactions through Structure Network Analysis. *PLOS Computational Biology* 4:1–15.

53. Roe, D. R., and T. E. Cheatham, 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 9:3084–3095.

54. Humphrey, W., A. Dalke, and K. Schulten, 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14:33 – 38.

55. 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.

56. Biswas, M., J. Langowski, and T. C. Bishop, 2013. Atomistic simulations of nucleosomes. *WIREs Computational Molecular Science* 3:378–392.

57. Ikebe, J., S. Sakuraba, and H. Kono, 2016. H3 Histone Tail Conformation within the Nucleosome and the Impact of K14 Acetylation Studied Using Enhanced Sampling Simulation. *PLoS Comput. Biol.* 12:e1004788.

58. Shaytan, A. K., G. A. Armeev, A. Goncearenco, V. B. Zhurkin, D. Landsman, and A. R. Panchenko, 2016. Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. *J. Mol. Biol.* 428:221–237.

59. Bowerman, S., and J. Wereszczynski, 2016. Effects of MacroH2A and H2A.Z on Nucleosome Dynamics as Elucidated by Molecular Dynamics Simulations. *Biophys. J.* 110:327–337.

60. Woods, D. C., and J. Wereszczynski, 2020. Elucidating the influence of linker histone variants on chromatosome dynamics and energetics. *Nucleic Acids Res.* 48:3591–3604.

61. Chen, Y., J. M. Tokuda, T. Topping, S. P. Meisburger, S. A. Pabit, L. M. Gloss, and L. Pollack, 2017. Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. *Proceedings of the National Academy of Sciences* 114:334–339.

62. Hansen, J. C., K. E. van Holde, and D. Lohr, 1991. The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence. *J. Biol. Chem.* 266:4276–4282.

63. Hazan, N. P., T. E. Tomov, R. Tsukanov, M. Liber, Y. Berger, R. Masoud, K. Toth, J. Langowski, and E. Nir, 2015. Nucleosome Core Particle Disassembly and Assembly Kinetics Studied Using Single-Molecule Fluorescence. *Biophys. J.* 109:1676–1685.

64. Genheden, S., and U. Ryde, 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461.

65. Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

66. Sun, H., Y. Li, M. Shen, S. Tian, L. Xu, P. Pan, Y. Guan, and T. Hou, 2014. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys Chem Chem Phys* 16:22035–22045.

67. Homeyer, N., and H. Gohlke, 2012. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* 31:114–122.

68. Genheden, S., J. Kongsted, P. Soderhjelm, and U. Ryde, 2010. Nonpolar Solvation Free Energies of Protein-Ligand Complexes. *J Chem Theory Comput* 6:3558–3568.

69. Morrison, E. A., S. Bowerman, K. L. Sylvers, J. Wereszczynski, and C. A. Musselman, 2018. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* 7.

70. Kullback, S., and R. A. Leibler, 1951. On Information and Sufficiency. *Ann. Math. Statist.* 22:79–86.

71. McClendon, C. L., L. Hua, G. Barreiro, and M. P. Jacobson, 2012. Comparing Conformational Ensembles Using the Kullback–Leibler Divergence Expansion. *Journal of Chemical Theory and Computation* 8:2115–2126.

72. Gansen, A., S. Felekyan, R. Kuhnemuth, K. Lehmann, K. Toth, C. A. M. Seidel, and J. Langowski, 2018. High precision FRET studies reveal reversible transitions in nucleosomes between microseconds and minutes. *Nat Commun* 9:4628.

73. Gautier, T., D. W. Abbott, A. Molla, A. Verdel, J. Ausio, and S. Dimitrov, 2004. Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep.* 5:715–720.

74. Read, C. M., J. P. Baldwin, and C. Crane-Robinson, 1985. Structure of subnucleosomal particles. Tetrameric (H3/H4)2 146 base pair DNA and hexameric (H3/H4)2(H2A/H2B)1 146 base pair DNA complexes. *Biochemistry* 24:4435–4450.

75. Simon, M., J. A. North, J. C. Shimko, R. A. Forties, M. B. Ferdinand, M. Manohar, M. Zhang, R. Fishel, J. J. Ottesen, and M. G. Poirier, 2011. Histone fold modifications control nucleosome unwrapping and disassembly. *Proc. Natl. Acad. Sci. U.S.A.* 108:12711–12716.

76. Chen, Y., J. M. Tokuda, T. Topping, J. L. Sutton, S. P. Meisburger, S. A. Pabit, L. M. Gloss, and L. Pollack, 2014. Revealing transient structures of nucleosomes as DNA unwinds. *Nucleic Acids Research* 42:8767–8776.