Article

The Role of Histone Tails in the Nucleosome: A Computational Study

Jochen Erler,[1] Ruihan Zhang,[1] Loukas Petridis,[2] Xiaolin Cheng,[2] Jeremy C. Smith,[2] and Jörg Langowski[1,*]
[1]Division of Biophysics of Macromolecules, German Cancer Research Center, Heidelberg, Germany; and [2]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee

ABSTRACT Histone tails play an important role in gene transcription and expression. We present here a systematic computational study of the role of histone tails in the nucleosome, using replica exchange molecular dynamics simulations with an implicit solvent model and different well-established force fields. We performed simulations for all four histone tails, H4, H3, H2A, and H2B, isolated and with inclusion of the nucleosome. The results confirm predictions of previous theoretical studies for the secondary structure of the isolated tails but show a strong dependence on the force field used. In the presence of the entire nucleosome for all force fields, the secondary structure of the histone tails is destabilized. Specific contacts are found between charged lysine and arginine residues and DNA phosphate groups and other binding sites in the minor and major DNA grooves. Using cluster analysis, we found a single dominant configuration of binding to DNA for the H4 and H2A histone tails, whereas H3 and H2B show multiple binding configurations with an equal probability. The leading stabilizing contribution for those binding configurations is the attractive interaction between the positively charged lysine and arginine residues and the negatively charged phosphate groups, and thus the resulting charge neutralization. Finally, we present results of molecular dynamics simulations in explicit solvent to confirm our conclusions. Results from both implicit and explicit solvent models show that large portions of the histone tails are not bound to DNA, supporting the complex role of these tails in gene transcription and expression and making them possible candidates for binding sites of transcription factors, enzymes, and other proteins.

INTRODUCTION

The nucleosome, discovered in 1974 (1,2), constitutes the lowest-order packing of DNA. Since its discovery, considerable effort has been made to understand its structure and dynamics. High-resolution x-ray measurements have determined the nucleosome structure in detail to near-atomic resolution (1.9 Å) (3,4), but its dynamics remain elusive. The nucleosome consists of 147 basepairs (bp) of DNA wrapped around a protein core. The core is made up of two copies each of four histone proteins, H3, H4, H2A and H2B, resulting in an octamer (see Fig. 1). Although this packing fits the genomic DNA into the small volume of the cell nucleus, it also hinders gene expression and gene transcription. For access to genomic DNA by proteins controlling and expressing the genome, the nucleosome must be remodeled, and such remodeling must include unwrapping of DNA from the nucleosome core.

The positively charged N(C)-terminal tails of the histone proteins play a central role in this process; for a general discussion of histone tails, see Preez and Patterton (5). The compaction of chromatin into higher-order structures is mediated by the N-terminal tails of histones H3 and H4 through binding to DNA and/or to acidic regions on the histone octamer surface of neighboring nucleosomes (6–8). Knowing the structure and dynamics of these tails and their role in nucleosome architecture will help us to understand regulation of gene expression. Whether the tails are intrinsically disordered (9,10) or have stable secondary structures, as proposed by circular dichroism experiments (11) and theoretical studies (12,13), remains an open question. Intrinsically disordered histone tails might be able to change their conformation easily upon binding to other proteins or DNA, providing a framework to explain their multiple and complex roles.

Posttranslational modifications (PTMs) of the tails, i.e., acetylation, phosphorylation, and methylation, will change their structure and thus their function. Acetylation of H4-tail lysines results in reduced compactness of chromatin, allowing transcription factors to access the DNA in acetylated chromatin regions (14–16). A modification of histone tail H3 (K27M) has been correlated with the occurrence of cancer (17). A recently published study suggests that chromatin condensation on mitosis is triggered by a cascade of PTMs of the H3 and H4 tails (18).

The conformation and role of histone tails in the nucleosome have been examined in experimental studies using, e.g., circular dichroism (CD) (11,19) or NMR and hydrogen-deuterium (H/D) exchange (20–22). Since the structural propensity of individual amino acid residues strongly depends on their environment (23), electrostatic interactions between histone tails and the nucleosomal DNA may affect the structure adopted by the histone domains in the nucleosome. Whether histone tails bind to DNA, and whether this interaction stabilizes or destabilizes a specific

Submitted April 25, 2014, and accepted for publication October 15, 2014.
*Correspondence: joerg.langowski@dkfz-heidelberg.de
Editor: Nathan Baker.

0006-3495/14/12/2911/12 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.10.065

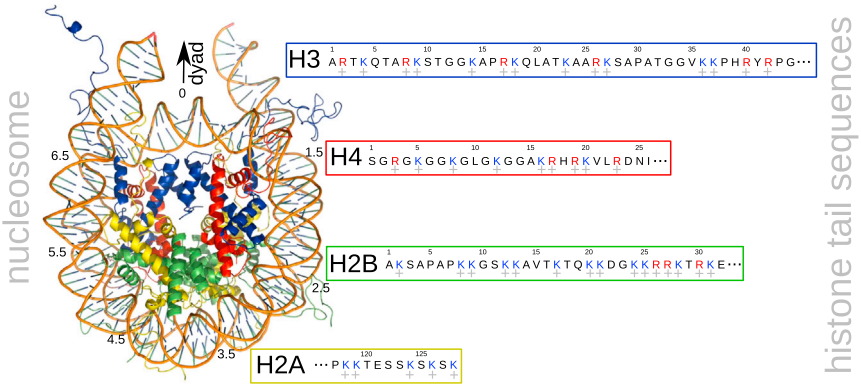

FIGURE 1 The nucleosome and histone tail sequences, showing 147 bp of DNA wrapped around a protein core. The protein core is made up of four histone proteins, H3, H4, H2A, and H2B (*blue*, *red*, *yellow*, and *green*, respectively). We extended the nucleosome by 10 bp of DNA on each side. Sequences of N-terminal tails of histones H3, H4, and H2B and the C-terminal tail of histone H2A are also shown, with charged residues lysine and arginine indicated in blue and red, respectively. To see this figure in color, go online.

secondary structure (19), is still unclear. Wang et al. concluded that although in native tails the α-helical content appears to depend on interaction with DNA, acetylated tails show an increase in α-helical content relative to native tails, independent of DNA interaction (11).

Experiments on single nucleosomes using H/D exchange NMR methods (21) found that residues 16–22 of the H4 tail bind to DNA whereas the rest, i.e., residues 1–15, are mobile. Furthermore, a large part of the H3 tail was found to be highly flexible and disordered, a finding also confirmed by results from nucleosome arrays (22). However, the binding of residues 16–21 of the H4 tail to DNA of single nucleosomes was not observed in the arrays (22), in which the whole H4 tail was found to be highly flexible and mobile, suggesting that these domains have somewhat different properties in single nucleosomes (low concentration) than in highly concentrated arrays. The finding that H3 and H4 tails in condensed nucleosome arrays are highly flexible and mobile indicates that chromatin compaction does not require specific protein-protein or protein-DNA contacts. However, the authors of an H/D exchange NMR study of 12-mer nucleosome arrays (20) concluded that the H3 tail forms stable folded structures in highly condensed chromatin fibers, leaving the question open.

Other studies using various coarse-grained models have focused on internucleosomal interaction and chromatin compaction, including the role of histone tails. Although it might be asked to what extent modeling histone tails as simple beads on a chain can provide a correct picture of their role in chromatin compaction, these studies have provided a preliminary idea of the dynamics of this complex system (24–29). Another coarse-grained study of DNA unwrapping showed long-lived DNA detachment (30), part of which involves binding of the H3 tail to the protein core, which blocks the DNA from rebinding to it. Removal of the H3 tails caused the long-lived detachment to disappear. This suggests that the H3 tail may stabilize DNA detachment during the initial stages of nucleosome remodeling.

There are a limited number of all-atom simulations (for a recent review, see Biswas et al. (31)), for example, the implicit solvent MD simulation by Ruscio and Onufriev (32), which focused on the flexibility of nucleosomal DNA, and the explicit solvent MD simulations of Roccatano et al. (33) and Biswas et al. (34), which examined the dynamics and the impact of a truncation of the histone tails. The truncation of the tails, similar to PTMs, resulted in a destabilization of the histone core, probably due to the absence of histone-histone and histone-DNA polar contacts (34). In a further study, by Ettig et al., steered MD simulations were used to investigate DNA-histone interactions and unwrapping of nucleosomal DNA (35).

Further all-atom studies focused on the structure of isolated histone tails (neglecting the nucleosome core particle) and the effect of PTMs (10,12,13). These studies used replica-exchange molecular dynamics (REMD) simulations together with explicit or implicit solvent models. The results suggested that specific secondary structures for histone tails: H3 and H2B were partly α-helical and that the H4 tail formed a stable β-sheet or α-helix. Modification of lysine residues was shown to alter the structure of the H3 and H4 N-terminal peptides. However, only one force field was used in these studies, making firm conclusions difficult, as secondary structural properties depend on the force field used (36,37). Furthermore, none of these studies included the nucleosome core particle and the effect of the negatively charged DNA surface on the tail structure was not considered. Two recent simulations of the H4 tail and part of the nucleosomal DNA focused on interactions of the H4 tail with DNA and the effect of PTM on H4 binding to the DNA surface (38,39). The latter study found that H4K16 acetylation partially orders the H4 tail and increases the probability of adopting α-helical configurations, resulting in stronger binding of the acetylated H4 tail to DNA compared to the nonacetylated form, perhaps contrary to what one might expect from electrostatic interactions. The difference in binding affinities was attributed to nonelectrostatic contributions, i.e., hydrophobic interactions, leading to more contacts with the DNA surface.

In this work, we study the role of histone tails in the nucleosome. Isolated tails tend to form a secondary

structure in the simulation that disappears in the presence of the nucleosome. We observe that the positively charged tail residues, lysine and arginine, contact the phosphate groups or bind to the minor or major groove of DNA, resulting in stable binding configurations of the tails on the DNA surface. Special emphasis is placed on testing the robustness of results to variation of the force field.

MATERIALS AND METHODS

Model structure

The nucleosome model contains the eight histone proteins with their tails and 147 bp of DNA (PDB ID 1KX5 (4)) plus 10 bp of linker DNA added at each end (for details, see Voltz et al. (30)). The choice of simulating the nucleosome core particle with 10-bp DNA linkers is motivated by the fact that H3 tails, which are at the entrance and exit points of the nucleosome, can also bind fully to DNA to model the behavior in oligonucleosomes, as described by Arya and Schlick (28). We simulated isolated histone tails and the whole nucleosome. The structure of the nucleosome and the sequences of the histone tails are shown in Fig. 1.

Simulation methods

We used REMD simulations to sample the configurational space (40–42). In REMD, multiple copies of the system are simulated at different temperatures and attempts are made routinely to exchange the configurations of two systems at neighboring temperatures based on the Monte Carlo Metropolis criterion. This enables the system to overcome high energy barriers between different states in conformational space (folded/unfolded or bound/unbound) and thus improves sampling. The simulations spanned a temperature range from 300 K to 550 K and the temperature spacing was chosen to reach a uniform exchange probability of ~0.1. Using a step size of 0.002 ps and an exchange attempt every 100 steps (0.2 ps) results in an expected exchange every 2 ps. To obtain this exchange rate, we had to simulate 8–12 replicas depending on the number of atoms in the simulated histone tail (for more details about each histone tail, see Fig. S1 in the Supporting Material). We checked the robustness of results for the H2A tail against variations in the REMD parameters, i.e., a higher exchange probability (0.2), and less frequent exchange attempts. Details of these simulations can be found in the Supporting Material.

To accelerate conformational sampling even more, we used the generalized Born solvation model developed by Onufriev et al. (43) to mimic solvent effects in the simulations. For the intrinsic radii, we used modified Bondi van der Waals radii (mbondi2) from the Amber package (44,45). The surface tension was set to 2.25936 kJ/mol/nm^2 (43). The exterior solvent dielectric constant was set to 78.5. The salt concentration was set to 0.15 M and thus mimics the physiological environment.

We used the GROMACS 4.5.7 package (46–49). The energy was minimized using the conjugate gradient method. To maintain the temperatures, a velocity rescaling with a stochastic term (50) was used (similar to Berendsen coupling), which ensures that a proper canonical ensemble is generated. The coupling constant, τ_t, was 2.0 ps. Conformations were saved every 10.0 ps. The LINCS algorithm (51) was applied to constrain the bonds to hydrogen atoms. The cutoff for nonbonded interactions and generalized Born pairwise summation was set to 1.5 nm. We performed two further REMD simulations for H2A and H3 using a cutoff of 2.5 nm to preclude a dependence of our results on the cutoff, as described in Anandakrishnan et al. (52). For both cutoffs, the tails bound to DNA and converged to the same number of DNA:tail contacts after 5 ns (see Supporting Material).

To save computational resources, we constrained the degrees of freedom of the whole nucleosome except the histone tail under investigation by defining them as a freeze group. Atoms that belong to this group were kept stationary in the simulation. Constraining the nucleosome results in a much smaller number of replicas needed to obtain the expected exchange probability. In this way, we include all interactions between the core and the tail residues and save resources at the same time. For H2A, we extended the tail (the nonfrozen part) by 13 residues (up to Q104) to give it more freedom to bind to the DNA surface. We also performed a test run relaxing a further 16 residues of the H4 histone to assess the robustness of the results with respect to the number of nonfixed residues (see the Supporting Material).

Simulations of all four histone tails were performed with and without the nucleosome core particle. The REMD simulations were conducted up to 160 ns for tails H3, H2A, and H2B and up to 120 ns for H4, depending on convergence. For analysis we have chosen the lowest temperature level of the REMD simulation, i.e., $T = 300$ K.

Currently available force fields are known to differ in their reproducing of experimental data and their predictions for observables like secondary structure. Some force fields are known to overstabilize certain types of secondary structure, e.g., helical structures with CHARMM27 (53) or Amber03 (54). For extensive studies comparing existing force fields, we refer the reader to Cino et al. (36) and Lindorff-Larsen et al. (37) or Best et al. (55). Further optimization of these force fields with direct use of the helical content of short peptides as fitting data resulted in an improvement of these deficiencies (see Tzanov et al. (56,), Huang and MacKerrell (57), and citations therein). It remains unclear whether these force fields provide a good description when extrapolating to the unknown region of intrinsically disordered proteins. This makes it necessary to test how robust the results are against variation of the force field. In our study, we were limited to force fields that also include parameters for nucleic acid. This limits our choice to different versions of Amber and Charmm force fields. Since first test runs of isolated histone tails using CHARMM27 have shown a high overstabilization of an α-helix, we excluded them. Finally, we limited our study to force fields Amber99SB (58), Amber03 (54), and Amber99SB-ILDN (59).

To examine the possible effect of explicit treatment of the solvent on the conclusions drawn from REMD with implicit solvent, we also performed MD simulations with an explicit solvent model (TIP3P). For this, we used the stable structure of the tails found in the REMD run as the initial structure. The water box was chosen such that the distances between the most distant atoms of the nucleosome and the box edges were 1.2 nm. The final size of the system, including water, was 594,113 atoms. The cutoff for nonbond interactions was set to 1.2 nm, and we used periodic boundary conditions and the particle mesh Ewald summation technique. Ions were added to neutralize the system, and further salt (NaCl) was added to represent 0.15 M ionic concentration. After the system was minimized using the conjugate gradient method, we equilibrated the system in the NVT ensemble (2 fs time step). Velocity rescaling with a stochastic term (50) and a coupling constant of $\tau_t = 1.0$ ps was used. After 200 ps of NVT simulation, the system was equilibrated for 1 ns in an NPT ensemble. The LINCS algorithm (51) was applied to constrain the bonds to hydrogen atoms. In contrast to the implicit REMD simulations, no further constraints were used in the explicit REMD simulations.

REMD temperature space and convergence

The temperature trajectories of the REMD calculation are shown in Fig. S1. For the four simulations, all trajectories visit all temperature levels, indicating a sufficient exchange between temperature levels.

All simulations converged with respect to secondary structure (see Fig. S10), yielding a 120 ns trajectory for H4 and 160 ns trajectories for H3, H2A, and H2B.

Data analysis

We have chosen the first 10 ns as the equilibration phase.

Definition of secondary structure

The secondary structure content of histone tails was calculated using the DSSP algorithm (60), which measures the backbone hydrogen bonding pattern.

Cluster analysis

Structures with pairwise root mean-square deviations below the 0.2 nm cutoff were clustered in the same group, and the cluster center was selected as the representative structure. The root mean-square deviation was calculated based on C_α atom coordinates.

Definition of contacts

To define contacts between histone tail residues and the DNA surface, we used the VMD measure contacts tool. The threshold distance for a contact was set to 0.3 nm. We used the same cutoff to study the neutralization of positively charged lysine and arginine residues and the negative phosphate groups of DNA. When a positive side chain made contacts with two phosphate groups, or one phosphate group was in contact with two side chains, we always chose the pair with the smaller distance to count the neutralized charges.

Hydrogen bonds

For an analysis of hydrogen bonds within the histone tails, we used as criteria a donor-acceptor distance of <0.35 nm and a hydrogen-donor-acceptor angle of $<30°$.

RESULTS AND DISCUSSION

Isolated tail results and comparison with previous results

We first focus on the isolated histone tails and compare our results with those of previous studies (10,12,13,39). Fig. 2 shows the secondary structure of the isolated histone tail H4 as a function of time from simulations using three force fields: Amber99SB (58), Amber99SB-ILDN (59), and Amber03 (54). Results for H2A and H2B can be found in the Supporting Material. Amber99SB-ILDN, and to some extent Amber99SB, forms a transient β-sheet in the H4 tail. However, although the β-sheet was stable in Potoyan and Papoian (10) using Amber99SB, it is not stable in our simulation of the isolated H4 tail. It disappears or switches to an α-helix for Amber99SB. Possible reasons for the higher stability found by Potoyan and Papoian (10) could be their use of explicit solvent or a shorter trajectory (60 ns). Results for the Amber03 force field in Fig. 2 clearly converge to stable α-helices between residues A15 and K20 and another one around residue 24, confirming the results reported in Yang and Arya (13) using Amber03. Amber03 seems to produce an energy surface with a distinct and clear minimum for an α-helix, whereas Amber99SB and Amber99SB-ILDN produce a more widely distributed configurational space for the secondary structure (showing a β-sheet, α-helix or coil).

Fig. 3 shows the secondary structure of the isolated H3 tail. As reported in Liu and Duan (12), using force field Amber03, residues K4 and K9 are part of an α-helix. However, in our study, this α-helix is not stable with force fields

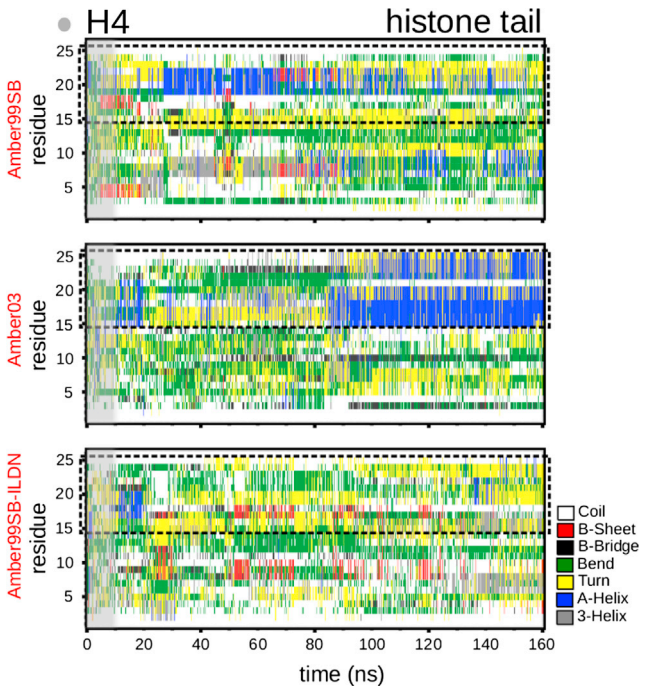

FIGURE 2 Secondary structure of the isolated H4 histone tail. Shown are results for the lowest temperature, 300 K, of the REMD calculation using Amber99SB (58), Amber03 (54), and Amber99SB-ILDN (59) force field. The gray area corresponds to 10 ns of equilibration. To see this figure in color, go online.

Amber99SB and Amber99SB-ILDN, as it disappears after 80 ns. Further results in Fig. 3 show a stable α-helix for all three force fields between residues 14 and 26 (*black box*), also confirming previous results (10).

Figs. S11 and S12 show the secondary structure of histone tails H2B and H2A, respectively. H2B forms a stable α-helix between residues 15 and 26 (*black box*) for Amber03 and Amber99SB-ILDN. Using AmberSB99 also tends to produce a helix in this region. The H2A tail forms a helix between residues 2 and 7 for all three force fields, in disagreement with the results of Potoyan and Papoian (10). A reason for this could again be the limited sampling in that study (10).

The results presented here indicate that it is risky to draw a conclusion about the secondary structure of histone tails based on only one force field, particularly for H4. The H4 tail may fold into an α-helix or a β-sheet, or may not have any ordered structure. H3, H2A, and H2B show a high tendency to form an α-helix for all tested force fields. The bias of certain force fields has been widely discussed in the literature (see, e.g., Lindorff-Larsen et al. (37)). Amber03 force field is known to overstabilize helical structures, whereas Amber99SB-ILDN is known to understabilize them (37). We therefore cannot draw firm conclusions regarding the stability of the secondary structures or in which configuration any given isolated tail exists. However, the results above are consistent with previous results for histone tails

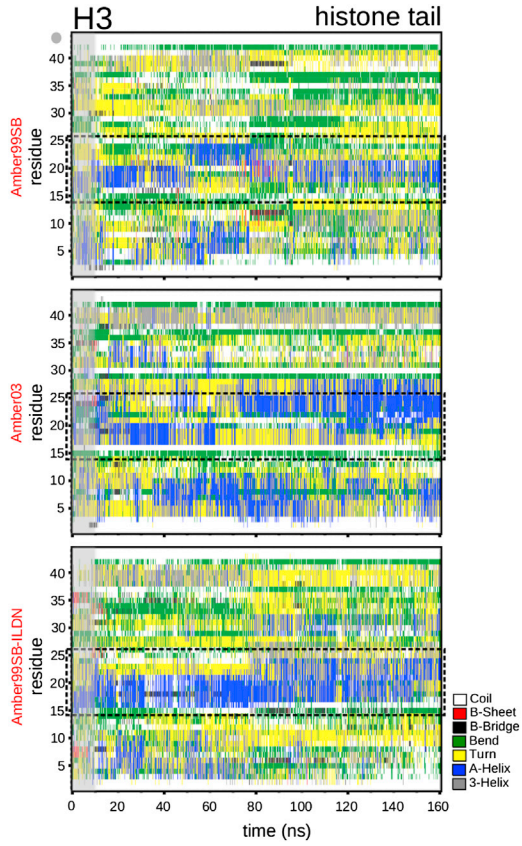

FIGURE 3 Secondary structure of the isolated H3 histone tail. Shown are results for the lowest temperature level, 300 K, of the REMD calculation using force field Amber99SB (58), Amber03 (54), and Amber99SB-ILDN (59). The gray area corresponds to 10 ns of equilibration. To see this figure in color, go online.

(12,13), as well as with previous studies using an explicit solvent model (10), and they should be considered a benchmark justifying our methods (REMD and implicit solvent).

Secondary structure of histone tails on the nucleosome

Histone tails are known to play an important role in nucleosome dynamics and thus in gene expression and transcription. We expect that the rest of the nucleosome, especially the negatively charged surface of the DNA, has a strong effect on the structure of the N(C)-terminal tails. Fig. 4 illustrates the secondary structure of all four histone tails in simulations of the whole nucleosome using force field Amber99SB (58). The configurations of all histone tails are dominated by bends (*green*), turns (*yellow*), and coils (*white*). In particular, H4 shows a stable bend for residues 7–11 over almost the whole trajectory. To preclude that in this case the tail is trapped in a local minimum, Fig. S13 shows the RMSD, dihedral angles, end-to-end distance, and radius of gyration of this segment (for results for other tails, see the Supporting Material). A transient α-helix or a β-sheet can be seen, but none is stable, with the exception of the H2A tail. For H2A, coils and bends dominate initially up to 90 ns, but afterward, a stable 3_{10}-helix forms between residues (4), 5, and 7, corresponding to residues (T120), E121, S122, and S123 situated between charged residues K119 and K124 (see also Fig. 1). In the isolated H2A tail, exactly this part forms a stable α-helix (see Fig. S12). Here, the binding of the lysines to the DNA surface shifts the α-helix to a 3_{10}-helix, with only the latter structure being compatible with the constraints given by the charged DNA phosphate groups. The effects of the charged DNA surface are even stronger for the H3, H4, and H2B tails. For these, although a high tendency to adopt secondary structure was noted in simulations of isolated tails, no secondary structure is found when simulating the whole nucleosome. The presence of the DNA (the negatively-charged phosphate groups) destroys the secondary structure of these tails (also predicted by the coarse-grained study (29)). The reason for this is the attraction of the side chains of lysines and arginines to the DNA phosphate groups, which is not compatible with a secondary structure. The tail configuration is thus governed by the charged phosphate groups. An explanation of this will be given later in the text. This shows that the whole nucleosome has to be considered for an understanding of the role of tails in nucleosome dynamics and PTM.

To examine whether the above results depend on the force fields and the initial structures used, we repeated the simulations for the H4 tail starting from a conformation in which residues 3–19 participate in an α-helix (the preparation was done by running the nucleosome without DNA using the CHARMM27 force field). Fig. 5 shows the results of 80 ns REMD simulations using Amber99SB (58), Amber03 (54), and Amber99SB-ILDN (59). For all force fields, the initial α-helix disappears. In Amber99SB-ILDN and Amber03, a small helical content remains, but a stable helical configuration is not found. The higher structure content for Amber99SB-ILDN and Amber03 is mainly due to their having a higher content of turns compared to Amber99SB. The α-helical content disappears with an increasing number of DNA:tail contacts, but no direct correlation exists between these two (see Fig. S15).

The reason for the disappearance of the α-helix is that in the helical configuration shown in Fig. 5, the attractive interaction between phosphate groups and charged tail residues outcompetes the stabilization of the helix by hydrogen bonds. We show later that hydrogen bonds exist between histone tail residues and can be important for the stability of certain binding configurations of histone tails on the DNA surface, but do not result in an ordered structure such as an α-helix.

The secondary structure of histone tails derived from REMD simulation may be compared with the results of CD experiments. For H2A and H2B, we find no secondary structure, in agreement with experimental results presented by Banères et al. (19). However, our simulation results do

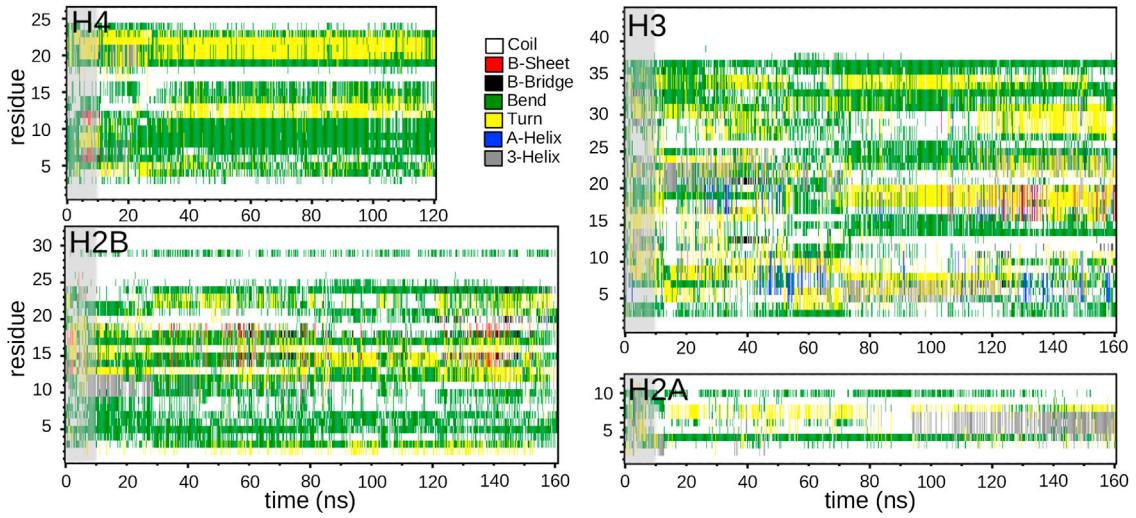

FIGURE 4 Secondary structure of H4, H3, H2B, and H2A histone tails. The whole nucleosome was simulated and the Amber99SB force field (58) was used. Different types of structure are colored as indicated. The gray area corresponds to 10 ns of equilibration. To see this figure in color, go online.

not agree with the finding in that study (19), as well as a study by Wang et al. (11), that H3 and H4 tails have an α-helical conformation attributable to their interaction with DNA. Nevertheless, a variation of the salt concentration in Wang et al. (11) did not support the conclusion that the H4 tail adopts an α-helical conformation upon binding to DNA, as the salt-dependent dissociation should lead to a decrease of the α-helical content.

Contacts and cluster analysis

To determine at which DNA sites the charged and other residues of the histone tails bind, Fig. 6 shows the probability of contacts between H4 tail residues and DNA phosphates during the last 40 ns of the trajectory (for other histone-tail binding probabilities, see Figs. S17–S19). H4 tail residues make specific contacts with the two DNA strands (see also Fig. 7). Most contacts are between arginine and lysine residues and DNA. Arginines and lysines are also found between two phosphate groups. The remaining contacts are made by threonine and glycine, which are in most cases next to a lysine or arginine. Examples for the H4 tail are: G2, G7, G9, G13, and G14 (Gly^{137}, Gly^{142}, Gly^{147}, Gly^{148}, and Gly^{149}). Fig. S16. shows a more detailed picture of these last 40 ns split into 5 ns intervals. The contacts are stable over the whole 40 ns interval apart from a

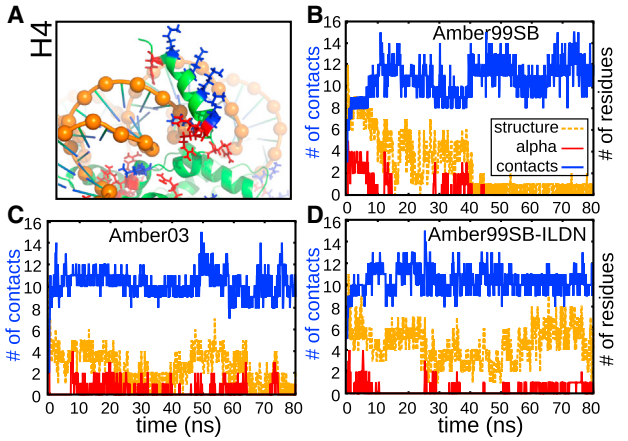

FIGURE 5 Independence of results from initial configuration and force field used. (*A*) H4 tail initial configuration, consisting of DNA (*orange*), phosphate groups (*orange beads*), histone proteins (*green*), lysine (*blue*), and arginine (*red*). (*B–D*) Number of contacts with DNA (atom-based) and number of residues participating in a structure (α-helix + β-sheet + β-bridge + turn) or in an α-helix. Used are the Amber99SB, Amber03, and Amber99SB-ILDN force fields. Data were averaged over 10 frames (0.1 ns). To see this figure in color, go online.

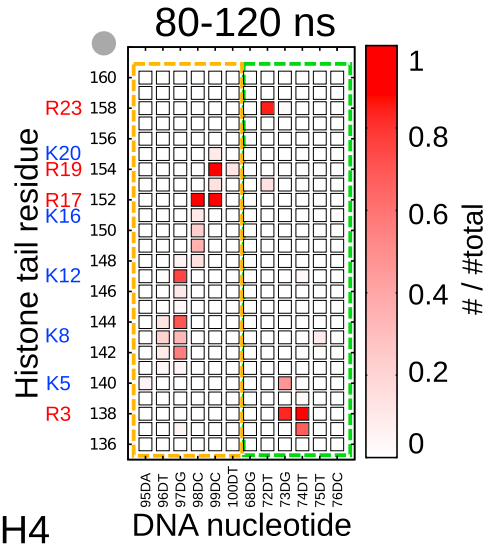

FIGURE 6 Probability of contacts between histone tail residues and DNA phosphate for the H4 tail. The last part of the trajectory (80–120 ns) was studied and the threshold for a contact was 0.3 nm. Colored boxes indicate different DNA strands. To see this figure in color, go online.

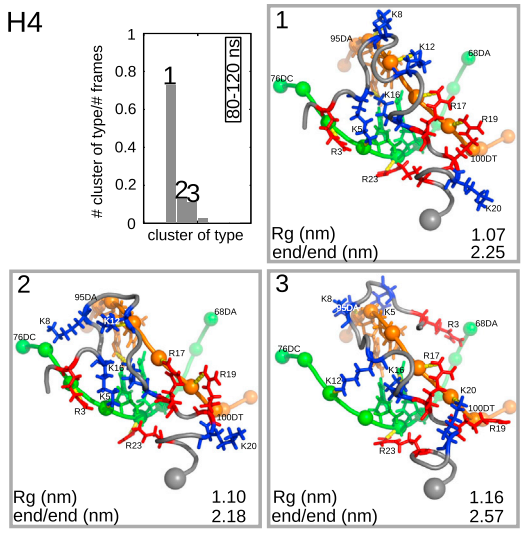

FIGURE 7 Results of a cluster analysis of the H4 tail trajectory. Probability and conformations of the cluster representative of the three most populated clusters are shown. The cutoff was 0.2 nm and the last part of the trajectory (80–120 ns) was used. The two DNA strands involved in binding of the H4 tail are colored orange and green. Spheres represent phosphate groups of DNA, and lysine and arginine residues are colored blue and red, respectively. The gray sphere is the end of the tail connected to the histone core. Yellow lines indicate interactions between the histone tail and the DNA. Radius of gyration, R_g, and end-to-end distance are given for each cluster. To see this figure in color, go online.

fast jump back and forth of K8, G7, and G8 to neighboring nucleotides.

The same is true for histone tails H2A, H2B, and H3. For the latter two tails, however, contacts are distributed over more residues, especially for the free end of the tail (Figs. S17 and S18). H3 and H2B contain more charged lysine or arginine residues, thus enabling more combinations with phosphate groups. Interestingly, there are also charged lysine and arginine residues that do not make contacts with a phosphate group, i.e., K16 of the H4 tail and K24 and R30 of the H2B tail. The side chains of these residues are found in the DNA groove.

It is interesting to examine whether there are preferred stable binding configurations of the histone tails on the nucleosome surface. Fig. 7 shows the result of a cluster analysis over the last 40 ns of the trajectory of the H4 tail, illustrating the probability distribution and conformations of the three most populated clusters. The radius of gyration, R_g, and end-to-end distance of these clusters are also indicated. As expected, the side chains of lysine and arginine residues make contacts with the DNA phosphate groups. We recall, however, that K16 is found in the DNA groove in all clusters (as was found also by Potoyan and Papoian (39)). Although contacts between charged side chains and the phosphate groups originate from charge-charge interactions, the origin of the binding of K16 to the groove is cation-π interactions with the surrounding basepairs (see Fig. S20). For a study of cation-π interactions in protein-DNA complexes, we refer the reader to Gromiha et al. (61). Distances and angles between the cationic group of K16 and the basepair rings of surrounding adenine and guanine are in the same range found for ions (62). For further details about distances and angles, see Fig. S20.

To understand the occurrence of a dominant stable cluster in the H4 trajectory, we mapped the distances between tail residues and phosphate groups for each cluster in Fig. 7 (see Fig. 8). Such maps for the H2A, H2B, and H3 histone tails can be found in Figs. S21–S29. Furthermore, we also characterized the hydrogen bonds between the tail residues. Table 1 shows the data for all clusters of H4, as well as the other histone tails, including the probability of each cluster, the radius of gyration, R_g, the number of neutralized charges, and the number of tail residues in a DNA groove. The data for the other histone tails are shown in the same table.

For the H4 tail, we find a total of 7, 6, and 6 neutralized charges for clusters 1–3, respectively. In addition, we find 9, 8, and 8 hydrogen bonds between tail residues. Thus, the number of bound residues for both types of interaction agrees with the ranking of clusters found for H4 in Fig. 7. Similar behavior is seen for histone tail H2A. Cluster 1 has a population of 0.71 and has all charged side chains neutralized by contacts to phosphate groups, whereas cluster 2 has a population of 0.20 with only four charges neutralized. In the latter configuration, the remaining charged residue K126 is found in the major groove of DNA, indicating that binding to a phosphate group is more favored than binding to the DNA groove. Comparing the radii of gyration one can conclude that the more charges are neutralized, the more compact are the configurations, since the repulsive interaction between the positively charged tail residues decreases.

The same applies to the results of the H2B tail. H2B contains many more charged residues than H4 and H2A (see above), enabling more combinations with phosphate groups and resulting in a broader distribution of the cluster probability (see Table 1). Comparison of the number of neutralized charges does not explain the ranking of clusters: cluster 2, with nine neutralized charges, has a slightly higher probability than clusters 3 or 4, which have 10 neutralized charges. Here, hydrogen bonds between histone tail residues seem to become important, and could explain the probability distribution of the clusters found.

For the H3 histone tail, 13, 12, and 12 of the 14 charged residues are bound to a phosphate group in clusters 1–3, respectively. This, together with the number of hydrogen bonds, explains the ranking of the clusters. The high number of hydrogen bonds seems to lead to the opposite trend in the radius of gyration for H3 compared to the other three tails.

Comparing distance maps for different clusters of the H4 tail in Fig. 8, one can see that between residues A15 and I26 (Ala^{150} and Ile^{161}), the maps do not differ. This indicates that this part of the tail is always bound to the DNA and

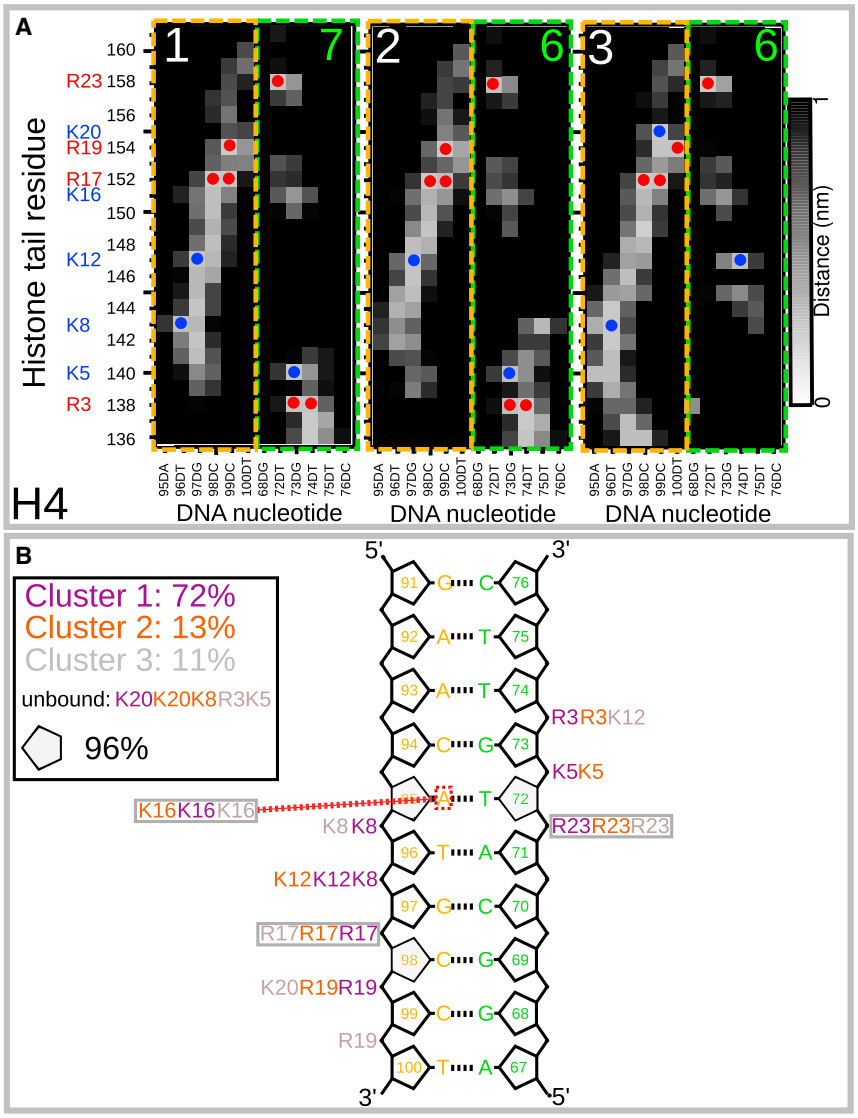

FIGURE 8 Key tail:DNA interactions of the H4 tail. (*A*) Distance map of tail residues and the phosphate group of each nucleotide for the three most populated clusters of the H4 tail. Orange and green boxes indicate the two strands of DNA, as in Fig. 7. Blue circles and red circles indicate contacts between lysines and arginines, respectively, and DNA phosphate (cutoff 0.3 nm). Green numbers denote the number of neutralized charges. (*B*) Schematic view of tail:DNA interactions. Gray boxes indicate interactions found in all three clusters. Red dashed line illustrate the cation-π interaction of K16. To see this figure in color, go online.

is less flexible than the rest of the tail. H4 residues are bound up to K16, which is also a common target for PTMs. This is consistent with the results of H/D exchange NMR measurements on the single nucleosome (21), in which residues 1–15 in H4 were found to be flexible, whereas residues 16–22 fold onto the nucleosome core. Furthermore, Zhou et al. (21) found that a K16Q mutation in H4, a mimic of K16 acetylation, leads to structural disorder of the basic patch (K16R17H18R19). One could expect that mutation of K16, which in our results is bound to the minor DNA groove, could lead to a similar destabilization of this basic patch. This could be the focus of a future study of PTMs of the H4 tail.

Zhou et al. (21) also found that a major part of the H3 tail, i.e., residues 3–36, was highly flexible (mobile). Gao et al. found that H3 tail residues 1–35 (38) are flexible and not attached to the DNA surface, both in nucleosome arrays and high densities (22). A similar conclusion can be drawn from our results for H3 in Fig. S22. Distance maps 1–3 agree well for residues 34–44, but differ for the rest of the residues.

Results for an explicit solvent model

To examine whether the above conclusions still hold in the presence of explicit solvent at physiological salt concentrations, we performed MD simulations of the whole nucleosome with explicit solvent and 0.15 M salt, focusing on the H4 tail. As the initial configuration for the H4 tail, we used a representative of cluster 1 in Fig. 7. Fig. 9 shows results for the secondary structure of the H4 tail (results for the other tails can be found in Fig. S30). In general, the explicit solvent MD trajectory does not exhibit stable secondary structure. Results for the structure of the implicit solvent REMD (see Fig. 4) and explicit solvent MD simulations are almost identical between K16 and I26.

TABLE 1 Results of a cluster analysis for all four histone tails

H4					
Cluster	Probability	H-bonds	R_g (nm)	Neutral (9)	Groove
1	0.72	9	1.07	7	1
2	0.13	8	1.10	6	1
3	0.11	8	1.16	6	1
H2A					
Cluster	Probability	H-bonds	R_g (nm)	Neutral (5)	Groove
1	0.71	2	0.75	5	0
2	0.20	2	0.87	4	1
H2B					
Cluster	Probability	H-bonds	R_g (nm)	Neutral (15)	Groove
1	0.29	10	1.06	11	2
2	0.27	12	1.11	9	2
3	0.22	11	1.08	10	2
4	0.11	6	1.09	10	2
H3					
Cluster	Probability	H-bonds	R_g (nm)	Neutral (14)	Groove
1	0.40	18	1.31	13	0
2	0.22	16	1.25	12	0
3	0.18	13	1.16	12	0

We present only clusters with a probability >0.1. Shown are number of hydrogen bonds between the histone tail residues (H-bonds), radius of gyration (R_g), and number of neutralized charges (Neutral) for each cluster representative, with the number of total positive charges on each histone tail in parentheses. The last column shows the number of residues found in the DNA groove.

Fig. 10 shows the results of cluster analysis of the simulation with explicit solvent. We split the trajectory into 40-ns-long segments. The cluster population evolves from a broad distribution, where the most populated cluster still resembles the starting configuration, to a distribution dominated by one single configuration with a probability of 0.94. Comparing this dominant conformation and the starting configuration in Fig. 7, one sees that some lysines and arginines are no longer bound to the phosphate groups. However, the basic patch K16R17H18R19, which was bound to DNA in the REMD results and in the study by Zhou et al. (21), remains attached after 120 ns of MD simulations with explicit water. This can be seen more clearly in the distance map in Fig. 10. In addition to K16, R17, and R19, R23 remains attached (compare the discussion above and the results of Zhou et al. (21)), yielding the same conformation for residues K16–I26 in the implicit as in the explicit solvent simulations (see Fig. 10 *B*). A similar result can be seen for the H3 tail from K36 to G44, confirming results reported by Gao et al. (22) and for the H2B tail from K378 to E397, since the charged residues remain attached to DNA (see the Supporting Material). In the explicit-solvent simulation, K16 of H4 is still bound in the minor groove. This again brings up the question of whether acetylation of K16 would lead to detachment of the whole H4 tail. In contrast, we find that K8, K12, and K20 are not bound to DNA (K20 is also unbound in the implicit REMD simulation). This supports the candidacy of these residues for involvement in interactions with other proteins. Results of Hirano et al., for example, suggest that the Lamin B receptor, an inner nuclear membrane protein, interacts with the dimethylated K20 (H4K20me2) (63), which is essential for chromatin compaction.

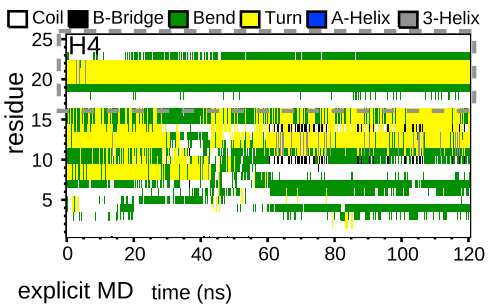

FIGURE 9 Secondary structure of the H4 tail of a MD simulation with explicit solvent. Force field Amber99SB (58) was used. Different types of structures are color-coded as indicated. The simulation was initialized with the most probable configuration identified by the cluster analysis in Fig. 7. To see this figure in color, go online.

Our simulation results disagree with an earlier experimental finding by Mangenot et al. (64), who observed a detachment of histone tails at physiological salt concentrations. They found, for a nucleosome containing 145 bp of DNA, a transition from a compact state, where all tails are attached to DNA, to a state where the tails are extended at 50 mM salt concentration. This transition was shifted to salt concentrations between 50 and 150 mM for nucleosomes with 165 bp, or to even higher concentrations for a longer DNA (65). We find a radius of gyration of $R_g = 4.3$ nm, slightly smaller than that of Mangenot and colleagues, $R_g = 4.6$ nm (64). The R_g value decreased during the first 20 ns, in agreement with the results of Roccatano et al. (33), and increased over the rest of the simulation time (see Fig. S34). Whether extending the simulation will increase the R_g further is not clear. Although the mesoscopic oligonucleosome model of Arya et al. (27) reproduced an increase in R_g, i.e., extension of the tails, with increasing salt concentration, no other computational study confirmed the extension of histone tails at physiological salt concentrations. Yang et al. (24), for example, found only a modest degree of detachment under a wide range of salt concentrations and concluded that a detachment is not induced by increased salt concentration but by the presence of other nucleosomes. However, all-atom MD simulations (33,35) find histone tails stably attached to DNA at physiological salt concentration, which may indicate a general problem with MD simulations caused by periodic boundary conditions and lattice-sum techniques.

CONCLUSIONS

In this study, we performed REMD simulations using an implicit solvent model for all four histone tails, H4, H3, H2A,

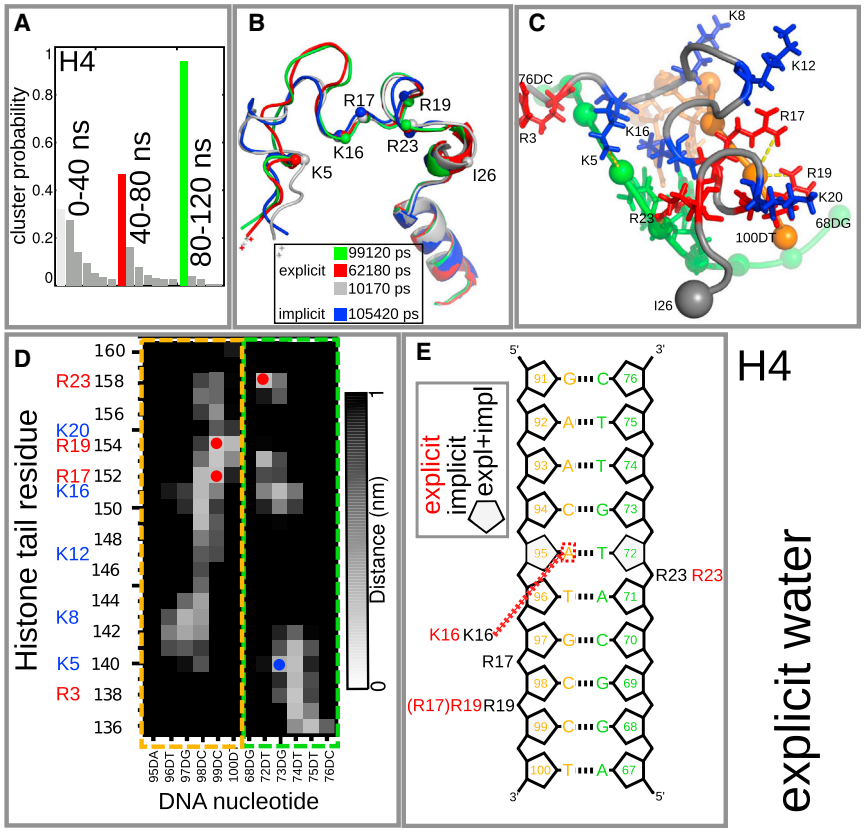

FIGURE 10 Results of the explicit MD simulation. (*A*) Evolution of cluster distribution. (*B*) Initial structure from implicit solvent REMD and cluster representatives of analysis in *A*. (*C–E*) Conformation (*C*), distance map (*D*), and DNA:tail interactions (*E*) of the cluster with highest population in the last time interval (including results from implicit solvent). To see this figure in color, go online.

and H2B. An important goal was to test results against variation of the force fields. Our results for the secondary structures of isolated histone tails confirm those of previous studies (10,12,13,39). We also show that for the H4 and H2B tails, the results depend strongly on the force field used. Further, we demonstrate that for most histone tails, a secondary structure is not compatible with the negatively charged DNA surface. An exception to this is H2A, which changes from an α-helix in the isolated tail to a 3_{10}-helix to fit onto the charged DNA surface. We also verified the robustness of the results for H4 by starting from an α-helical conformation, using different force fields, and using an explicit solvent model.

Results from contact maps, distance maps, and cluster analysis show that lysine and arginine residues make specific contacts with DNA. The positively charged side chains of histone tails were found at negatively charged phosphate groups or in the minor or major groove of the DNA, similar to the known behavior for ions (62). In cluster analysis, we found dominant binding conformations for H4 and H2A, whereas H3 and H2B show a broader probability distribution. The absence of a dominant binding conformation for the H3 tail supports the finding of Kan and Hayes (66) that H3 has multiple functions in chromatin compaction and regulation because it binds to neighboring nucleosomes and to the linker DNA of its own nucleosome. The H4 tail, however, is known to mediate internucleosomal interactions. We expect the binding behavior of the H4 tail in an oligonucleosome system to be different (see also Gao et al. (22)), which makes it difficult to draw a conclusion about general functions in the nucleosome from the binding configurations found. The ranking of the cluster, i.e., its population, is dominated mostly by the attractive interaction of lysine and arginine residues and phosphate groups, and thus the number of neutralized charges. However, hydrogen bonds between tail residues can also contribute to stabilizing the structures. Comparing the probabilities of clusters with different binding configurations of the H2A tail, we conclude that binding to a phosphate group is more favored than binding to the DNA groove.

Our results are compatible with H/D exchange NMR measurements on the single nucleosome (21) showing that part of the H4 tail is bound to DNA, whereas the rest is flexible and mobile. This part of the H4 tail stays attached to DNA in an explicit solvent MD simulation at physiological salt concentration. A similar result was found for the H3 tail. The flexibility and mobility of a major part of histone tails could enable the binding of enzymes or transcription factors, thereby explaining the complex role of these structures in nucleosome dynamics.

As a next step, we plan to investigate the role of PTMs. Further, spFRET measurements are planned to test our conclusions on conformations of histone tails and the binding of tails to the DNA surface.

SUPPORTING MATERIAL

Thirty-four figures and Supporting Materials and Methods are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)01156-4.

Resources for this research were provided by the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC05-00OR22725. Computational resources were also provided by the HELICS computer cluster at the Institute for Scientific Computing at the University of Heidelberg.

This work was supported by Deutsche Forschungsgemeinschaft under Grant LA 500/14-2.

REFERENCES

1. Olins, A. L., and D. E. Olins. 1974. Spheroid chromatin units (*ν* bodies). *Science.* 183:330–332.
2. Woodcock, C. L., L.-L. Frado, …, L. Ricciardiello. 1976. Fine structure of active ribosomal genes. *Chromosoma.* 58:33–39.
3. Luger, K., A. W. Mäder, …, T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 389:251–260.
4. Davey, C. A., and T. J. Richmond. 2002. DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl. Acad. Sci. USA.* 99:11169–11174.
5. Preez, L. L., and H.-G. Patterton. 2013. Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure. *In* Epigenetics: Development and Disease. T. K. Kundu, editor. Springer, New York, pp. 37–55.
6. Luger, K., and T. J. Richmond. 1998. The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8:140–146.
7. Dorigo, B., T. Schalch, …, T. J. Richmond. 2003. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* 327:85–96.
8. Zheng, C., X. Lu, …, J. J. Hayes. 2005. Salt-dependent intra- and internucleosomal interactions of the H3 tail domain in a model oligonucleosomal array. *J. Biol. Chem.* 280:33552–33557.
9. Hansen, J. C., X. Lu, …, R. W. Woody. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* 281:1853–1856.
10. 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.
11. Wang, X., S. C. Moore, …, J. Ausió. 2000. Acetylation increases the *α*-helical content of the histone tails of the nucleosome. *J. Biol. Chem.* 275:35013–35020.
12. Liu, H., and Y. Duan. 2008. Effects of posttranslational modifications on the structure and dynamics of histone H3 N-terminal peptide. *Biophys. J.* 94:4579–4585.
13. Yang, D., and G. Arya. 2011. Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation. *Phys. Chem. Chem. Phys.* 13:2911–2921.
14. Shogren-Knaak, M., H. Ishii, …, C. L. Peterson. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* 311:844–847.
15. Robinson, P. J., W. An, …, D. Rhodes. 2008. 30 nm Chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.* 381:816–825.
16. Allahverdi, A., R. Yang, …, L. Nordenskiöld. 2011. The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.* 39:1680–1691.
17. Schwartzentruber, J., A. Korshunov, …, N. Jabado. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature.* 482:226–231.
18. Wilkins, B. J., N. A. Rall, …, H. Neumann. 2014. A cascade of histone modifications induces chromatin condensation in mitosis. *Science.* 343:77–80.
19. Banères, J.-L., A. Martin, and J. Parello. 1997. The N tails of histones H3 and H4 adopt a highly structured conformation in the nucleosome. *J. Mol. Biol.* 273:503–508.
20. Kato, H., J. Gruschus, …, Y. Bai. 2009. Characterization of the N-terminal tail domain of histone H3 in condensed nucleosome arrays by hydrogen exchange and NMR. *J. Am. Chem. Soc.* 131:15104–15105.
21. Zhou, B.-R., H. Feng, …, Y. Bai. 2012. Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J. Mol. Biol.* 421:30–37.
22. Gao, M., P. S. Nadaud, …, C. P. Jaroniec. 2013. Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *J. Am. Chem. Soc.* 135:15278–15281.
23. Zhong, L., and W. C. Johnson, Jr. 1992. Environment affects amino acid preference for secondary structure. *Proc. Natl. Acad. Sci. USA.* 89:4462–4465.
24. Yang, Y., A. P. Lyubartsev, …, L. Nordenskiöld. 2009. Computer modeling reveals that modifications of the histone tail charges define salt-dependent interaction of the nucleosome core particles. *Biophys. J.* 96:2082–2094.
25. Arya, G., and T. Schlick. 2009. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A.* 113:4045–4059.
26. Sun, J., Q. Zhang, and T. Schlick. 2005. Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc. Natl. Acad. Sci. USA.* 102:8180–8185.
27. Arya, G., Q. Zhang, and T. Schlick. 2006. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys. J.* 91:133–150.
28. 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.
29. Sharma, S., F. Ding, and N. V. Dokholyan. 2007. Multiscale modeling of nucleosome dynamics. *Biophys. J.* 92:1457–1470.
30. Voltz, K., J. Trylska, …, J. Langowski. 2012. Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study. *Biophys. J.* 102:849–858.
31. Biswas, M., J. Langowski, and T. C. Bishop. 2013. Atomistic simulations of nucleosomes. *Wiley Interdiscip. Comput. Mol. Sci.* 3:378–392.
32. Ruscio, J. Z., and A. Onufriev. 2006. A computational study of nucleosomal DNA flexibility. *Biophys. J.* 91:4121–4132.
33. Roccatano, D., A. Barthel, and M. Zacharias. 2007. Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers.* 85:407–421.
34. Biswas, M., K. Voltz, …, J. Langowski. 2011. Role of histone tails in structural stability of the nucleosome. *PLOS Comput. Biol.* 7:e1002279.
35. Ettig, R., N. Kepper, …, K. Rippe. 2011. Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. *Biophys. J.* 101:1999–2008.
36. Cino, E. A., W. Y. Choy, and M. Karttunen. 2012. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.* 8:2725–2740.
37. Lindorff-Larsen, K., P. Maragakis, …, D. E. Shaw. 2012. Systematic validation of protein force fields against experimental data. *PLoS ONE.* 7:e32131.
38. Korolev, N., H. Yu, …, L. Nordenskiöld. 2014. Molecular dynamics simulations demonstrate the regulation of DNA-DNA attraction by H4 histone tail acetylations and mutations. *Biopolymers.* 101:1051–1064.
39. 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.

40. Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* 65:1604–1608.

41. Okabe, T., M. Kawata, …, M. Mikami. 2001. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chem. Phys. Lett.* 335:435–439.

42. Okamoto, Y. 2004. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graph. Model.* 22:425–439, (Conformational Sampling.).

43. 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.

44. Bondi, A. 1964. van der Waals volumes and radii. *J. Phys. Chem.* 68:441–451.

45. Tsui, V., and D. A. Case. 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122:2489–2498.

46. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.

47. Van Der Spoel, D., E. Lindahl, …, H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

48. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 306–317.

49. Berendsen, H., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

50. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

51. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

52. Anandakrishnan, R., M. Daga, and A. V. Onufriev. 2011. An n log n generalized Born approximation. *J. Chem. Theory Comput.* 7:544–559.

53. MacKerell, Jr., A. D., N. Banavali, and N. Foloppe. 2000-2001. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers.* 56:257–265.

54. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

55. Best, R. B., N. V. Buchete, and G. Hummer. 2008. Are current molecular dynamics force fields too helical? *Biophys. J.* 95:L07–L09.

56. Tzanov, A. T., M. A. Cuendet, and M. E. Tuckerman. 2014. How accurately do current force fields predict experimental peptide conformations? An adiabatic free energy dynamics study. *J. Phys. Chem. B.* 118:6539–6552.

57. Huang, J., and A. D. MacKerell, Jr. 2014. Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)3 peptide. *Biophys. J.* 107:991–997.

58. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 65:712–725.

59. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 78:1950–1958.

60. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

61. Gromiha, M. M., C. Santhosh, and M. Suwa. 2004. Influence of cation-π interactions in protein-DNA complexes. *Polymer.* 45:633–639.

62. Stewart, M., T. Dunlap, …, L. McFail-Isom. 2013. Cations form sequence selective motifs within DNA grooves via a combination of cation-pi and ion-dipole/hydrogen bond interactions. *PLoS ONE.* 8:e71420.

63. Hirano, Y., K. Hizume, …, Y. Hiraoka. 2012. Lamin B receptor recognizes specific modifications of histone H4 in heterochromatin formation. *J. Biol. Chem.* 287:42654–42663.

64. Mangenot, S., A. Leforestier, …, F. Livolant. 2002. Salt-induced conformation and interaction changes of nucleosome core particles. *Biophys. J.* 82:345–356.

65. Mutskov, V., D. Gerber, …, S. Dimitrov. 1998. Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Mol. Cell. Biol.* 18:6293–6304.

66. Kan, P.-Y., and J. J. Hayes. 2007. Detection of interactions between nucleosome arrays mediated by specific core histone tail domains. *Methods.* 41:278–285, (Methods Related to the Structure and Function of Eukaryotic Chromatin.).

Supporting Material: The role of histone tails in the nucleosome: A computational study

Jochen Erler[1], Ruihan Zhang[1], Loukas Petridis[2], Xiaolin Cheng[2], Jeremy C. Smith[2], Joerg Langowski[1]

1 Division Biophysics of Macromolecules, German Cancer Research Center, INF 580,

D-69120 Heidelberg, Germany

2 UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge,

Tennessee 37831, USA

* E-mail: Corresponding joerg.langowski@dkfz-heidelberg.de

Methods

REMD configurations

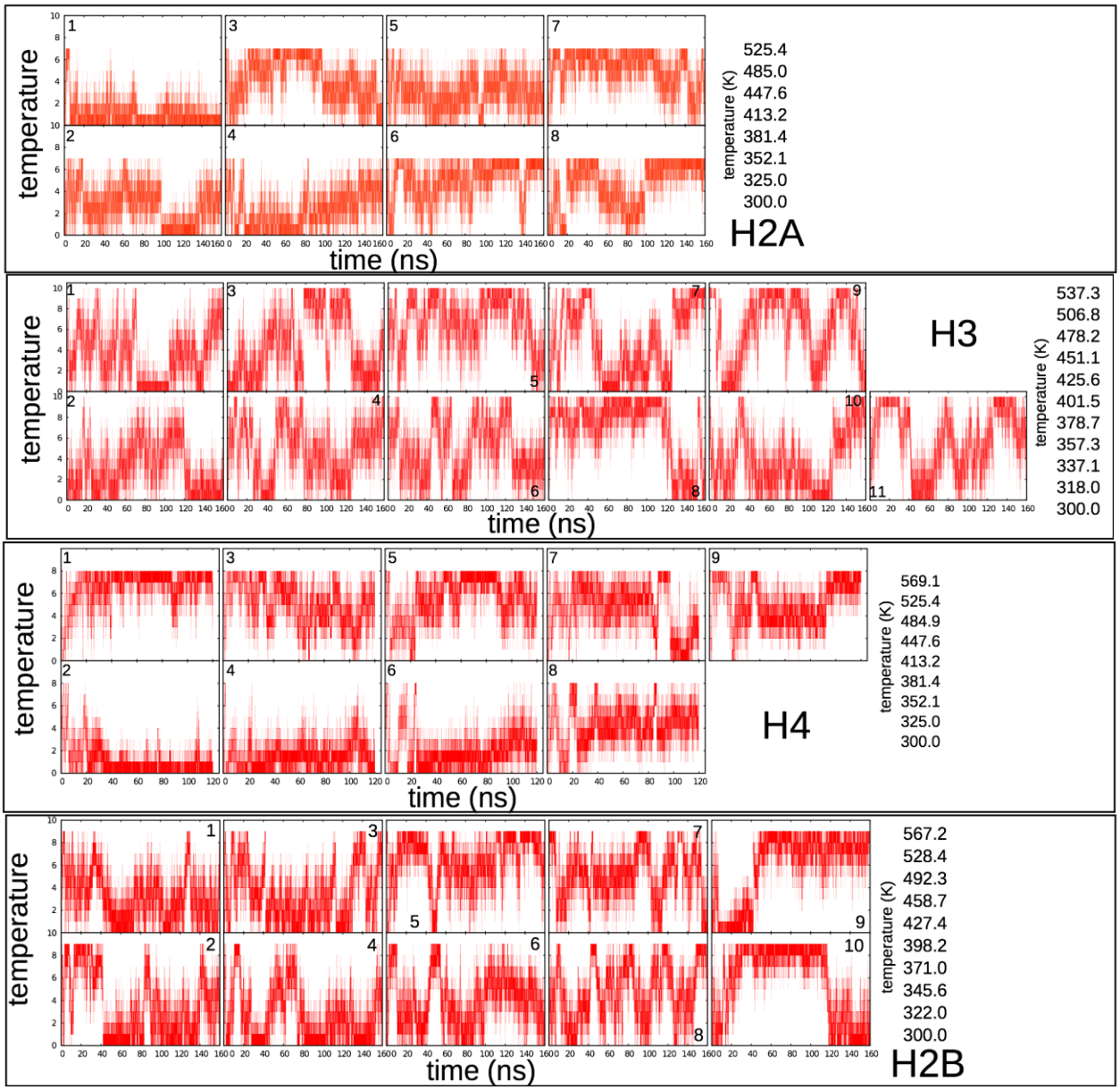

Figure S1. **Temperature space of the REMD simulations**. Results for trajectories of histone tails H3 (160 ns), H2B (160 ns), H4 (120 ns) and H2A (160 ns). Panels show the different replicas starting from the chosen temperature levels.

Test 1: Robustness of results against REMD conditions

To investigate a possible dependence of the results on the exchange rate (number of replicas) we performed a further simulation of the histone tail H2A on the nucleosome using 10 instead of 8 replicas in the same temperature range. The result is an exchange probability of 0.22. At the same time we reduced the frequency for exchange attempts from an attempt every 100 steps (0.2 ps) to an attempt every 400 steps (0.8 ps). Figure S2 shows the temperature space and exchange probabilities for both runs.

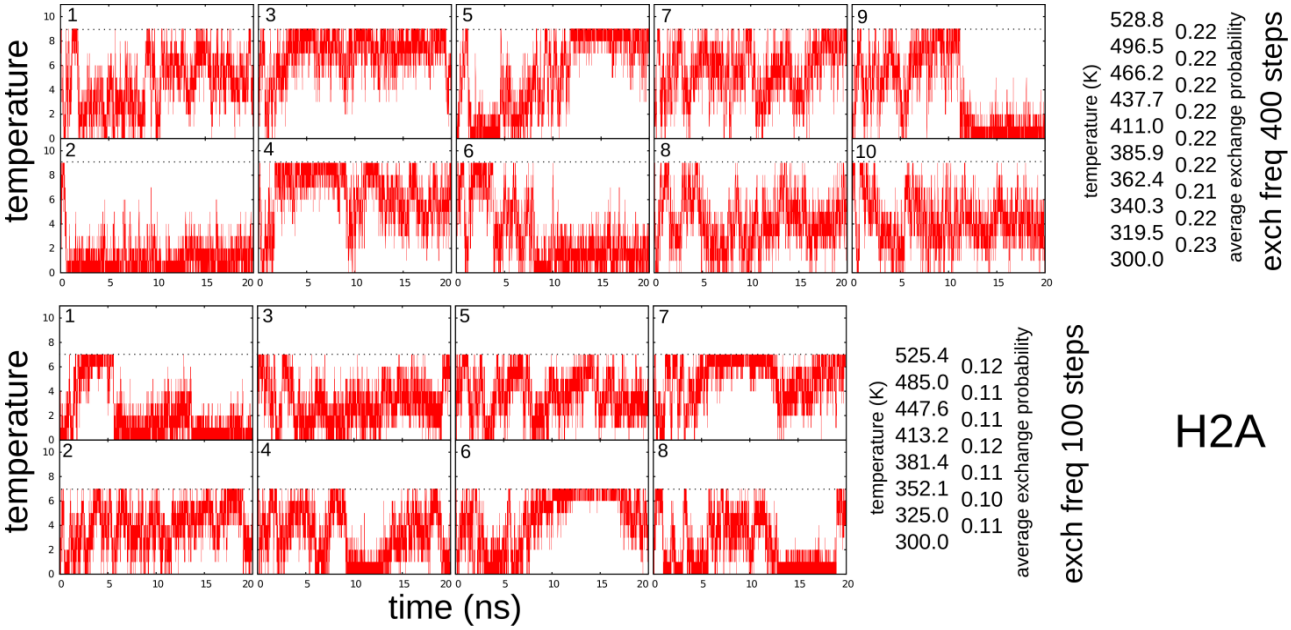

Figure S2**. Temperature space of REMD simulations of H2A tail**. Results of a REMD simulations with 10 replicas, a resulting exchange probability of about 0.22 and exchange attempts every 0.8 ps (400 steps) and the simulation of H2A with 8 replicas, an exchange probability of 0.1 and exchange attempts every 0.2 ps (100 steps). Panels show the different replicas starting from the chosen temperature levels.

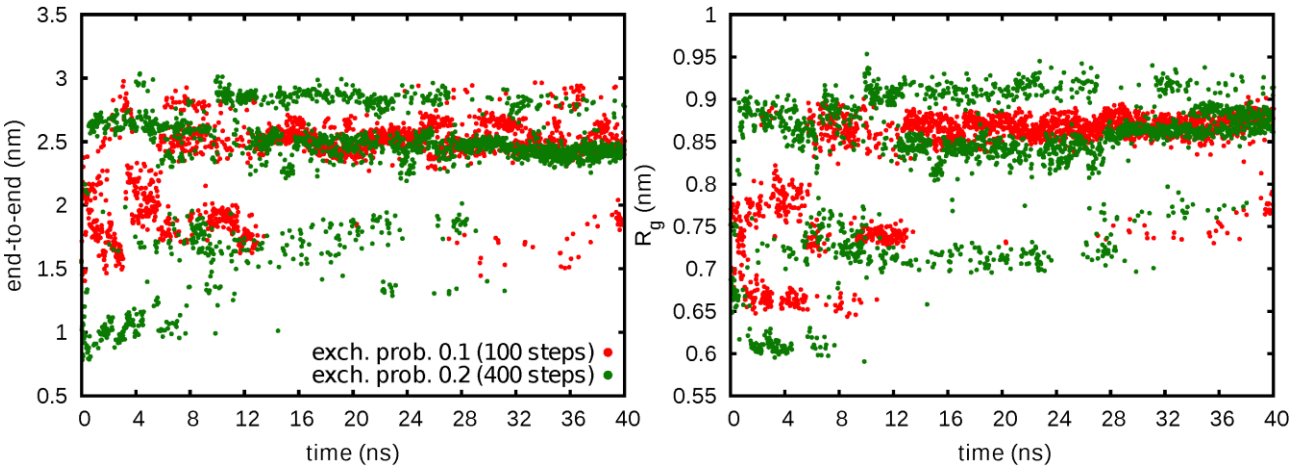

Figure S3. **End-to-end distance and radius of gyration**. Comparison of results of the two trajectories shown in Fig. S2.

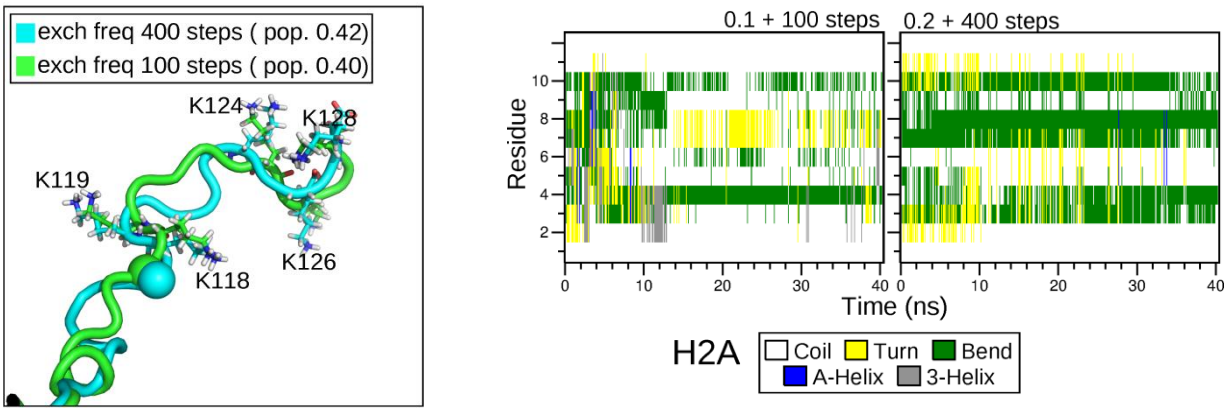

Figure S4**. Conformation and secondary structure of the H2A tail**. Comparison of results of the two trajectories shown in Fig. S2.

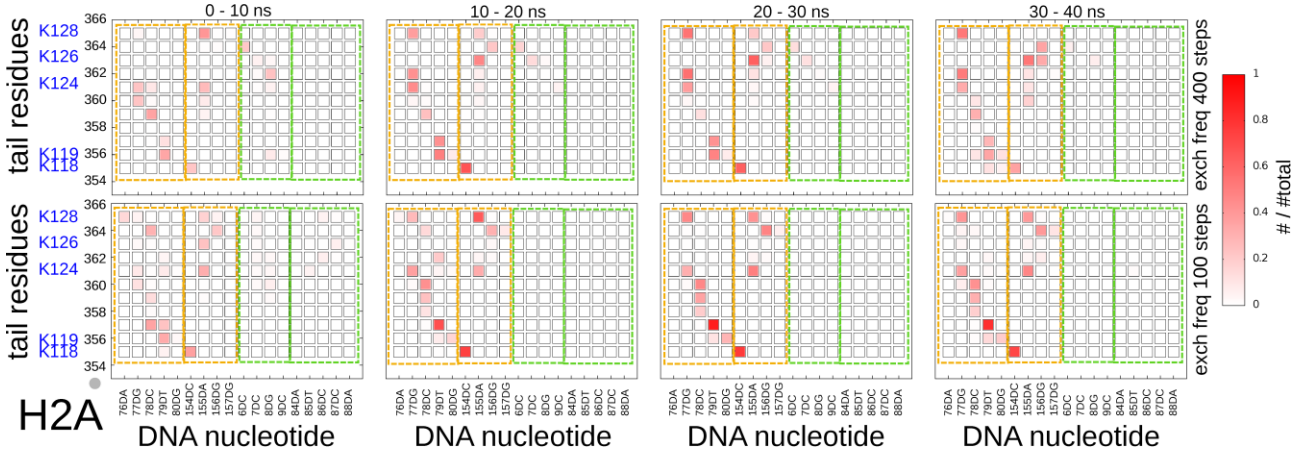

Figure S5. **Probability of contacts between histone tail residues and DNA phosphate for H2A tail.** Comparison of results of the two trajectories shown in Fig. S2.

The exchange is in both runs sufficient. For a comparison of the dynamics of both REMD runs we study the end-to-end distance and the radius of gyration (Fig. S3), results of a cluster analysis, the secondary structure (Fig. S4) and DNA:tail contacts (Fig. S5).

Both simulations converge to the same end-to-end distance and radius of gyration, whereas the 0.1 (100 steps) REMD seemed to converge faster. Results for secondary structure are also similar. In both cases we find bends around residue 4 and 10 and coil around the same residues.

Figure S4 shows also the highest populated clusters of both REMD runs. The probabilities are 0.42 and 0.40 for the 400 steps and 100 steps exchange frequency. Charged lysine residues, except the K126, make contacts with the same DNA phosphat groups for both runs.

For a comparison of DNA:tail contacts we devided the 40 ns trajectory into four parts and calculated the population of each contact (Fig. S5). Both results show the same DNA:tail contacts, but the run with 10 replicas and attempts every 400 steps seems to converge more slowly. We expect the results to converge to the same DNA:tail contacts for the whole tail when we would extend the simulation.

.

Test 2: Test of an longer cutoff in the GB simulation

A too short cutoff can have dramatic influences on the dynamics of the histone tails (see Ref. (1)). Compared to the example shown in (1) our simulations contains 10 bp of linker DNA. For this reason we do not expect such an drastic effect, since the H3 tail is close to the linker DNA.
To investigate a possible dependence on the chosen cutoff we performed further REMD simulations of H3 and H2A using a cutoff of 2.5 nm instead of 1.5 nm. This results in a longer simulation time (0.6 ns/day instead of 3 ns/day). Fig. S6 shows results for the number of DNA:tail contacts for H2A and H3. The number of contacts for the REMD simulation with a 2.5 nm cutoff agree well for both tails with results using a 1.5 nm and we do not expect any influence on the dynamics of the histone tails in the REMD simulations.

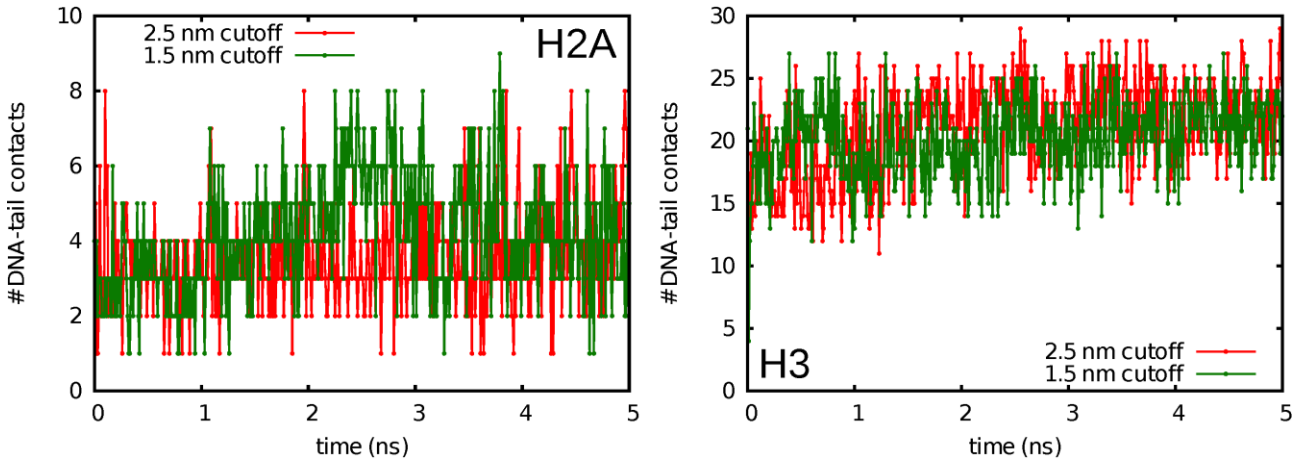

Figure S6. **Number of contacts between DNA and histone tail H2A and H3.** Comparison of results using a cutoff of 1.5 and 2.5 nm in the GB simulation.

Test 3: Dependence of results on the definition of the freeze group

To justify the approximation of freezing the whole nucleosome except the tail under investigation, we performed a further simulation of the H4 tail keeping 16 residues less fixed. The new system contains 42 free residues instead of 26 as in the tail definition in Fig. 1.
We have chosen H4 for the test, since we extended already H2A in the present REMD simulation. For H3 and H2B we do not expect any influence since both N-terminal tails protrude between two DNA strands and are thus fixed by binding to the phosphate groups (compare Fig. S31 and S32).
Fig. S7 shows results for end-to-end distance and the radius of gyration. Both runs converge to the same values, whereas the convergence for the shorter tail is faster.
The conformation of the cluster with the highest population and the secondary structure agree well for both system from residue 1 till 19 (see Fig. S8). The reason for this can be seen in Fig. S9 showing the probability of the DNA:tail contacts. Residues R3, K5, K8, K12, K16 and R17 make the same contacts for both systems (after 30 ns).

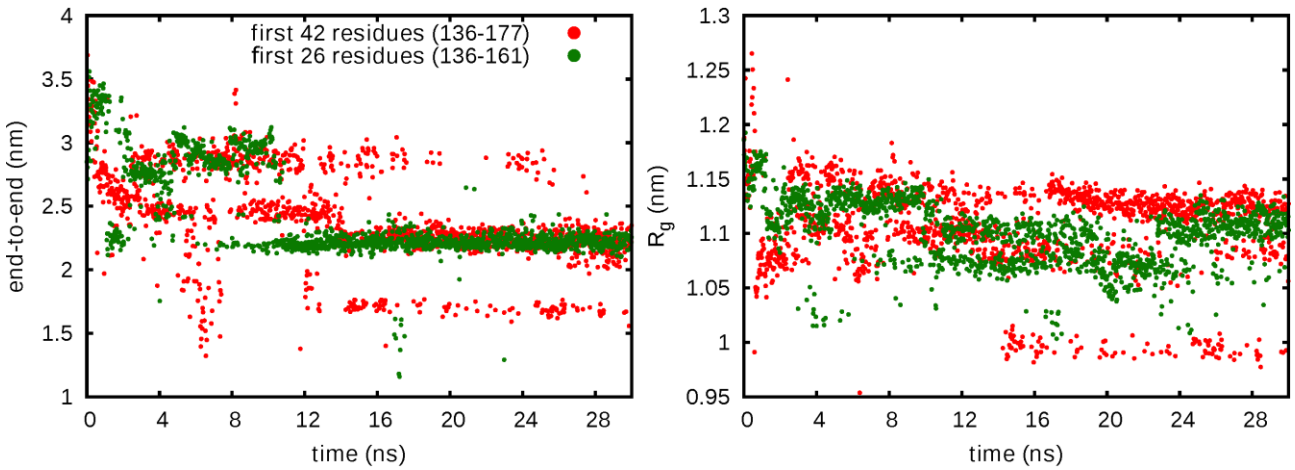

Figure S7. **End-to-end distance and radius of gyration**. Comparison of results of the two REMD simulation with a different number of free residues of the histone H4.

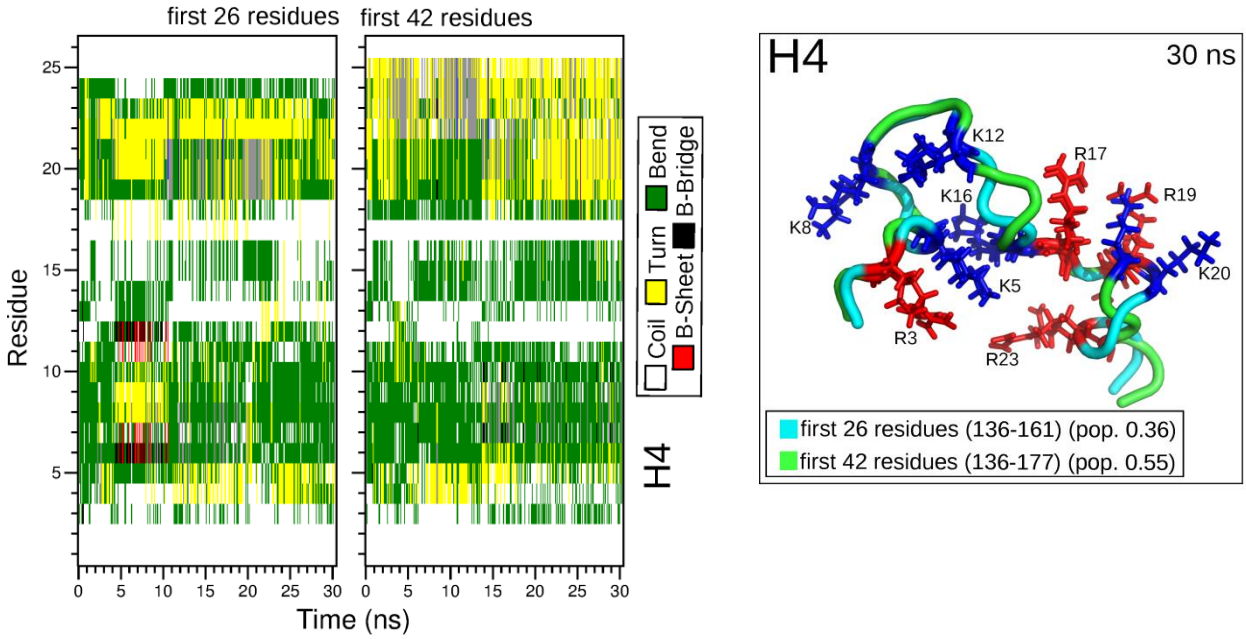

Figure S8. **Secondary structure and conformation of the H4 tail**. Comparison of results of the two REMD simulation with a different number of free residues of the histone H4.

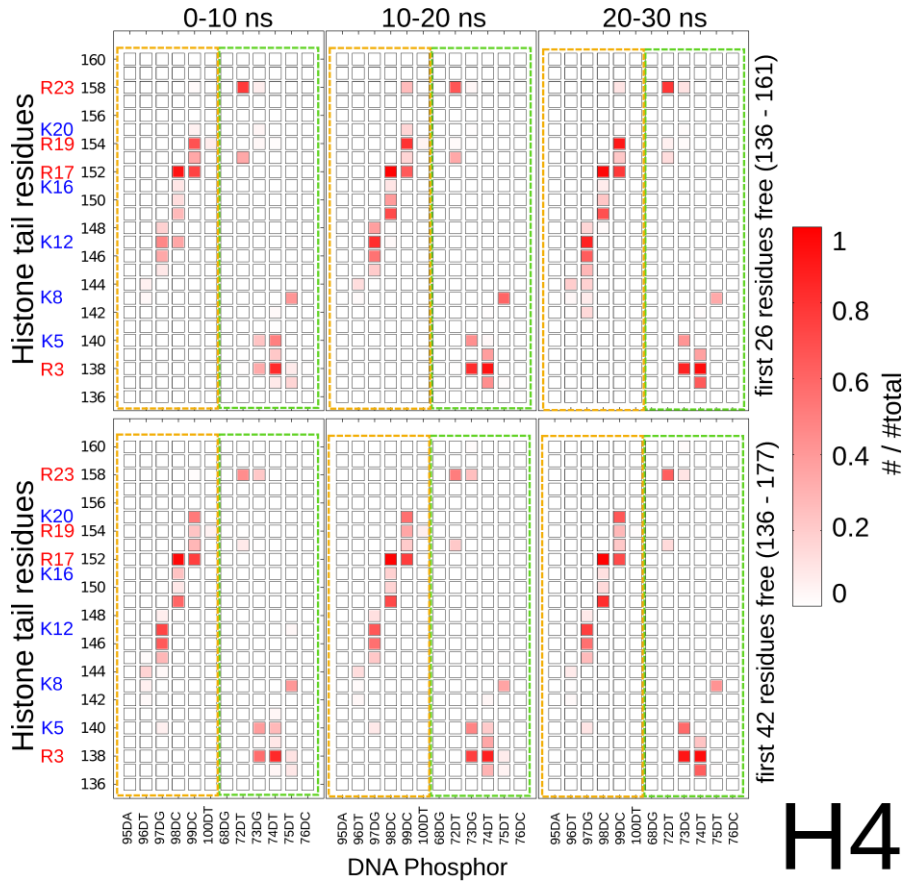

Figure S9. **Probability of contacts between histone tail residues and DNA phosphate for H4 tail.** Comparison of results of the two REMD simulation with a different number of free residues of the histone H4.

Convergence

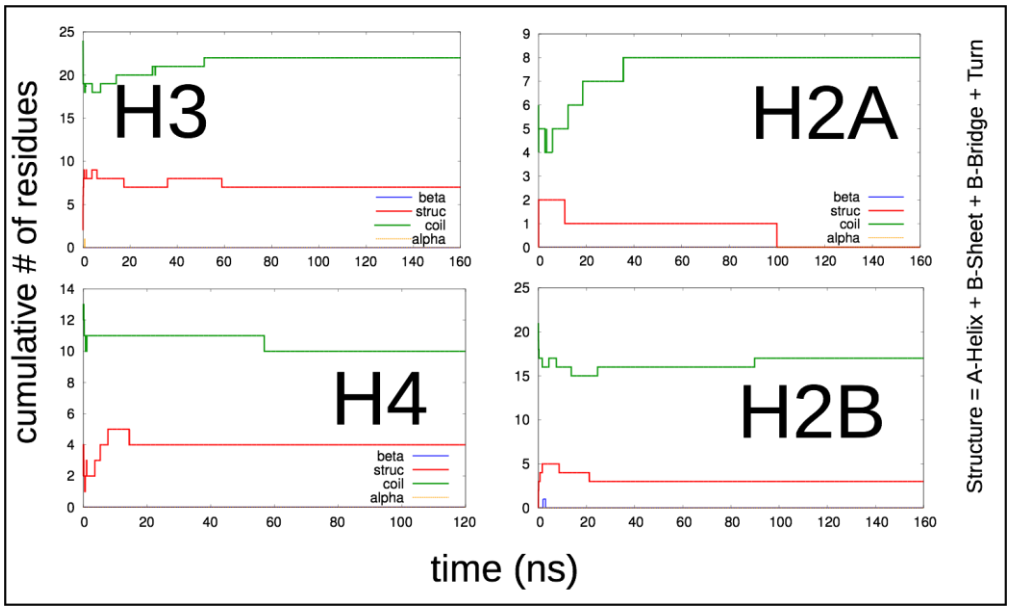

Figure S10. **Accumulated number of residues in a structure for all histone tails.**

Results of implicit REMD simulations

Isolated histone tails

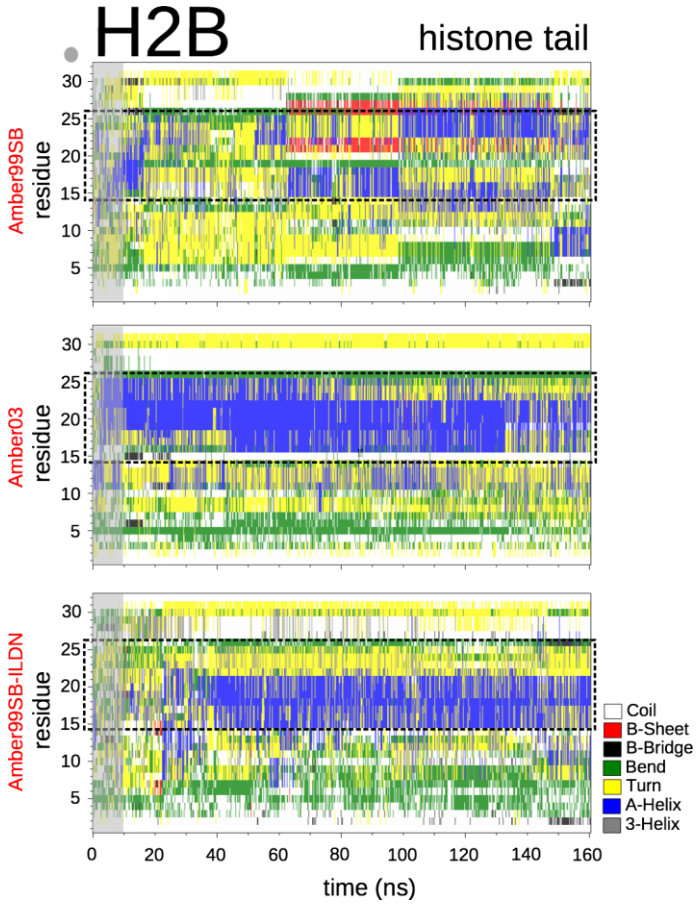

Figure S11. **Secondary structure of the isolated H2B histone tail**. Shown are results for the lowest temperature level 300 K of the REMD calculation. Used were the three force fields Amber99SB, Amber03, Amber99SB-ILDN . Black boxes are added to make a comparison between different force fields easier

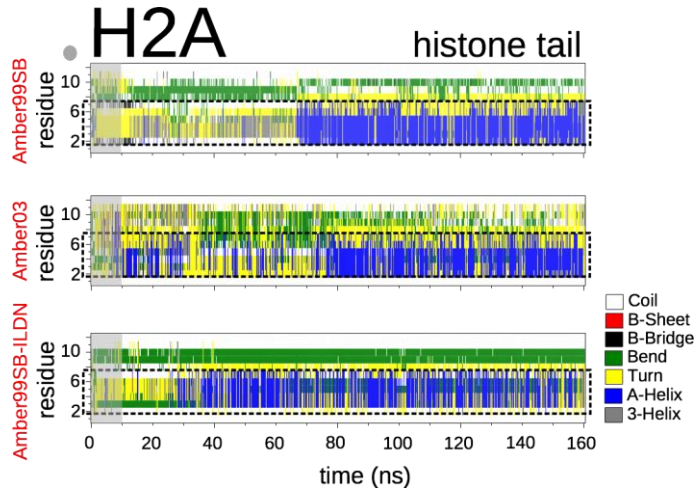

Figure S12. **Secondary structure of the isolated H2A histone tail**. Shown are results for the lowest temperature level 300 K of the REMD calculation. Used were the three force fields Amber99SB, Amber03, Amber99SB-ILDN . Black boxes are added to make a comparison between different force fields easier

Results for the nucleosome

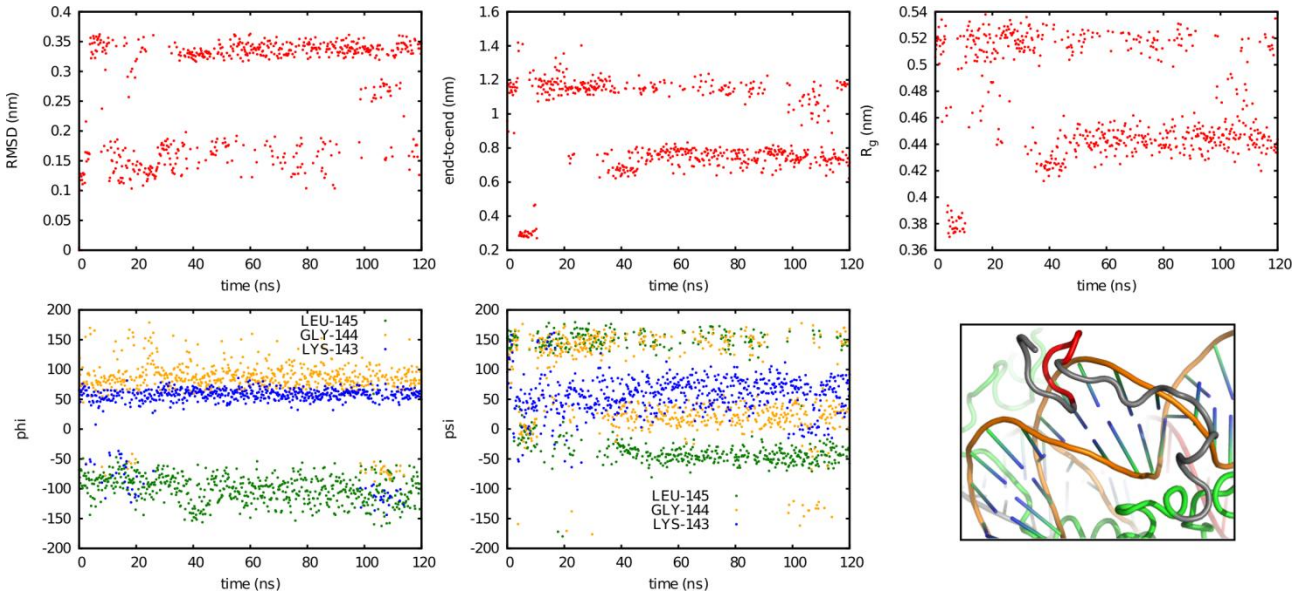

Figure S13. **RMSD, end-to-end distance, radius of gyration, and the dihedral angles *φ* and ψ for a part of the H4 tail** (residue 7 to 11; GKGLG, red in right lower panel).

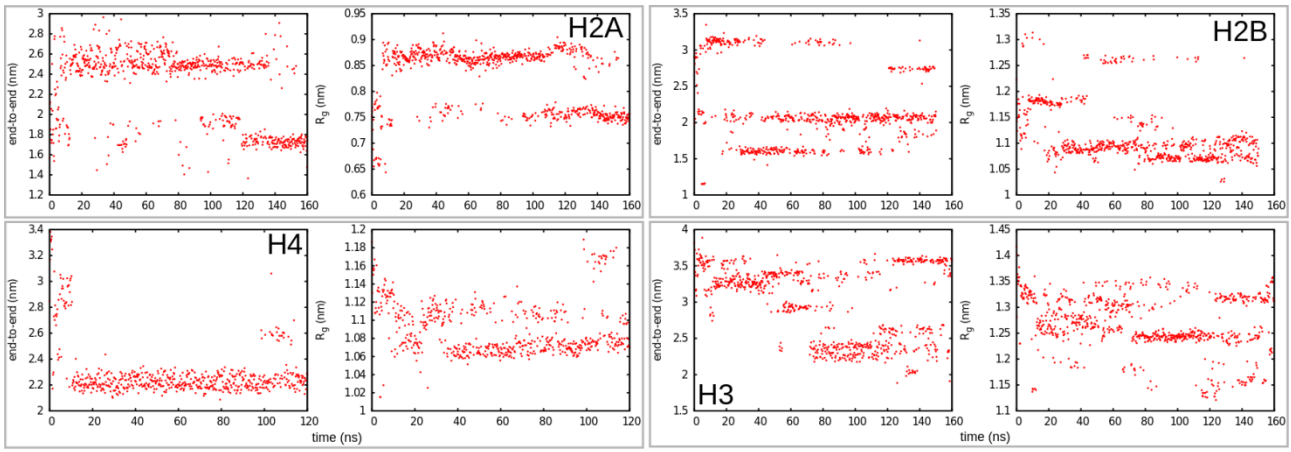

Figure S14**. End-to-end distance and radius of gyration for H4, H3, H2B and H2A tail**.

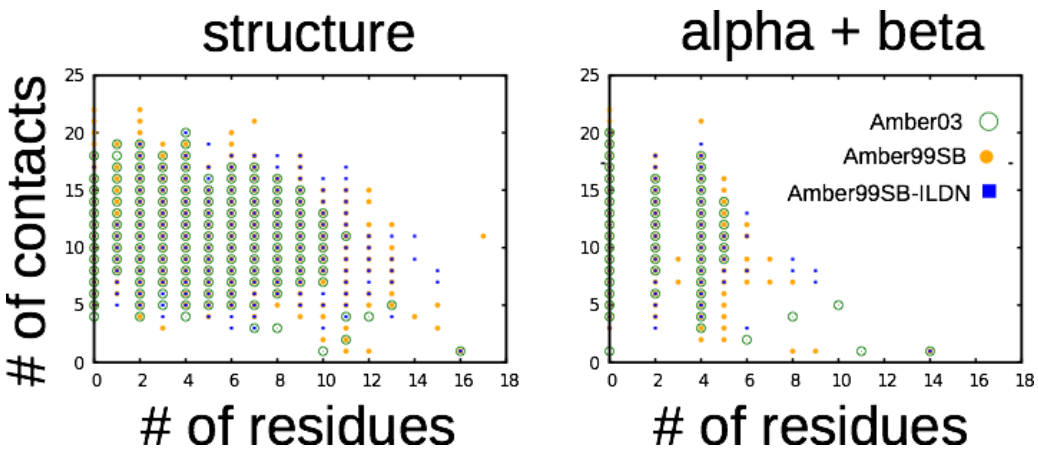

Figure S15. **Number of contacts versus number of residues in a secondary structure**. Results for force fields Amber99SB, Amber03 and Amber99SB-ILDN are shown.

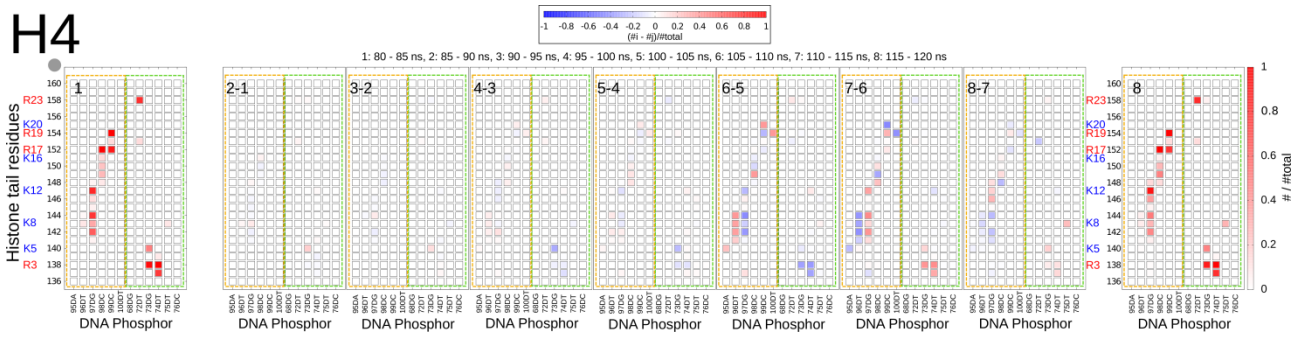

Figure S16. **Probability of contacts between histone tail residues and DNA phosphate for H4 tail**. The last part of the trajectory (80 to 120 ns) is divided in eight smaller intervals to see how the contacts evolve. The panels left and right show the total probabilities, the panels in-between the increase or decrease.

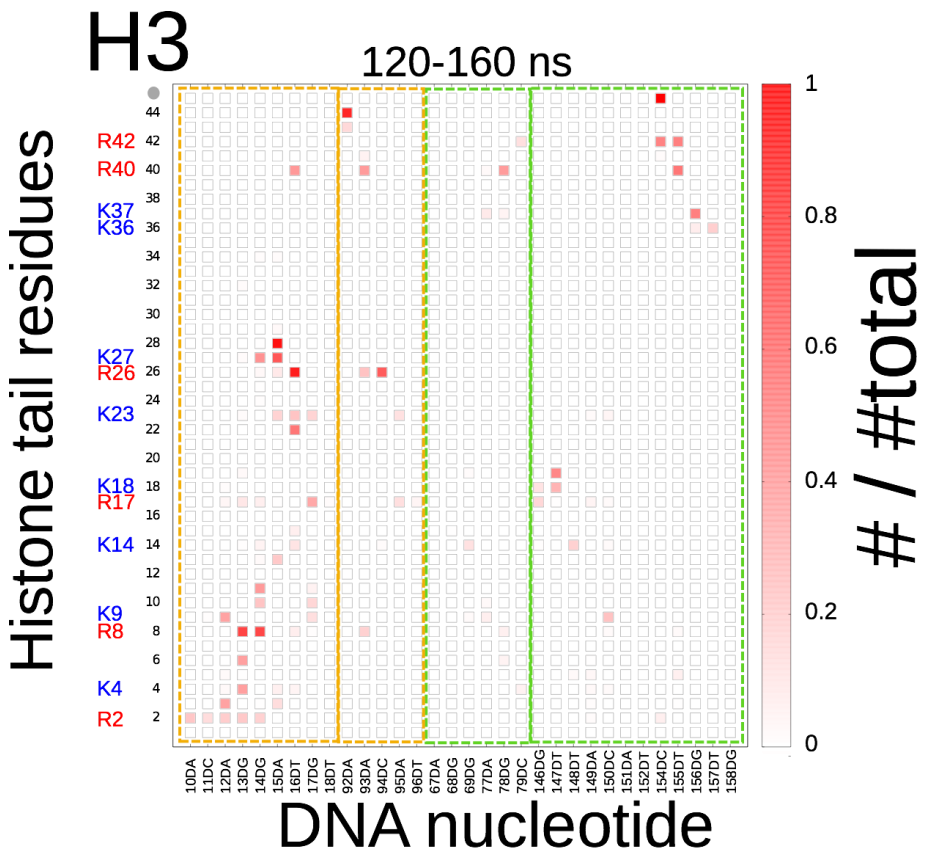

Figure S17. **Probability of contacts between histone tail residues and DNA phosphate for H3 tail**. The last part of the trajectory (120 to 160 ns) was studied and the threshold for a contact was 0.3 nm. Coloured boxes indicate different DNA strands.

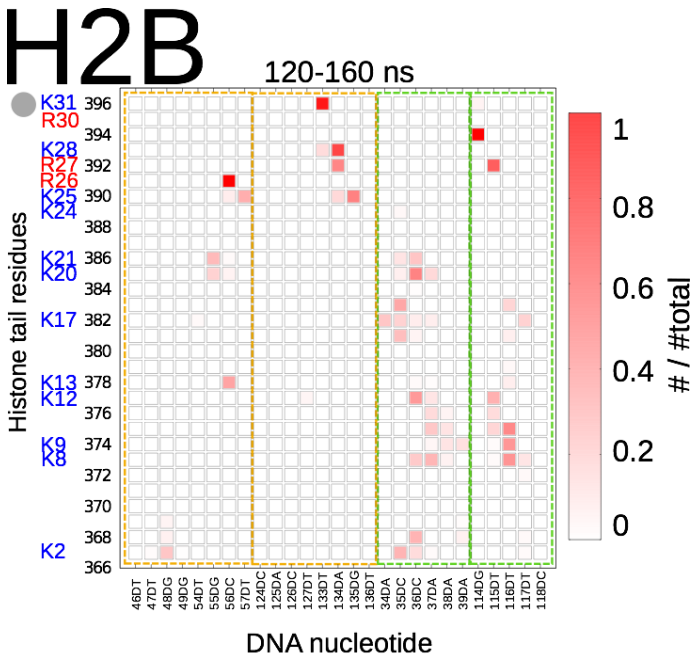

Figure S18. **Probability of contacts between histone tail residues and DNA phosphate for H2B tail**. The last part of the trajectory (120 to 160 ns) was studied and the threshold for a contact was 0.3 nm. Colored boxes indicate different DNA strands.

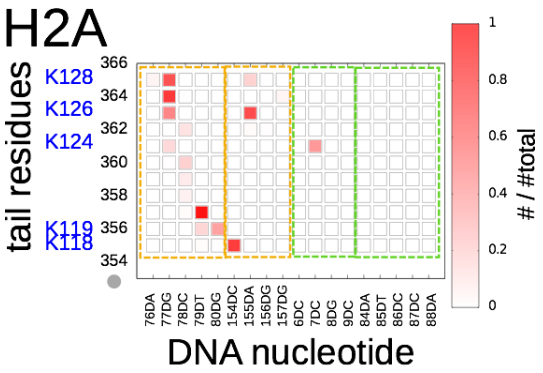

Figure S19. **Probability of contacts between histone tail residues and DNA phosphate for H2A tail**. The last part of the trajectory (120 to 160 ns) was studied and the threshold for a contact was 0.3 nm. Colored boxes indicate different DNA strands.

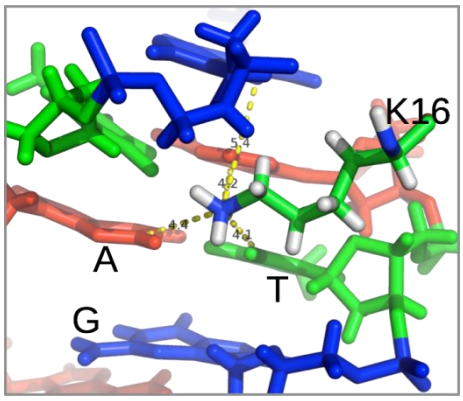

Figure S20. **Cations-π interaction of K16 of H4 tail**. Residue K16 and the surrounding base pairs adenine (red), guanine (blue) and thymine (green). Distances were calculated between the cationic group of K16 and the centroid of the base ring. Angles are defined as the angle formed between the cationic group, the base centroid, and a point positioned normal to the plane. Distances and angles are 0.42 nm, 54° or 0.44 nm, 48° between adenine and K16, 0.54 nm, 22° between guanine and K16 and 0.41 nm, 72° between thymine and K16.

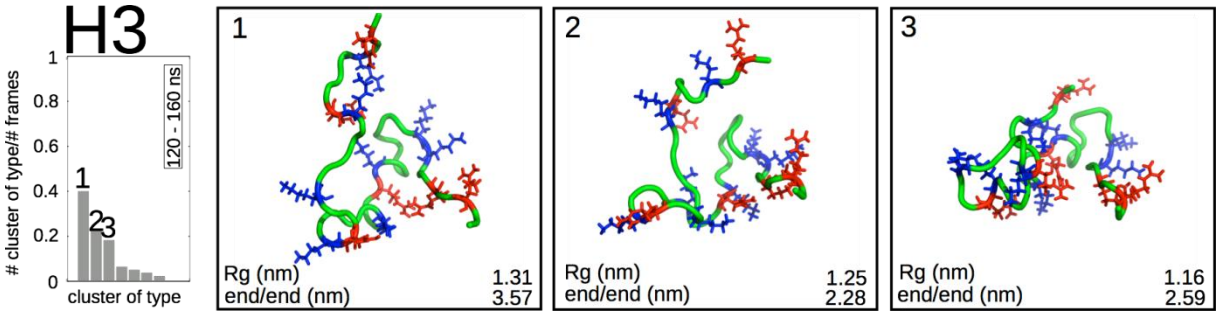

Figure S21. **Results of a cluster analysis of H3 tail**. The cutoff was 0.2 nm and the last part of the trajectory (120 to 160 ns) was used. The probability and the conformations of the cluster representative of the three most populated clusters are shown.

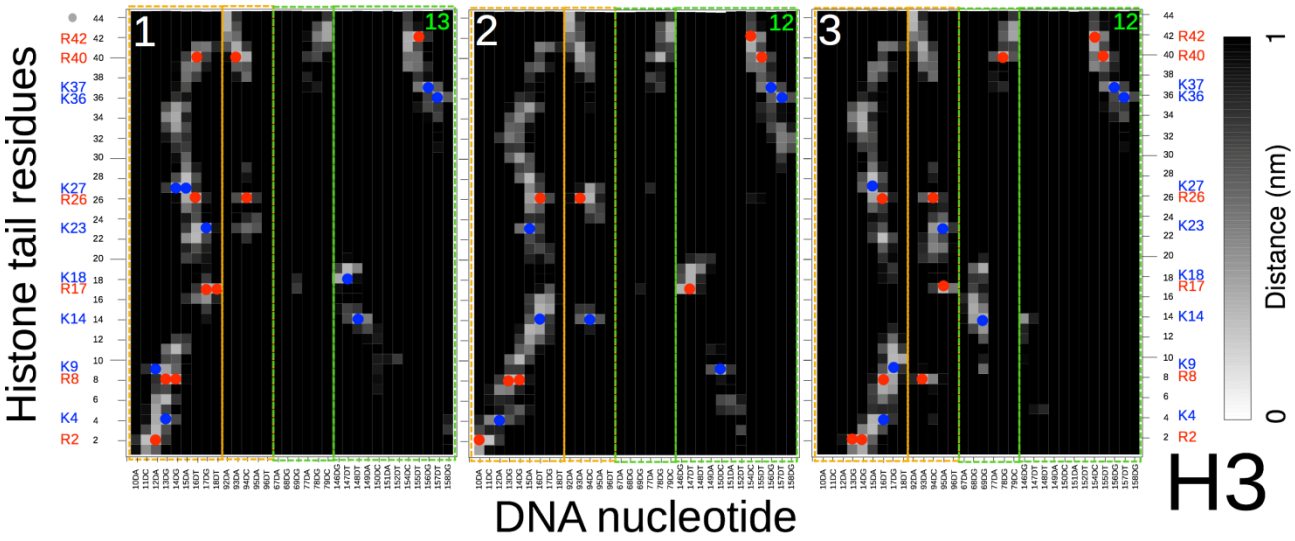

Figure S22. **Distance maps for the three most populated clusters of H3 tail**. Distances were measured between histone tail residues and the phosphate group of each nucleotide. Blue circles and red circles indicate contacts of lysines and arginines with DNA phosphate groups (cutoff 0.3 nm). Green numbers indicate the number of neutralized charges.

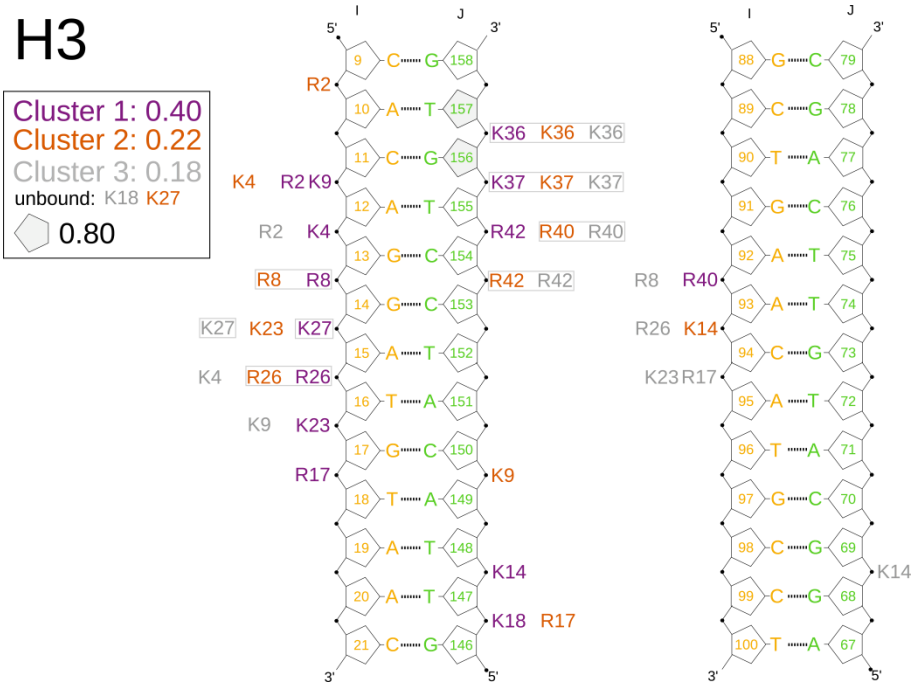

Figure S23. **Tail:DNA interactions of the H3 tail**. Interactions between histone tail and DNA for the clusters found in Fig. S8. Grey boxes indicate interactions found in all three clusters.

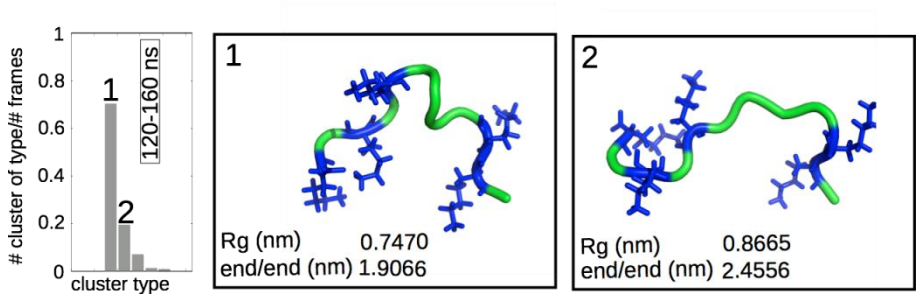

Figure S24. **Results of a cluster analysis of H2A tail**. The cutoff was 0.2 nm and the last part of the trajectory (120 to 160 ns) was used. The probability and the conformations of the cluster representative of the three most populated clusters are shown.

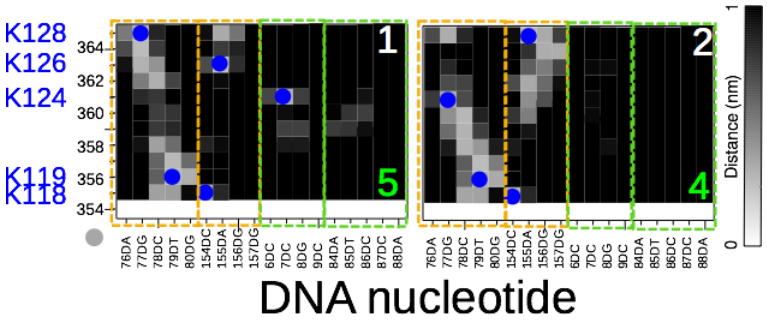

Figure S25. **Distance maps for the three most populated clusters of H2A tail**. Distances were measured between histone tail residues and the phosphate group of each nucleotide. Blue circles and red circles indicate contacts of lysines and arginines with DNA phosphate groups (cutoff 0.3 nm). Green numbers indicate the number of neutralized charges.

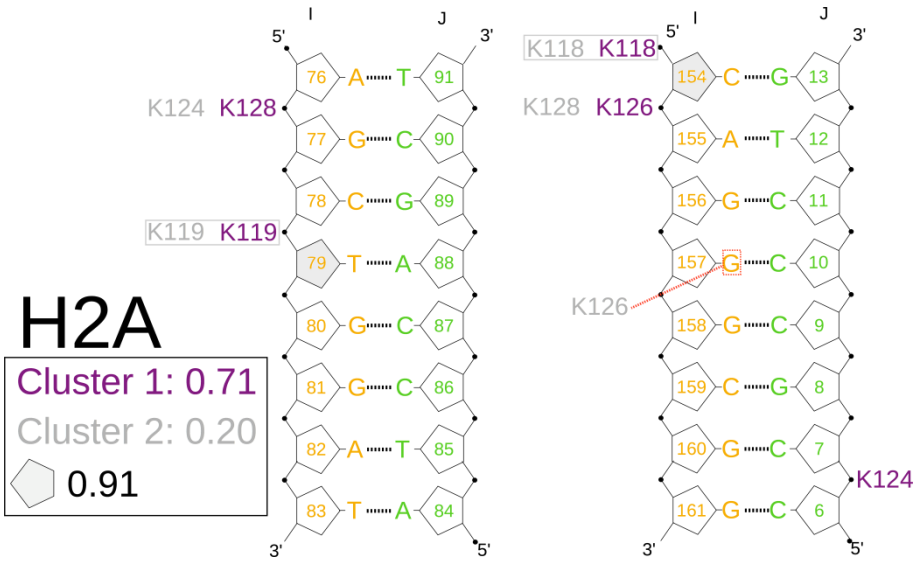

Figure S26. **Tail:DNA interactions of the H2A tail**. Interactions between histone tail and DNA for the clusters found in Fig. S11. Grey boxes indicate interactions found in both clusters. The red dashed line illustrate cation-π interactions.

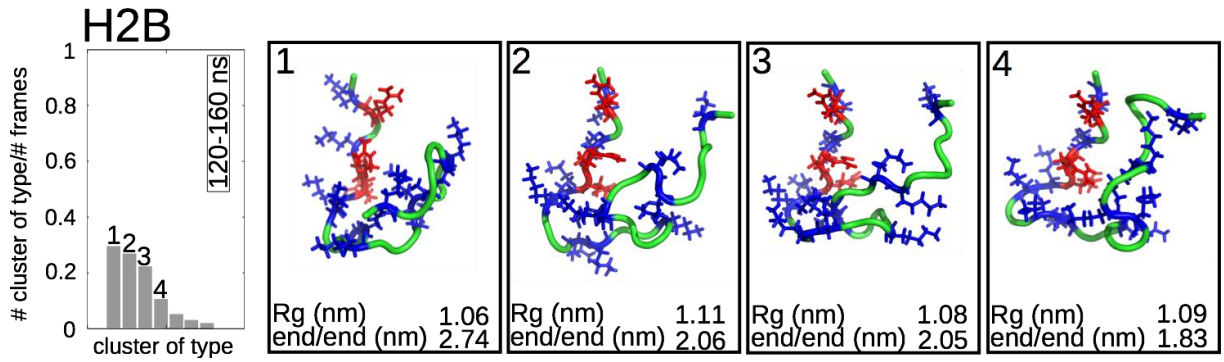

Figure S27**. Results of a cluster analysis of H2B tail**. The cutoff was 0.2 nm and the last part of the trajectory (120 to 160 ns) was used. The probability and the conformations of the cluster representative of the three most populated clusters are shown.

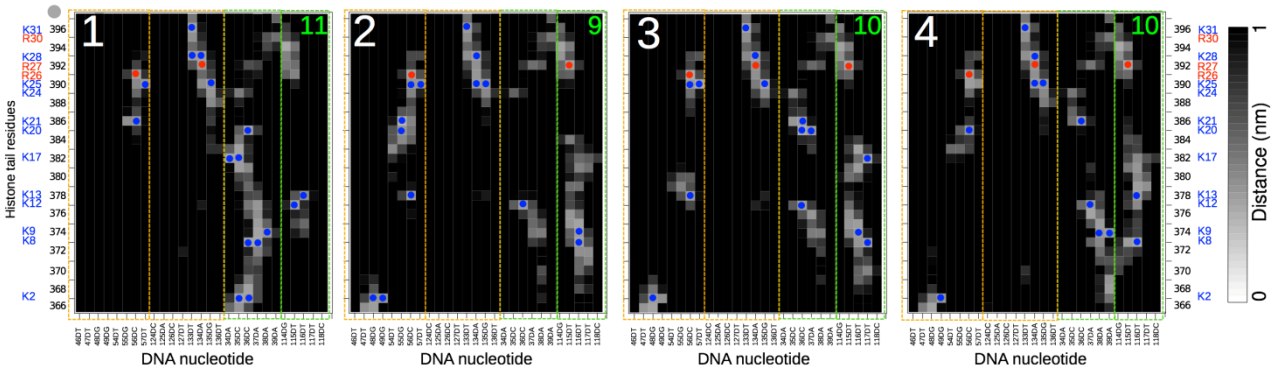

Figure S28. **Distance maps for the three most populated clusters of H2B tail**. Distances were measured between histone tail residues and the phosphate group of each nucleotide. Blue circles and red circles indicate contacts of lysines and arginines with DNA phosphate groups (cutoff 0.3 nm). Green numbers indicate the number of neutralized charges.

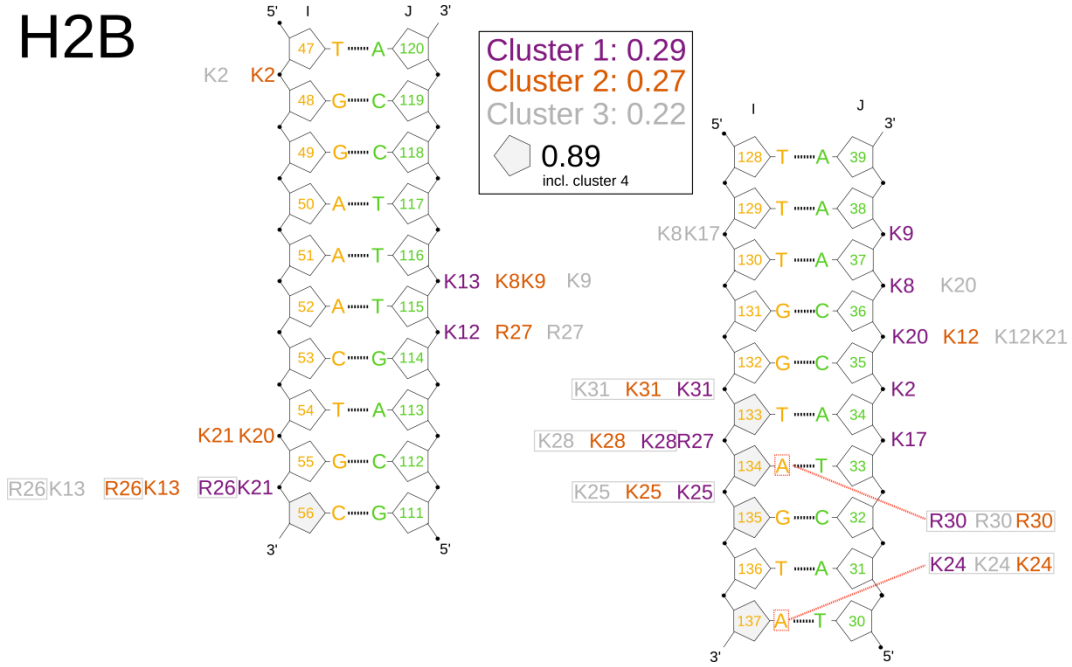

Figure S29. **Tail:DNA interactions of the H2B tail**. Interactions between histone tail and DNA for the clusters found in Fig. S14. Grey boxes indicate interactions found in all four clusters. The red dashed line illustrate cation-π interactions.

Explicit solvent MD results

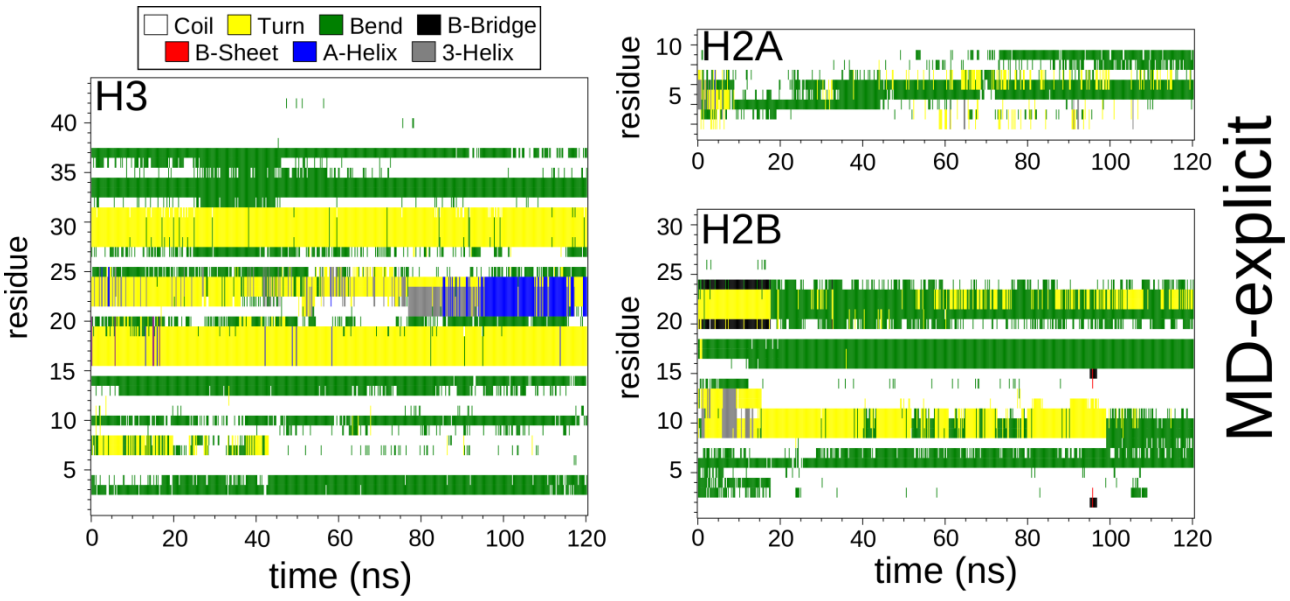

Figure S30. **Secondary structure of the H3, H2A and H2B histone tail of the MD simulation with explicit solvent**. Force field AMBER99SB was used. Different types of structure as indicated. The simulation was initialized with the most probable configuration identified by results of the REMD simulations.

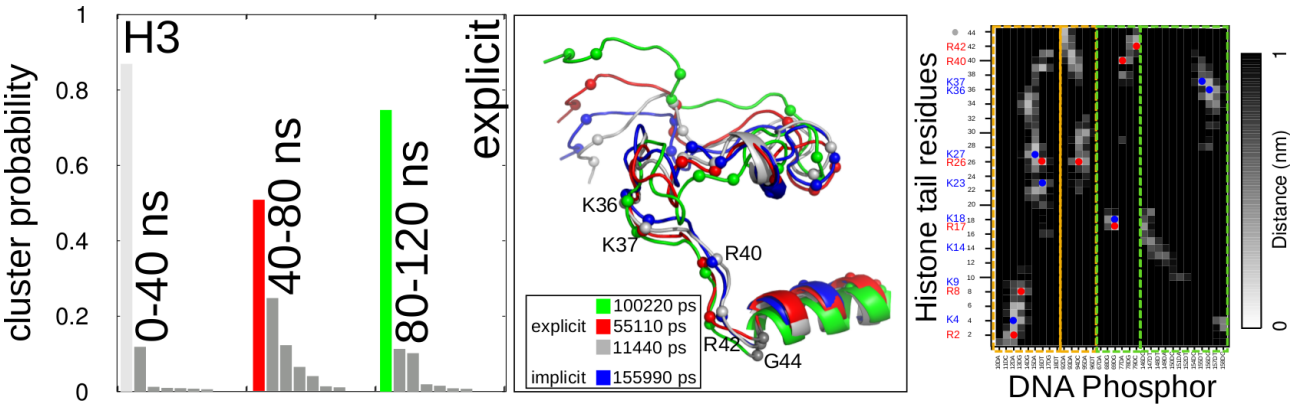

Figure S31. **Results of a cluster analysis of the MD simulations with an explicit solvent**. Shown are results for the H3 tail. The cutoff was 0.2 nm and the trajectory was divided into 3 intervals. Evolution of cluster populations, conformation of the dominant cluster of each interval (including starting configuration from the implicit REMD) and distance map of the dominant configuration in the last interval (80-120ns). Lysines, arginines and the end of the tail (grey) are indicated as beads.

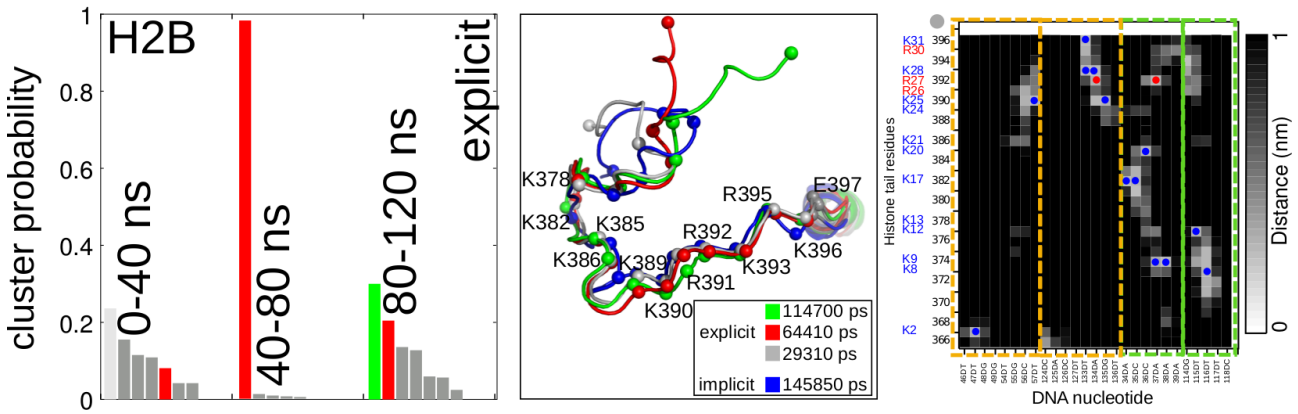

Figure S32. **Results of a cluster analysis of the MD simulations with an explicit solvent**. Shown are results for the H2B tail. The cutoff was 0.2 nm and the trajectory was divided into 3 intervals. Evolution of cluster populations, conformation of the dominant cluster of each interval (including starting configuration from the implicit REMD) and distance map of the dominant configuration in the last interval (80-120ns). Lysines, arginines and the end of the tail (grey) are indicated as beads.

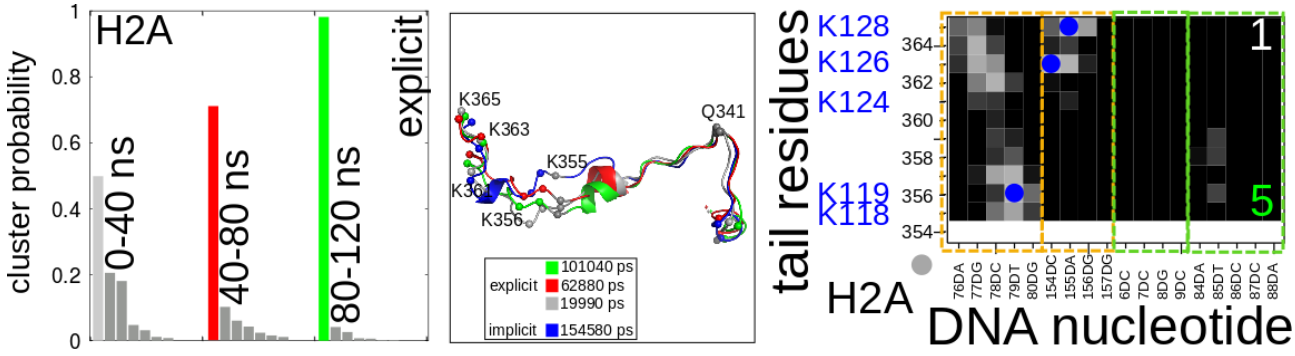

Figure S33. **Results of a cluster analysis of the MD simulations with an explicit solvent**. Shown are results for the H2A tail. The cutoff was 0.2 nm and the trajectory was divided into 3 intervals. Evolution of cluster populations, conformation of the dominant cluster of each interval (including starting configuration from the implicit REMD) and distance map of the dominant configuration in the last interval (80-120ns). Lysines, arginines and the end of the tail (grey) are indicated as beads.

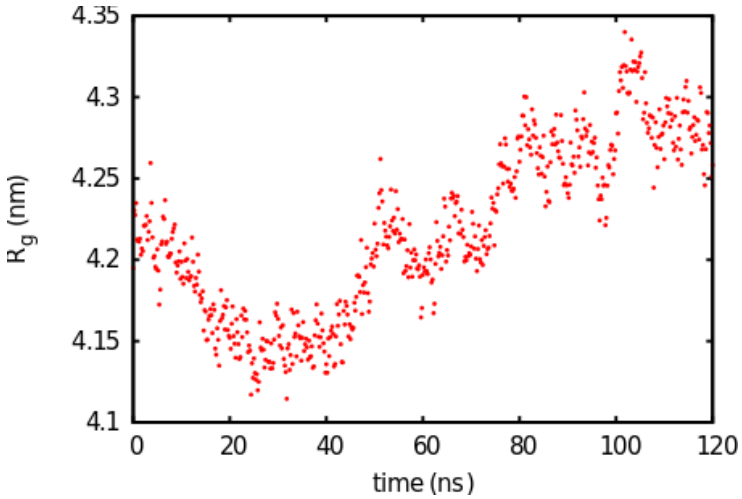

Figure S34. **Radius of gyration of the whole nucleosome**. Shown are results from the explicit solvent MD simulation.

Supporting References

1. Anandakrishnan, R., Daga, M., Onufriev, A.V. (2011). An n log n generalized born approximation. J. Chem. Theory. Comput. 7, 544-559.