Supporting Information: Supplementary figures.

Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale simulation: atomistic insights into nucleosome functions

Alexey K. Shaytan[1,2], Grigoriy A. Armeev[2], Alexander Goncearenco[1], Victor B. Zhurkin[3], David Landsman[1], Anna R. Panchenko[1,*]

[1]National Center for Biotechnology Information, NLM, NIH, Bethesda, Maryland 20894, United States
[2]Faculty of Biology, Lomonosov Moscow State University, Moscow 119991, Russia
[3]Laboratory of Cell Biology, National Cancer Institute, NIH, Bethesda, Maryland 20892, United States

*panch@ncbi.nlm.nih.gov

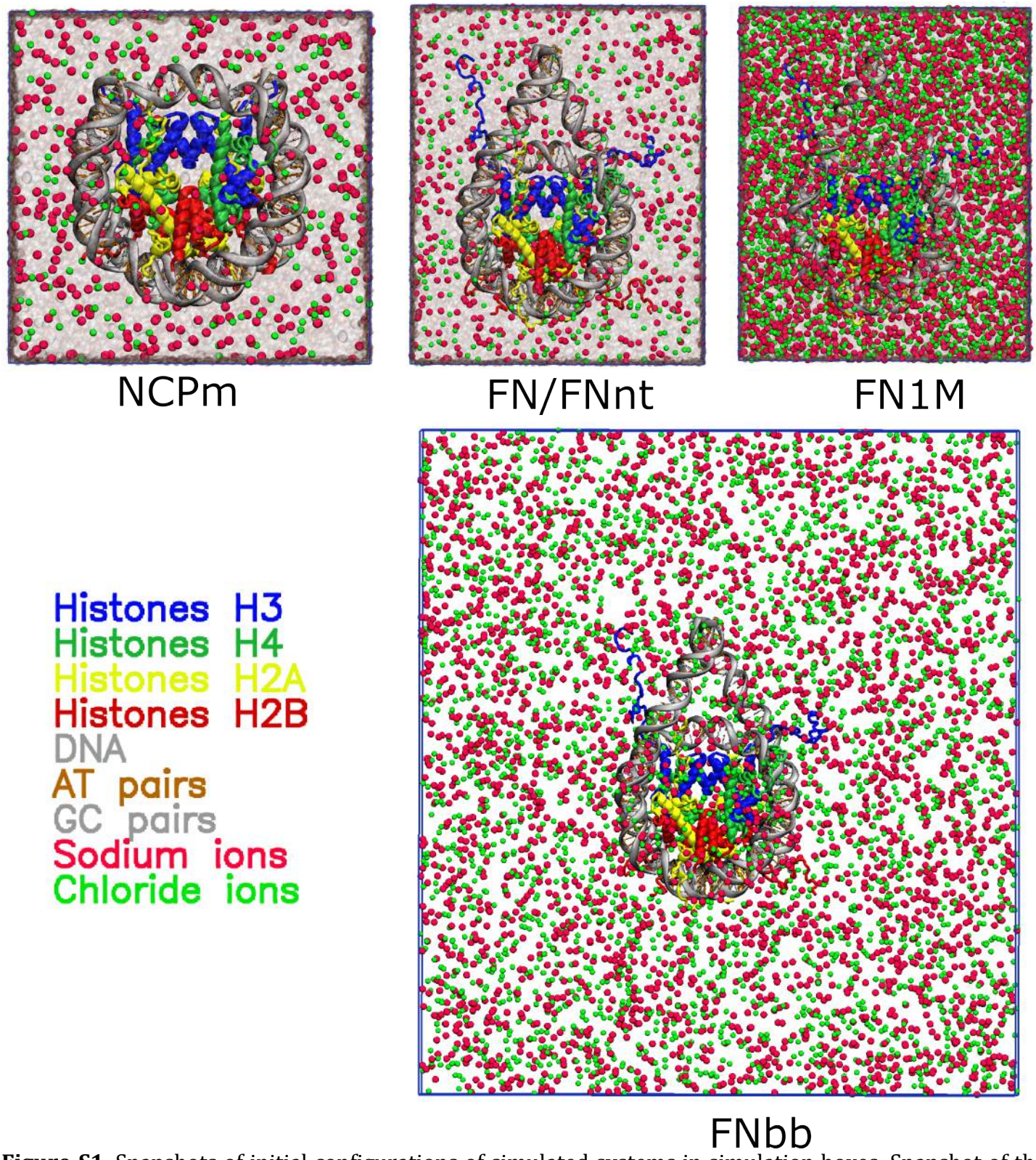

Figure S1. Snapshots of initial configurations of simulated systems in simulation boxes. Snapshot of the initial configuration of the atomistic model of nucleosome fully solvated in a simulation box. Sodium and chloride ions are shown with magenta and green spheres, respectively. See **Tables 1** and **S1** for system description and abbreviations.

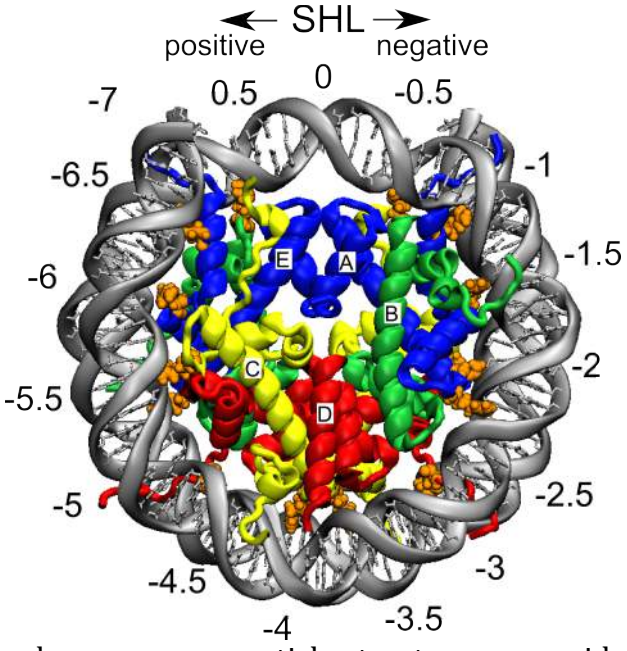

Figure S2. Detailed view of nucleosome core particle structure as provided by crystallography (PDB ID 1kx5). DNA, histones H3, H4, H2A and H2B are depicted in grey, blue, green, yellow, and red respectively. Histone tails are truncated at positions specified in **Figure 2**. The side chains of 14 arginines penetrating in DNA minor groove are shown in orange. The super helix location (SHL) parameters along the DNA are marked. The one-letter chain IDs of front facing histones are depicted.

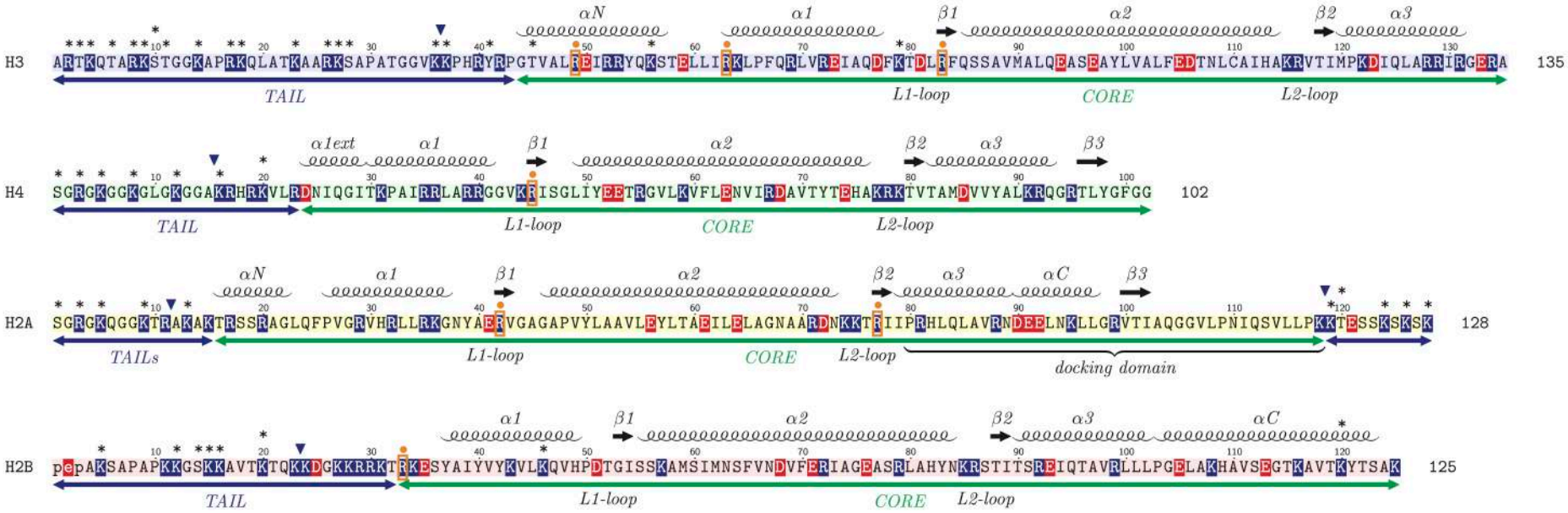

Figure S3. Histone sequences, location of key structural elements, designation of tail and globular core domains. Asterisks denote common post-translationally modified residues as provided by http://www.actrec.gov.in/histome/ [1].

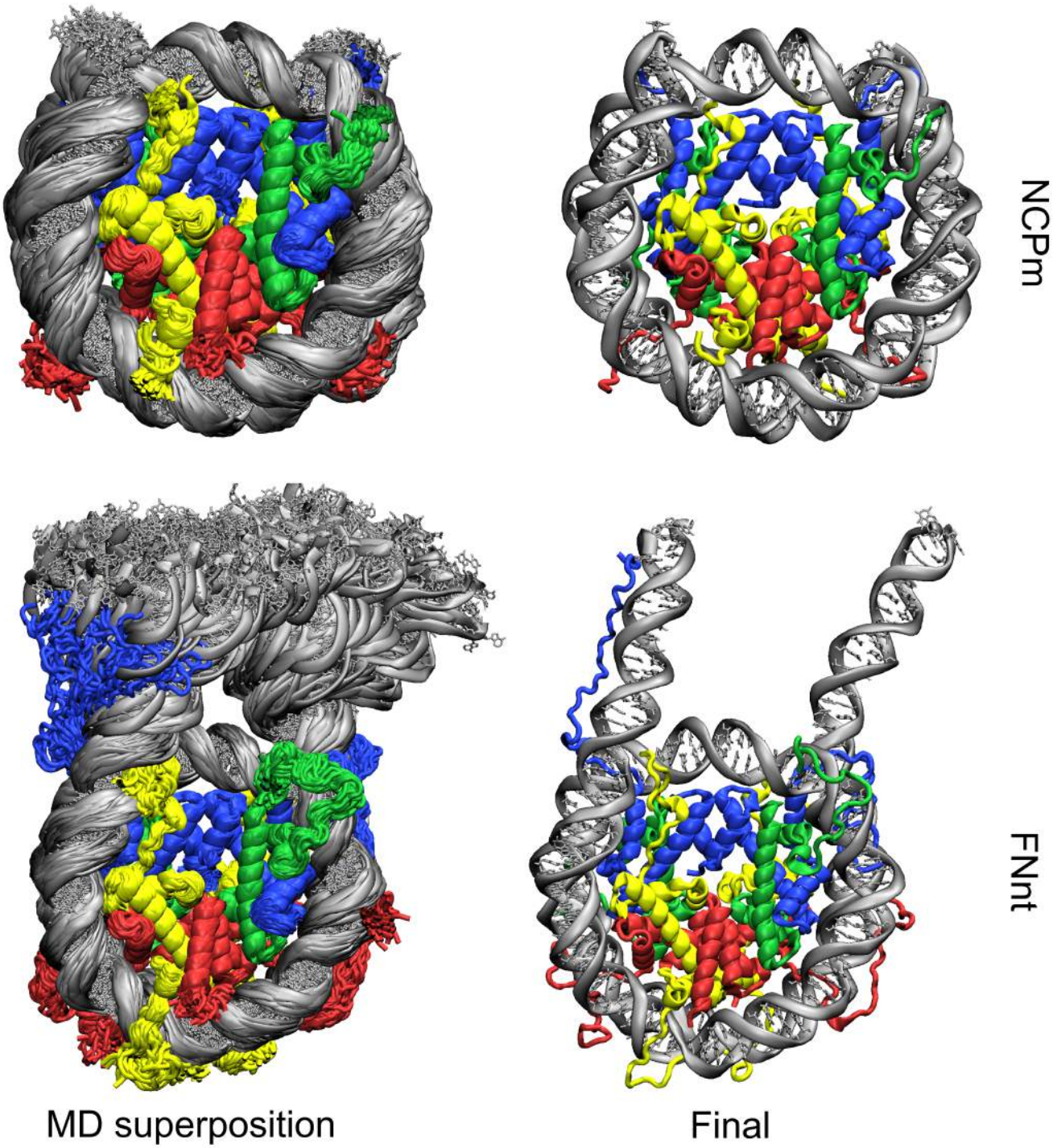

Figure S4. Overview of the dynamics in supporting simulation systems. 150 MD simulations snapshots equally spaced in time during the last 750 ns of the MD simulations trajectory overlaid on top of one another and the last snapshot after 1 μs of simulations are shown for FNnt and NCPm systems.

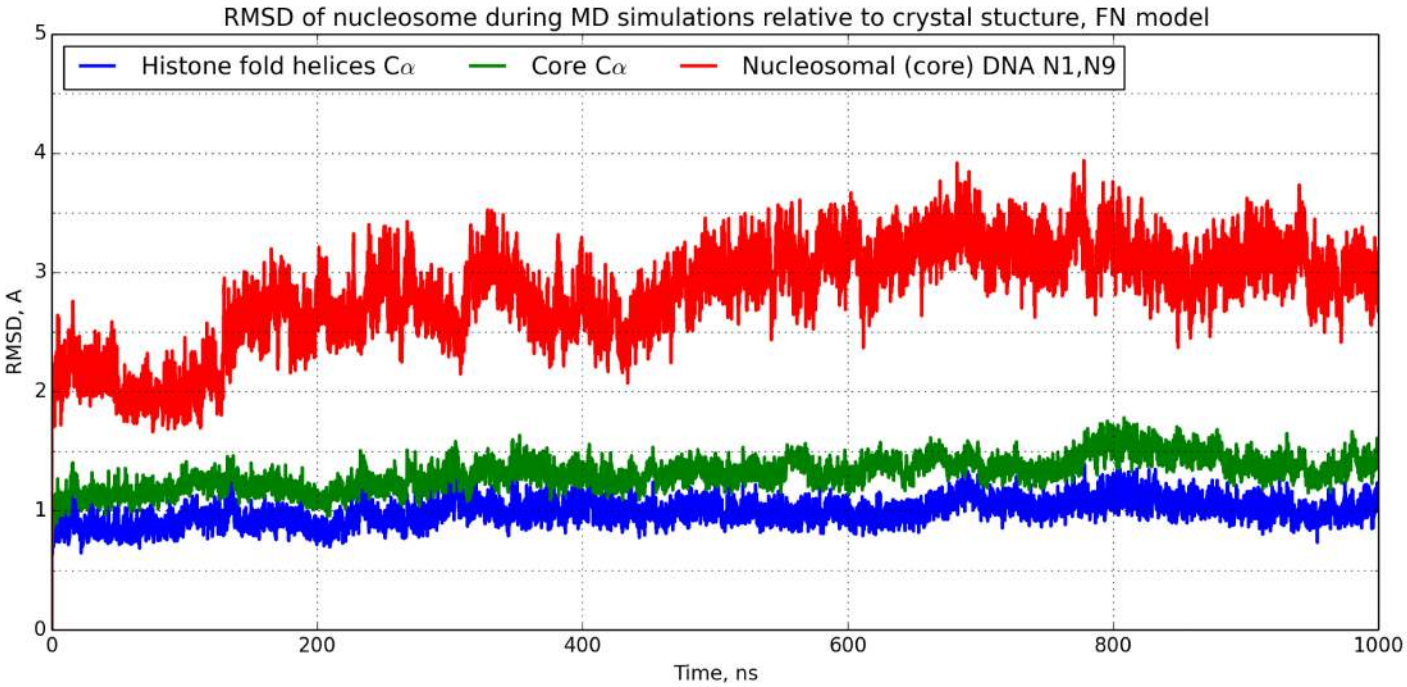

Figure S5. The root-mean-square deviation (RMSD) of certain atomic groups with respect to their X-ray positions throughout the FN model simulation. The plots for RMSD of Cα-atoms of the histone globular region (as defined in **Figure 2**), of the histone fold helices (α1,α2,α3), and N1, N9 atoms (nitrogen atoms of the bases participation in glycosidic bonds) of nucleosomal core DNA (linker DNA excluded) are shown.

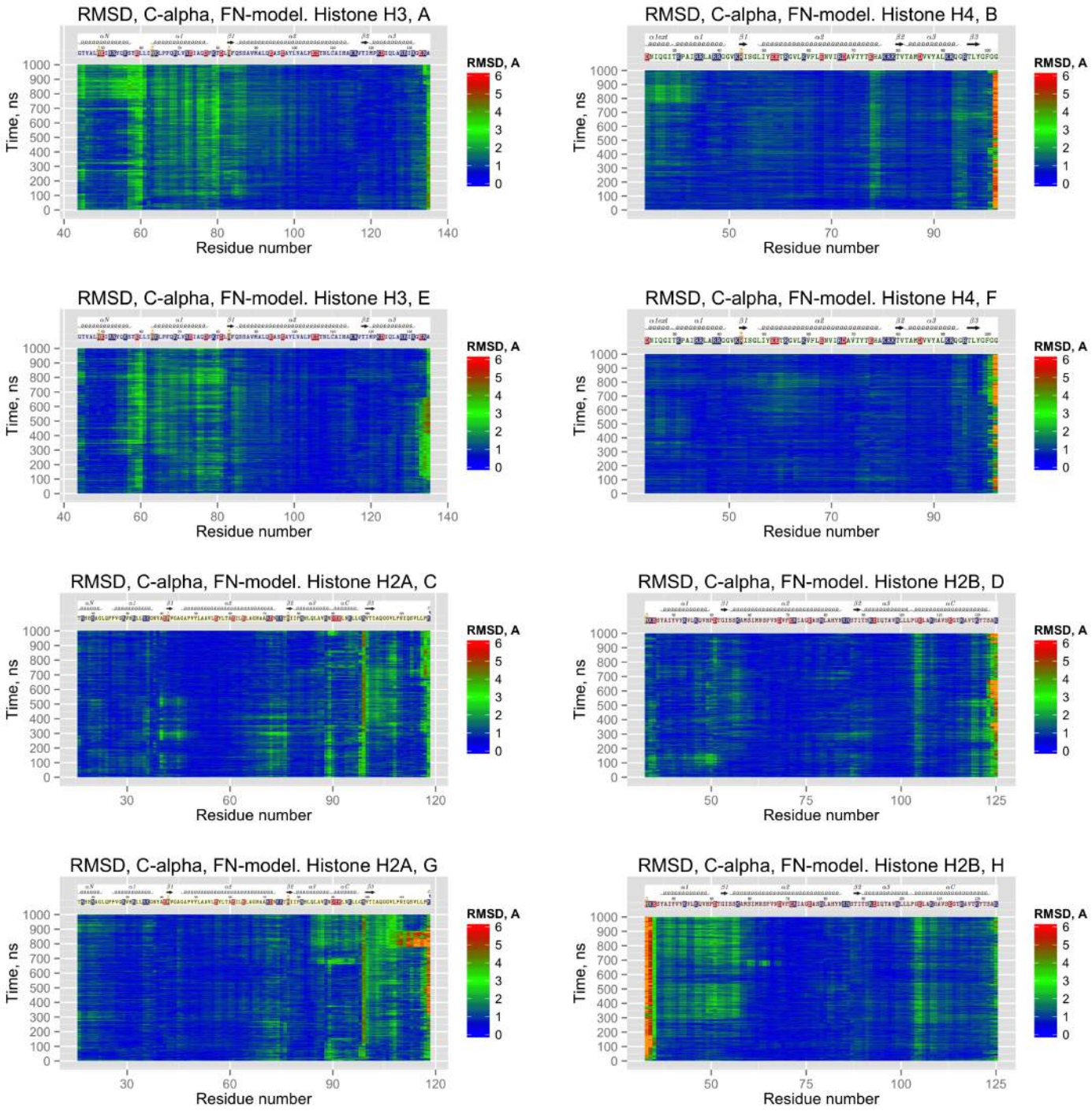

Figure S6. Heatplots of Cα-atoms position evolution relative to their positions in X-ray structure during simulations of FN model. Orange color highlights deviations higher than 6 Å. This is essentially a part of **Figure 2** unwrapped versus time and identical histone chains.

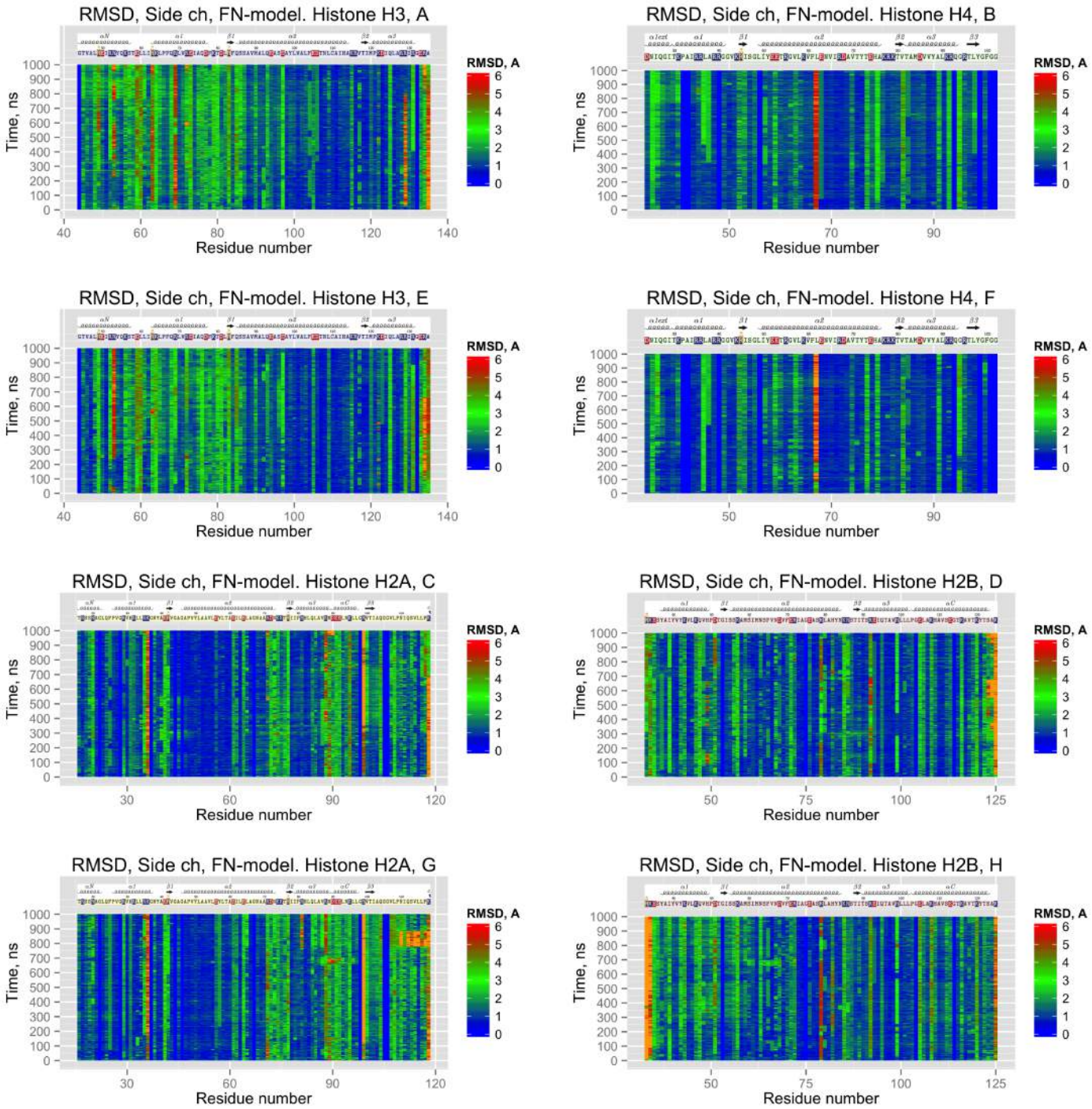

Figure S7. Heatplots of side chain atoms RMSD evolution relative to their positions in X-ray structure during simulations of FN model. Orange color highlights deviations higher than 6 Å. This is essentially a part of **Figure 2** unwrapped versus time and identical histone chains.

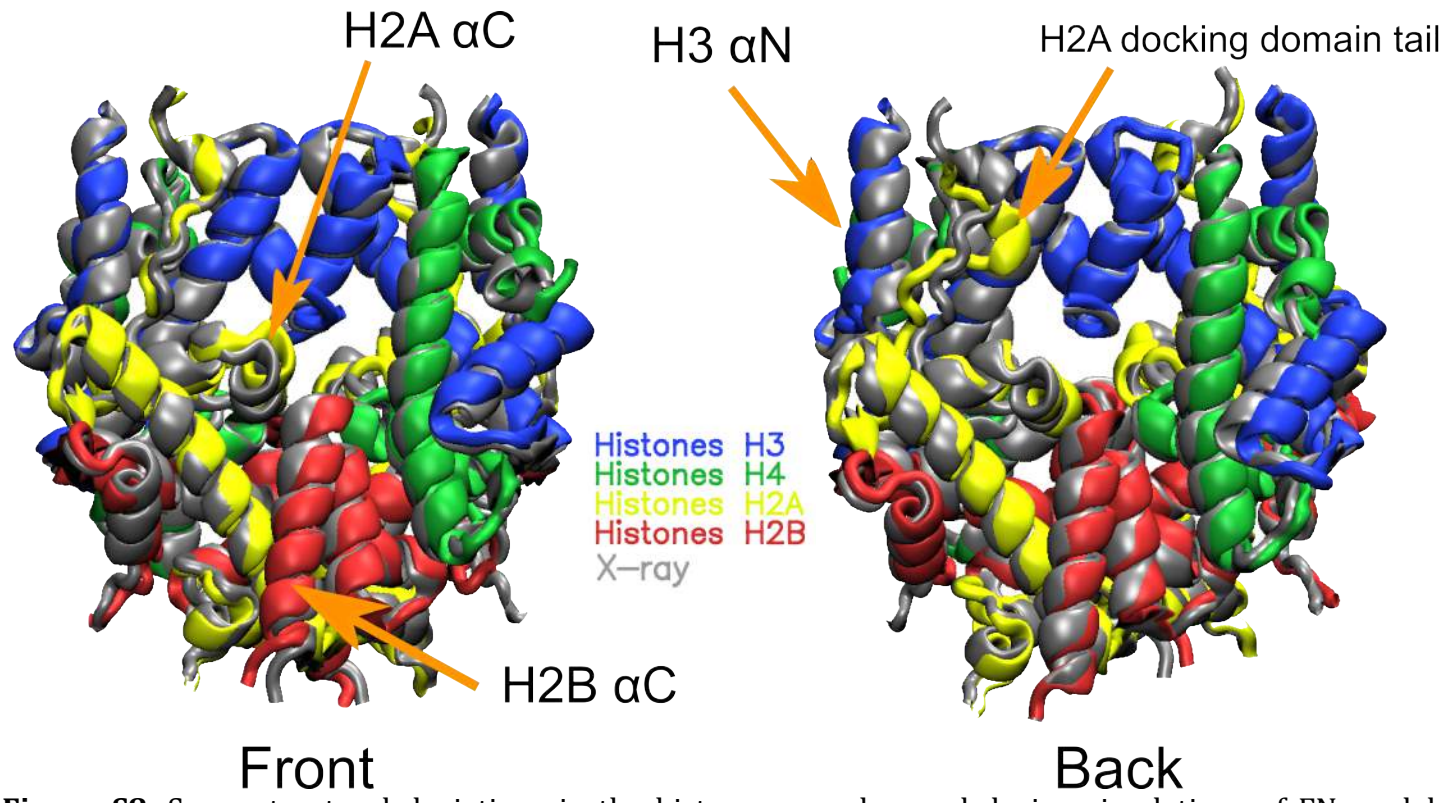

Figure S8. Some structural deviations in the histone core observed during simulations of FN model. Original crystal structure is shown in grey.

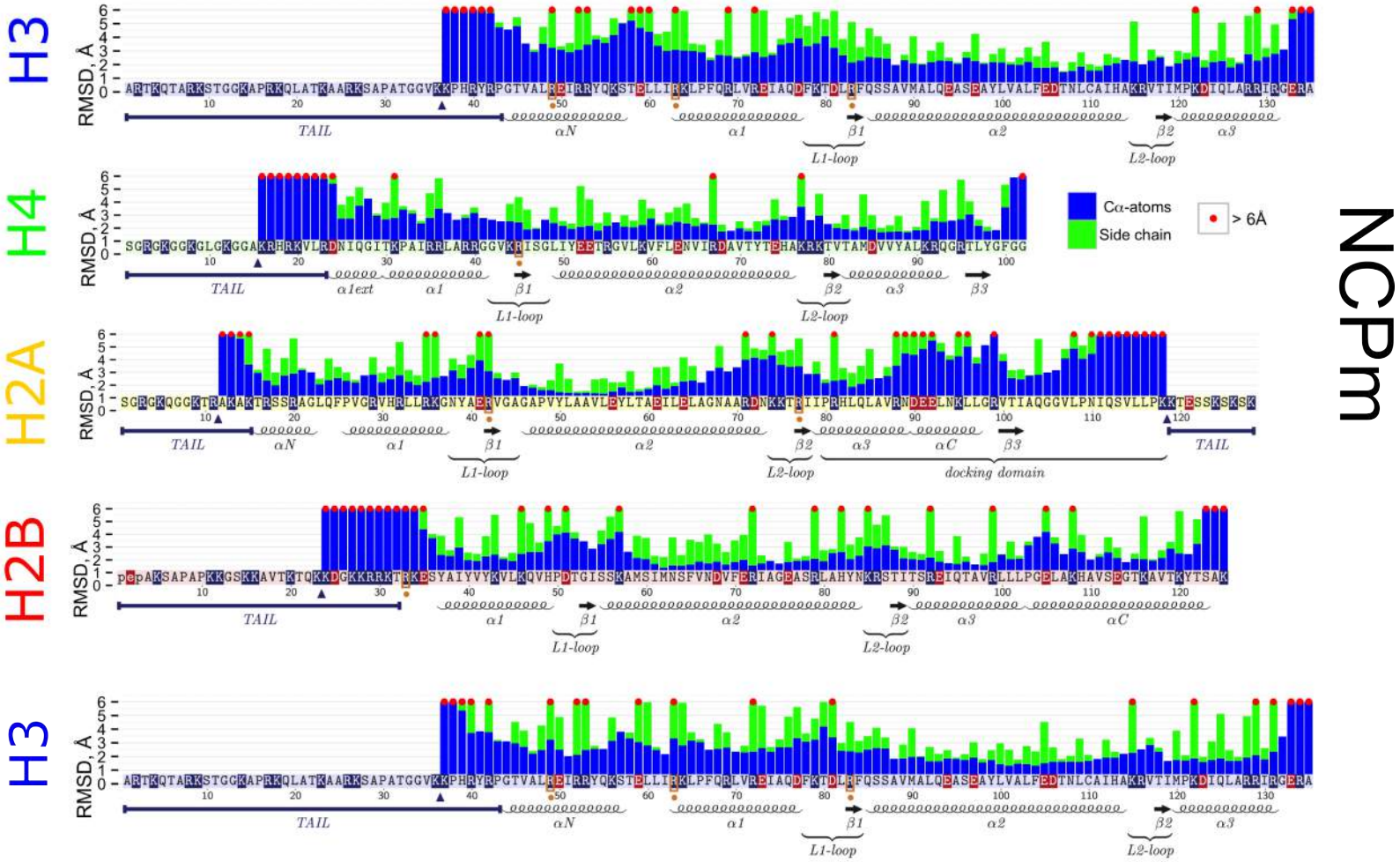

Figure S9. Deviation of atomic positions during 1 μs simulations from initial crystal structure along the histone sequence for NCPm model simulation. These plots are analogous to **Figure 2** in the main text, but for a different simulation system.

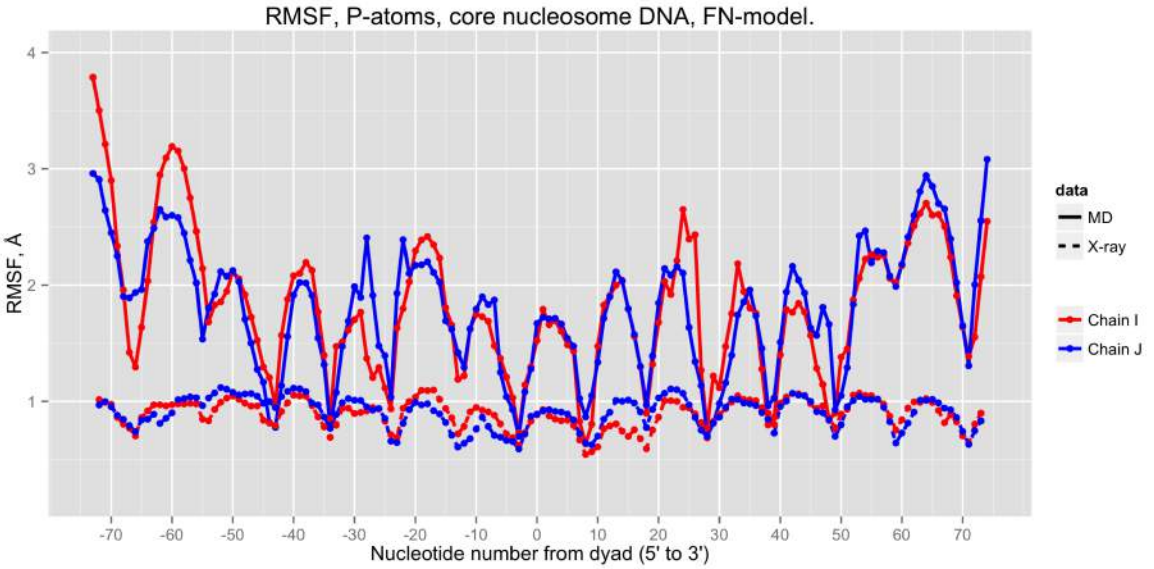

Figure S10. FN model: Root mean square fluctuation (RMSF) of atomic positions for phosphorus atoms along the backbone of DNA strands over the last 750 ns of MD trajectory and RMSF calculated from B-factor of X-ray structure according to the formula $RMSF=(B/(8\pi^2))^{0.5}$. Both DNA chains are numbers in 5' to 3' direction.

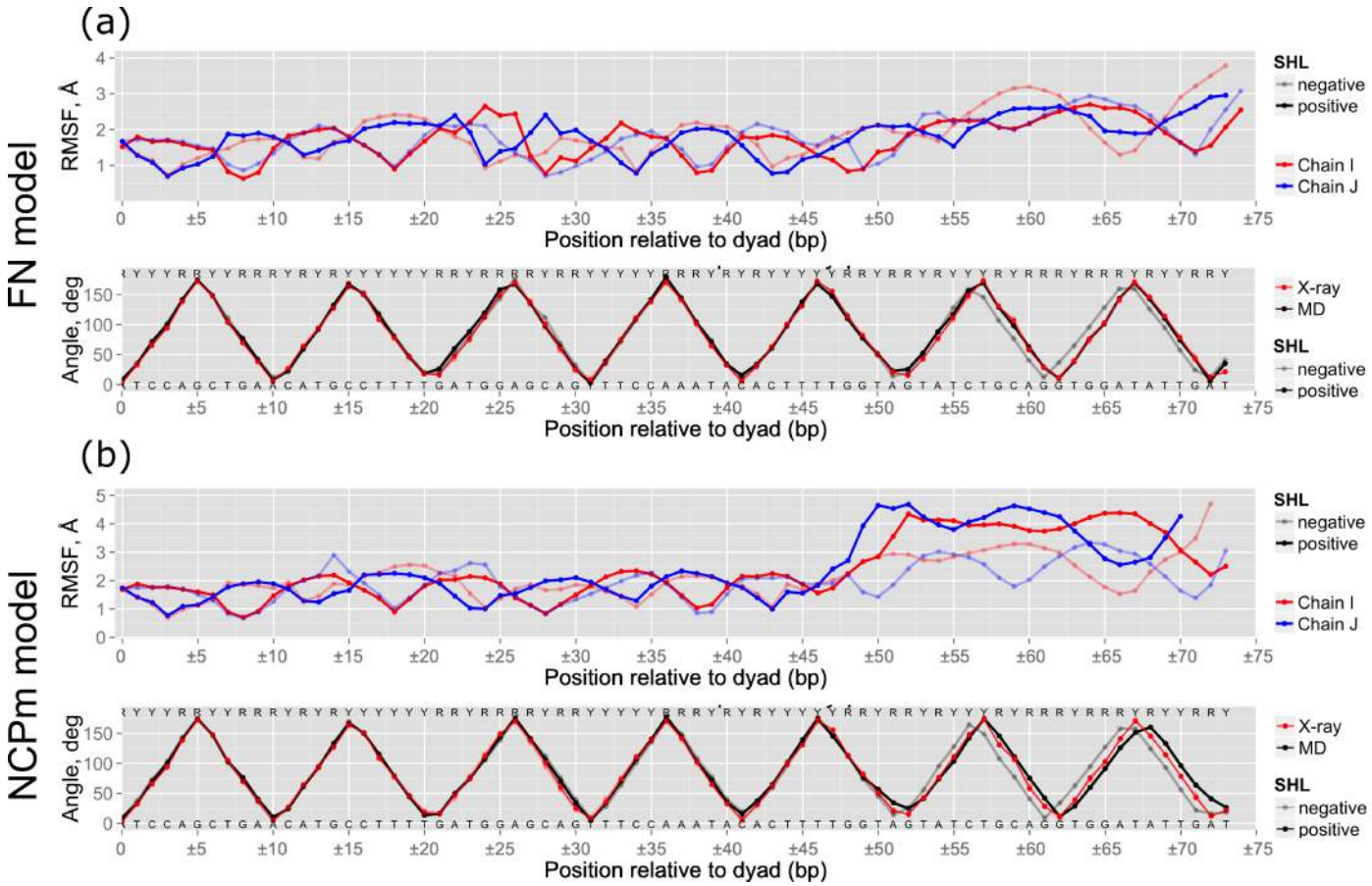

Figure S11. The symmetrized plots of DNA phosphorus atoms RMSF and DNA rotation for FN and NCPm models simulations. The RMSF plots are analogous to **Figure S10**. The DNA rotation is characterized by the angle between the nucleosome superhelical axis and the vectors connecting N1 and N9 atoms in base pairs (the nitrogen atoms of the glycosidic bonds). Note the fluctuations and distortions in the region SHL > 5.0.

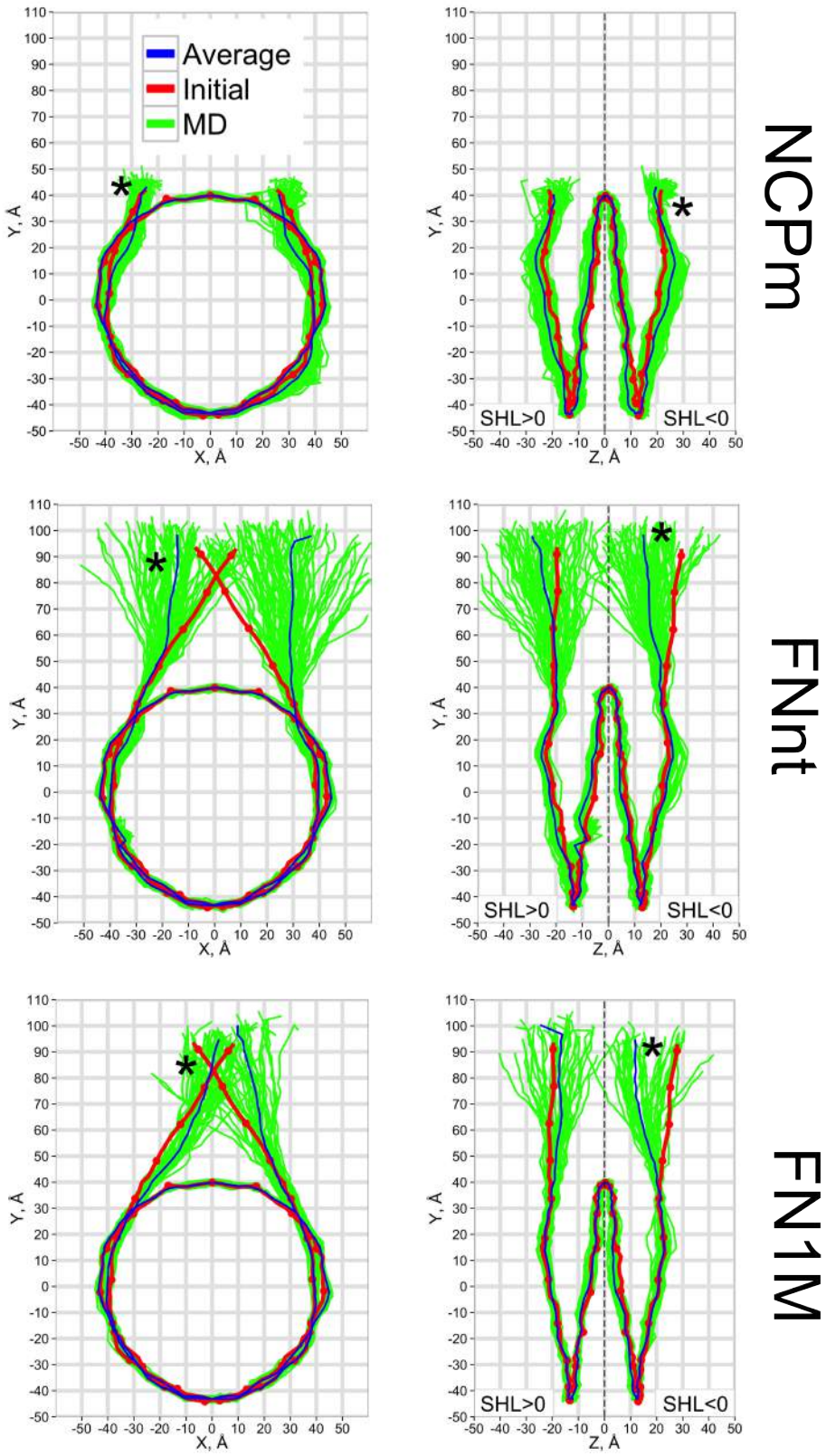

Figure S12. Conformational evolution of DNA during simulations in NCPm, FNnt and FN1M models. Dynamics of DNA path conformation is visualized by the polygons connecting the base pair centers' positions in two projections onto superhelix reference frame, (a) front projection, (b) side projection. The initial DNA conformation (red), ensemble of conformations depicted by snapshots captured every 10 ns along the whole trajectory (green) and average conformation during the last 3/4 of simulation trajectory (blue) are plotted. The red dots mark the integer and half integer SHL values. An asterisk highlights the same linker DNA segment on both plots for clarity.

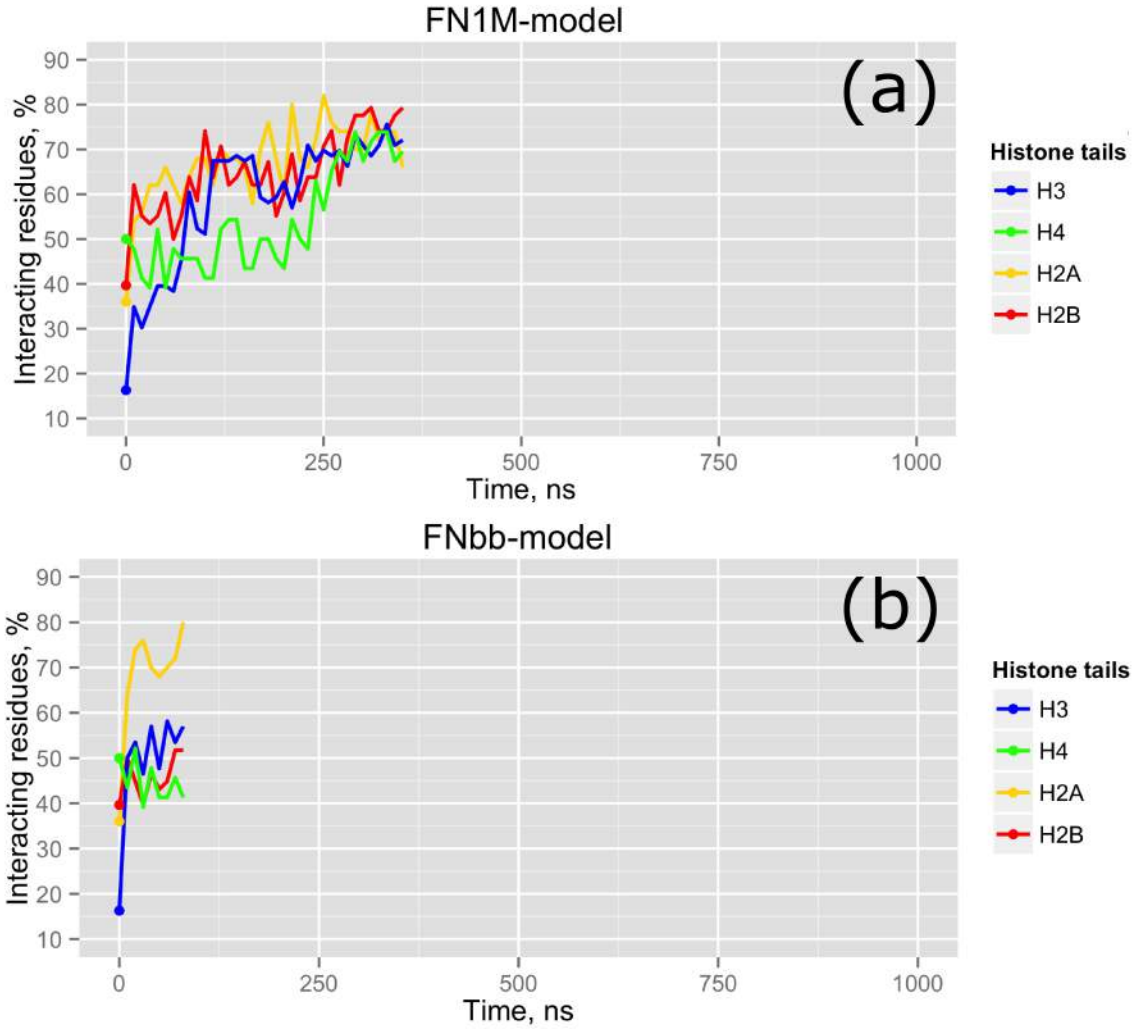

Figure S13. Evolution of histone tails condensation onto DNA during the simulations of (a) FN1M and (b) FNbb models: the percentage of amino acid residues within tail regions of histones (as defined in **Figure 2**) that have direct or water mediated interactions with DNA is plotted as a function of simulation time. Compare with **Figure 4**: even under high ionic salt conditions (a) or in a bigger solvent box (b) the tails show tendency to be adsorbed onto DNA

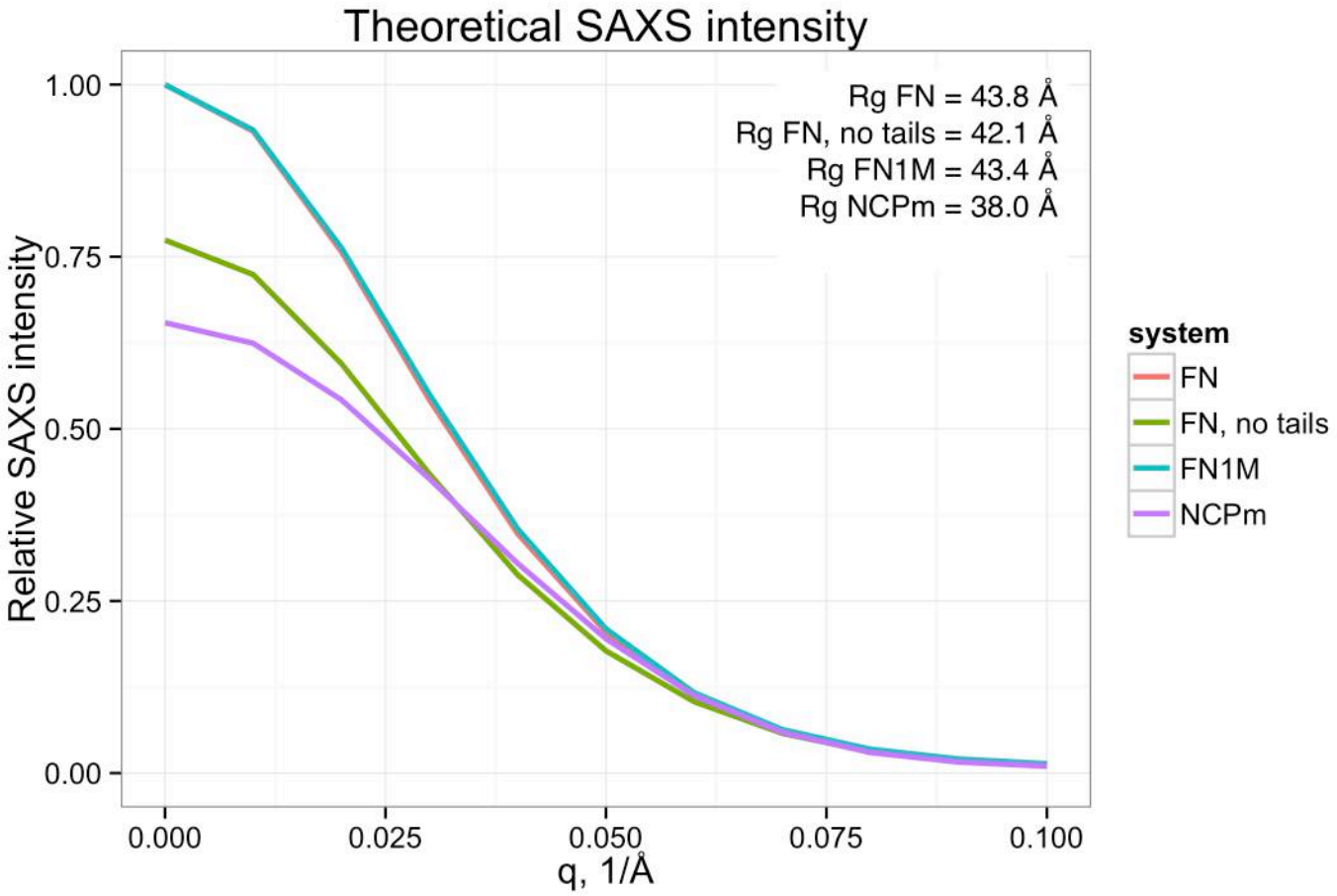

Figure S14. Theoretical small angle X-ray scattering curves for nucleosome were generated from MD simulations using CRYSOL software by averaging one hundred snapshots equally spaced along the trajectory [2]. Plots for FN, FN1M and NCPm systems are presented. "FN, no tails" curve was generated by omitting histone tails in calculations based on FN system. The corresponding gyration radius values are provided in the inset. While the absence of tails in the "FN, no tails" calculation may reduce SAXS intensity the gyration radius is not significantly affected due to presence of linker DNA strands.

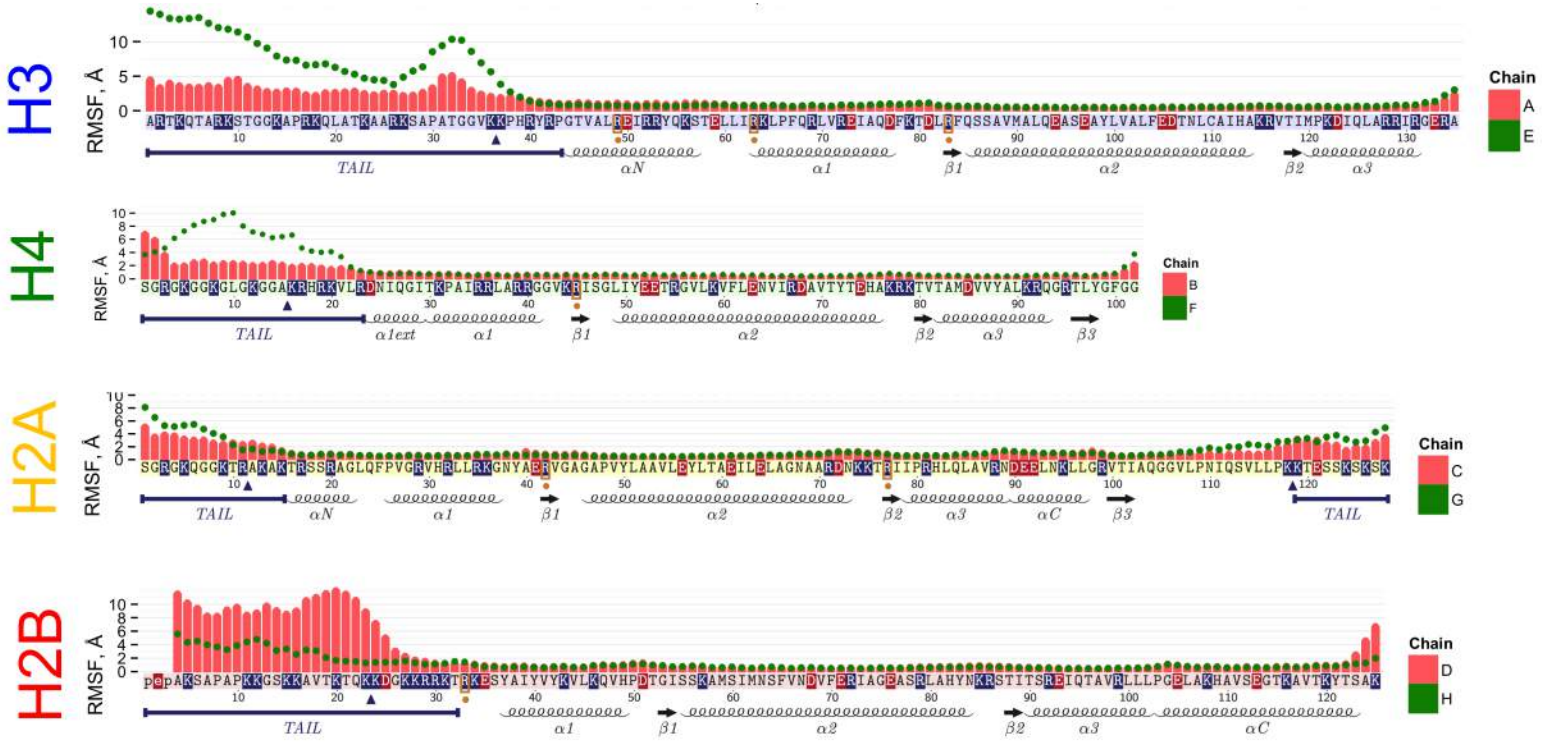

Figure S15. FN model: Root-mean-square fluctuations (RMSF) of Cα-atoms positions for each amino acid during last 750 ns of MD simulations.

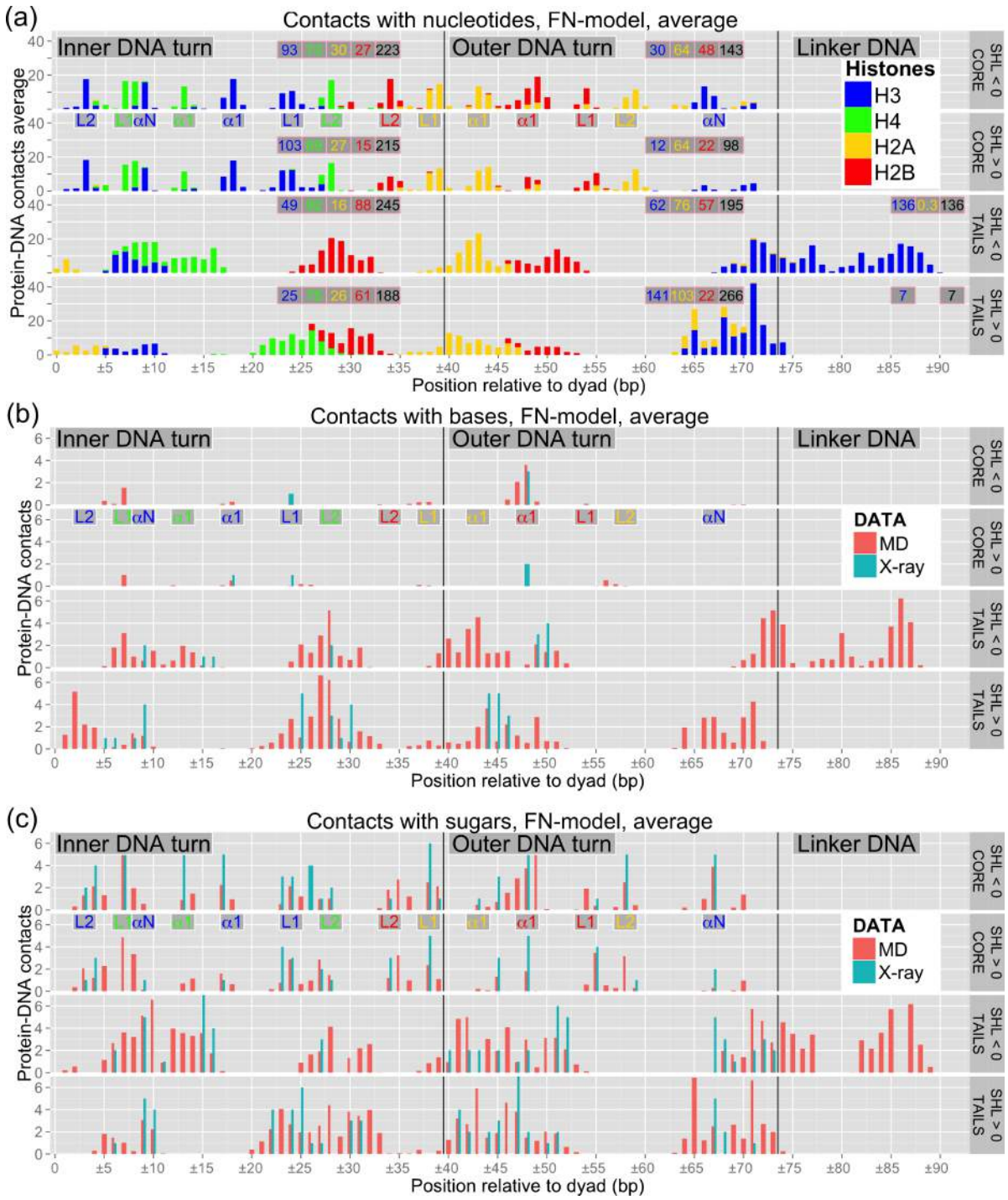

Figure S16. FN model: detailed analysis of protein-DNA interactions in nucleosome along the DNA sequence and comparison with X-ray structure. (a) The number of simple atom-atom contacts (as defined in methods) between histones and nucleotide pairs along the DNA double helix in nucleosome are represented with stacked bars colored according to the contributing histone type. Each plot consists of four subplots, presenting separately the interaction of histone core part and tails for both pseudo symmetric halves of the nucleosome. Quantitative data for the combined number of contacts in each DNA segment is shown in pink-framed boxes. The structural elements of the DNA binding sites are annotated on top of the second subplot. Number of contacts in averaged over the last 750 ns of simulation. Note DNA-protein interactions in the linker DNA region, as well, as asymmetry in the outer turn of nucleosomal DNA. (b) The plot showing the number of contacts with DNA bases average for MD simulations and for X-ray structure. (c) The plot showing the number of contacts with sugar moieties of DNA average for MD simulations and for X-ray structure.

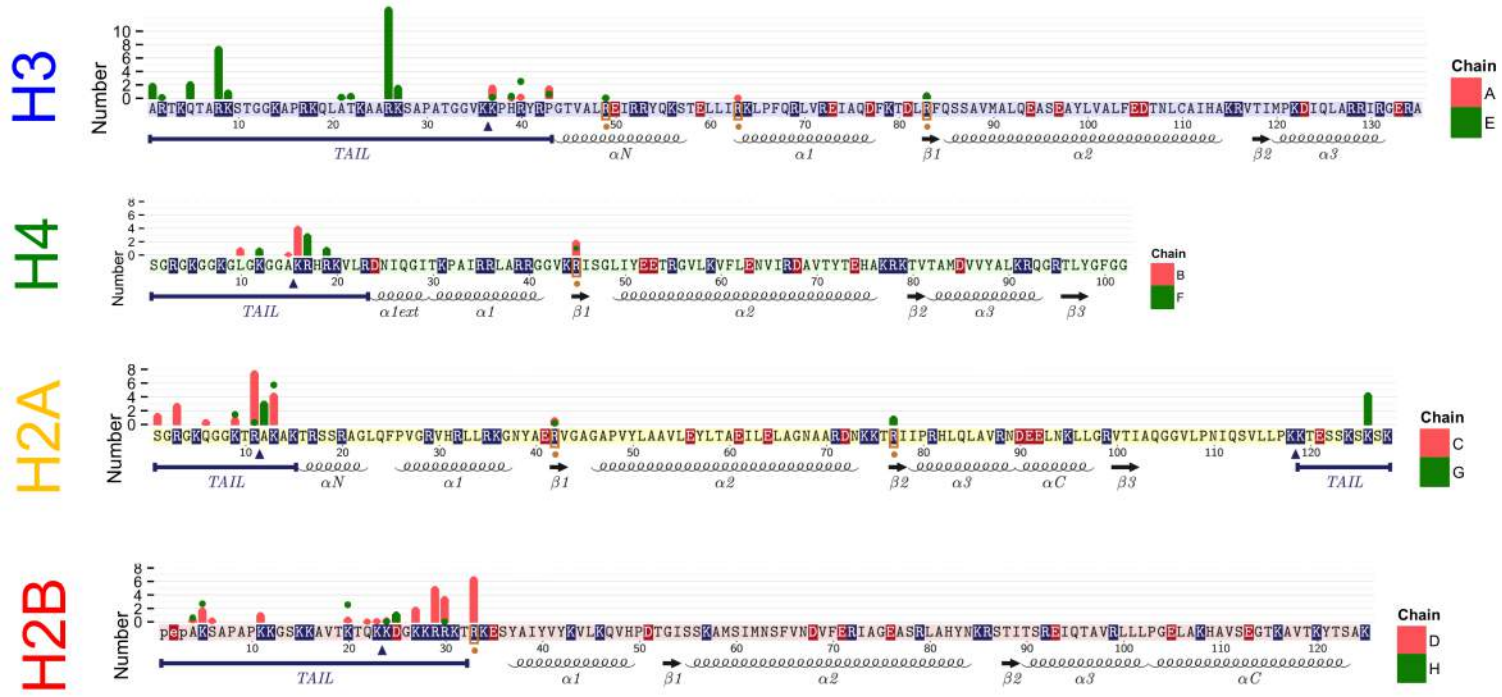

Figure S17. FN model: the average number of simple atom-atom contacts between protein side chains and bases in DNA minor grooves for every amino acid. The data is averaged over last 750 ns of simulations.

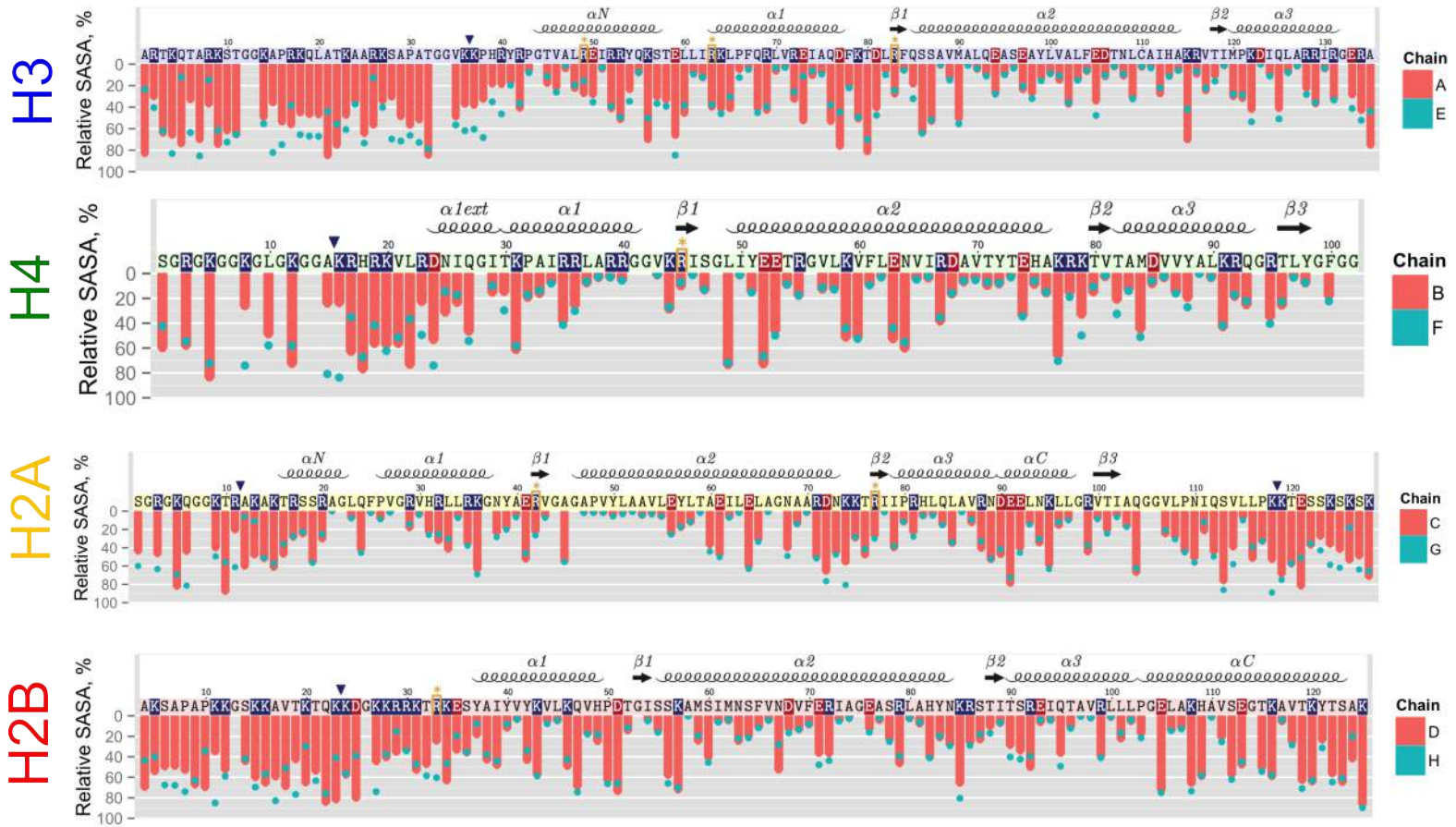

Figure S18. Solvent accessibility of histone side chains. The solvent accessible surface area (SASA) was determined for the amino acid side chains relative to its value obtained in a model linear β–strand peptide and averaged of last 750 ns of FN system simulation. Values for one of the chains is presented with bars, while for the other with dots. Note the potentially highly occluded state of H3Q5, H3R8, H3R26, H2AR11, H2BR29.

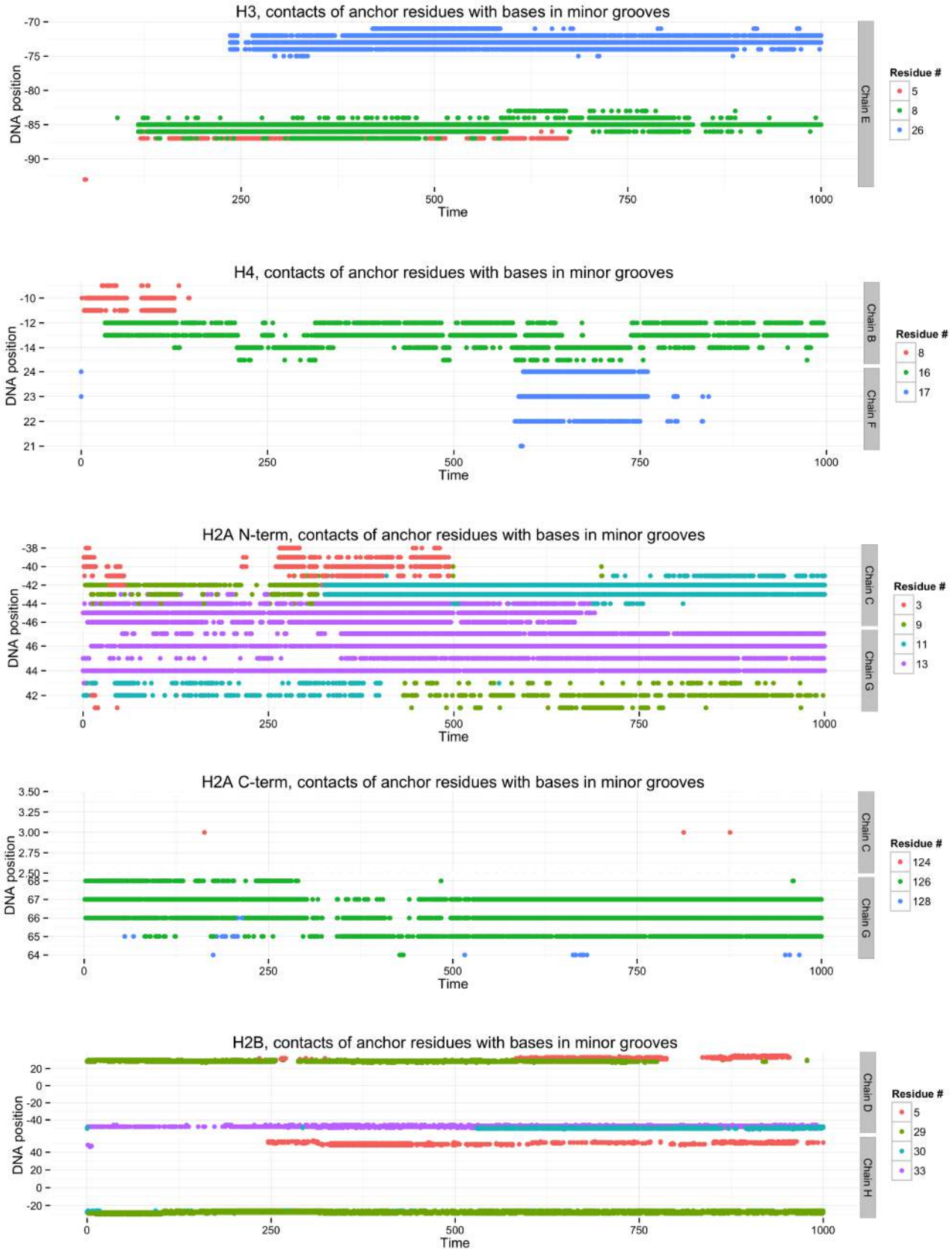

Figure S19. FN model: contact dynamics of certain "anchor" residues. The presence of contacts between side chains of selected residues and DNA bases in minor groove is plotted versus time. If any number of contacts is present a dot is plotted.

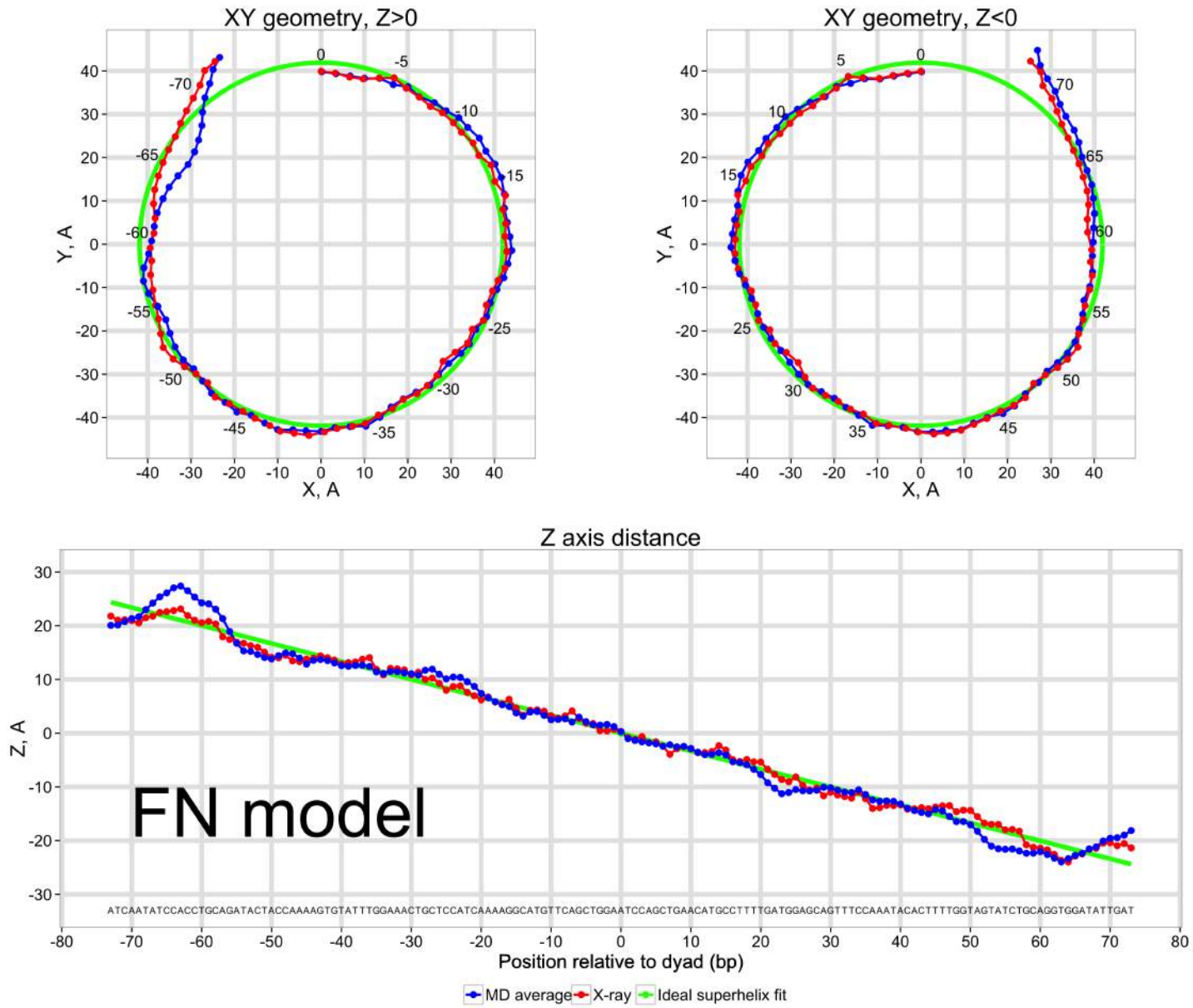

Figure S20. FN model: the average geometry of DNA path in different projections onto nucleosomal superhelical reference frame. The DNA path corresponds to the centers of base pairs as calculated by 3DNA. Averaged over last 750 ns of the trajectory. Positions of individual base pairs are depicted by dots.

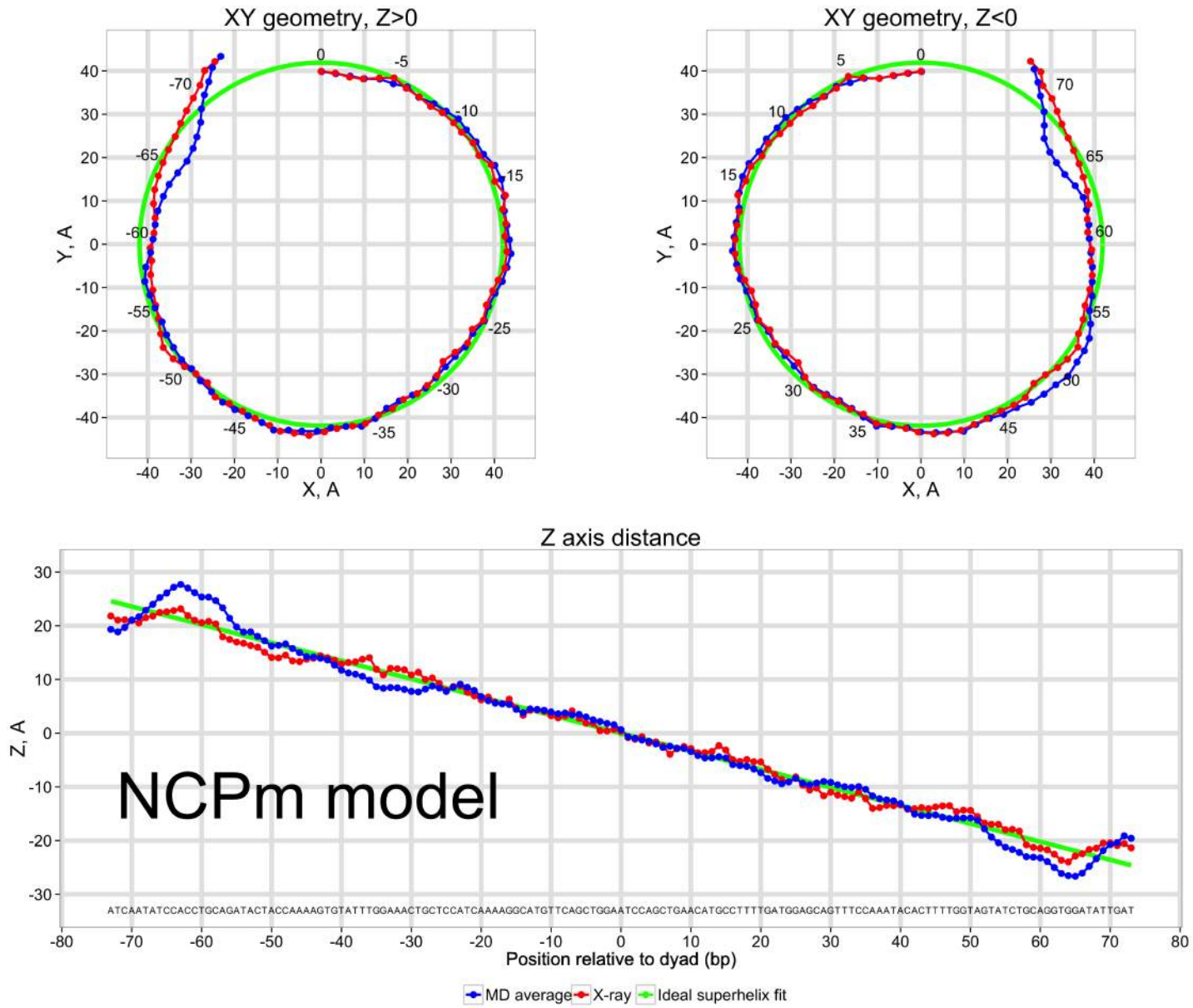

Figure S21. NCPm model: the average geometry of DNA path in different projections onto nucleosomal superhelical reference frame. The DNA path corresponds to the centers of base pairs as calculated by 3DNA. Averaged over last 750 ns of the trajectory. Positions of individual base pairs are depicted by dots.

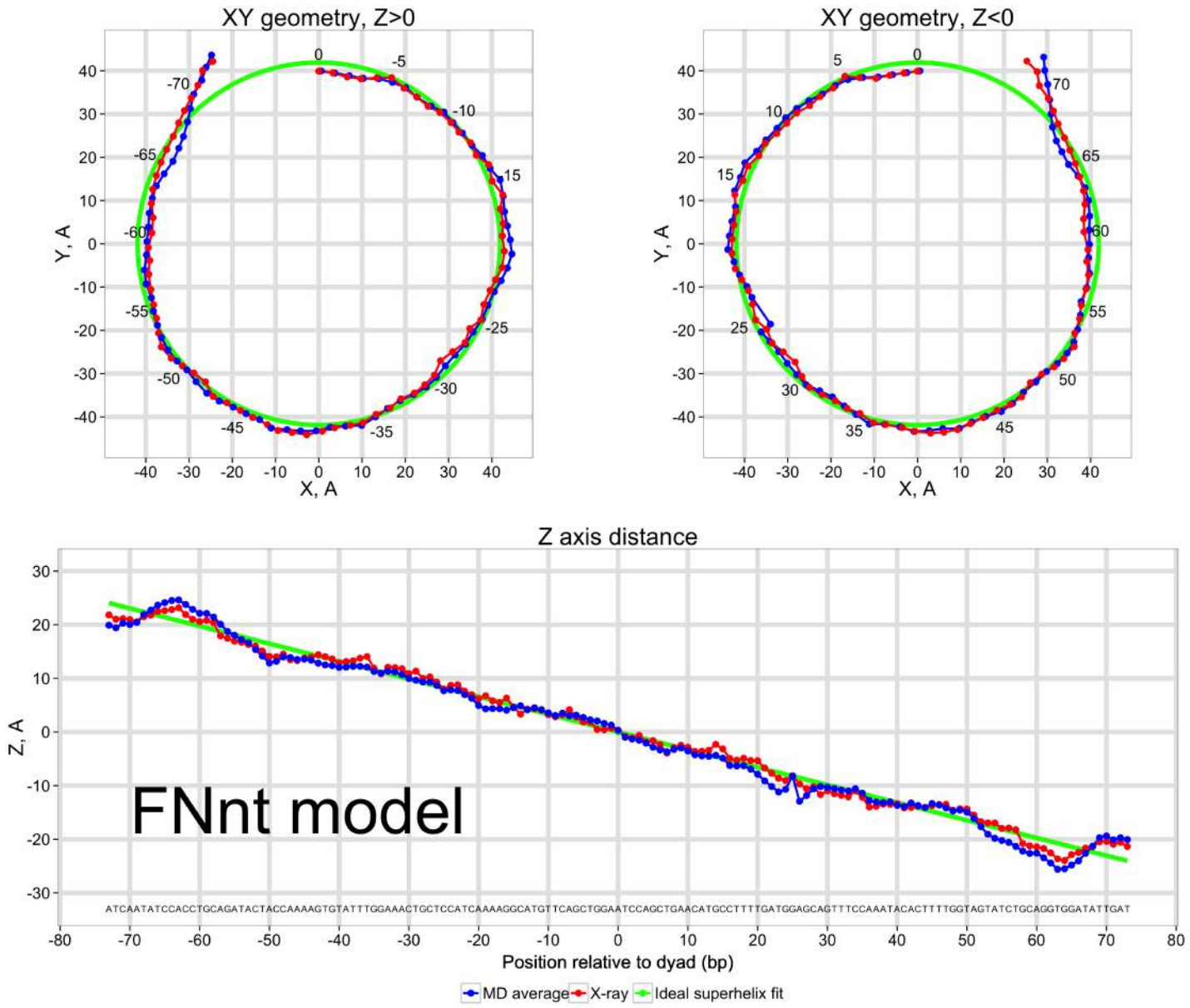

Figure S22. FNnt model: the average geometry of DNA path in different projections onto nucleosomal superhelical reference frame. The DNA path corresponds to the centers of base pairs as calculated by 3DNA. Averaged over last 750 ns of the trajectory. Positions of individual base pairs are depicted by dots.

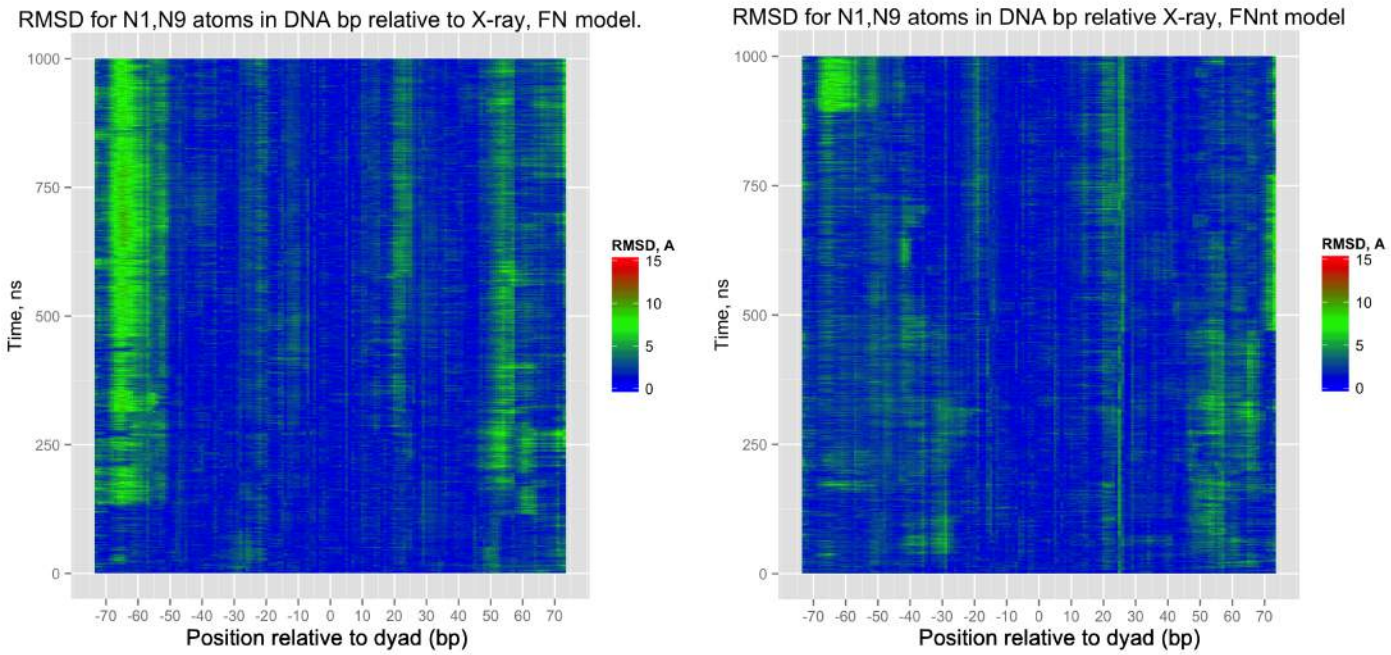

Figure S23. Time evolution of nucleosomal DNA geometry for the FN and FNnt models simulations with respect to its geometry in X-ray structure measured via the root mean square deviations (RMSD) of N1 and N9 atoms in base pairs along the DNA. Note the distortion of DNA around SHL -6/-6.5/-7, +5.5 and +2.5/+3.

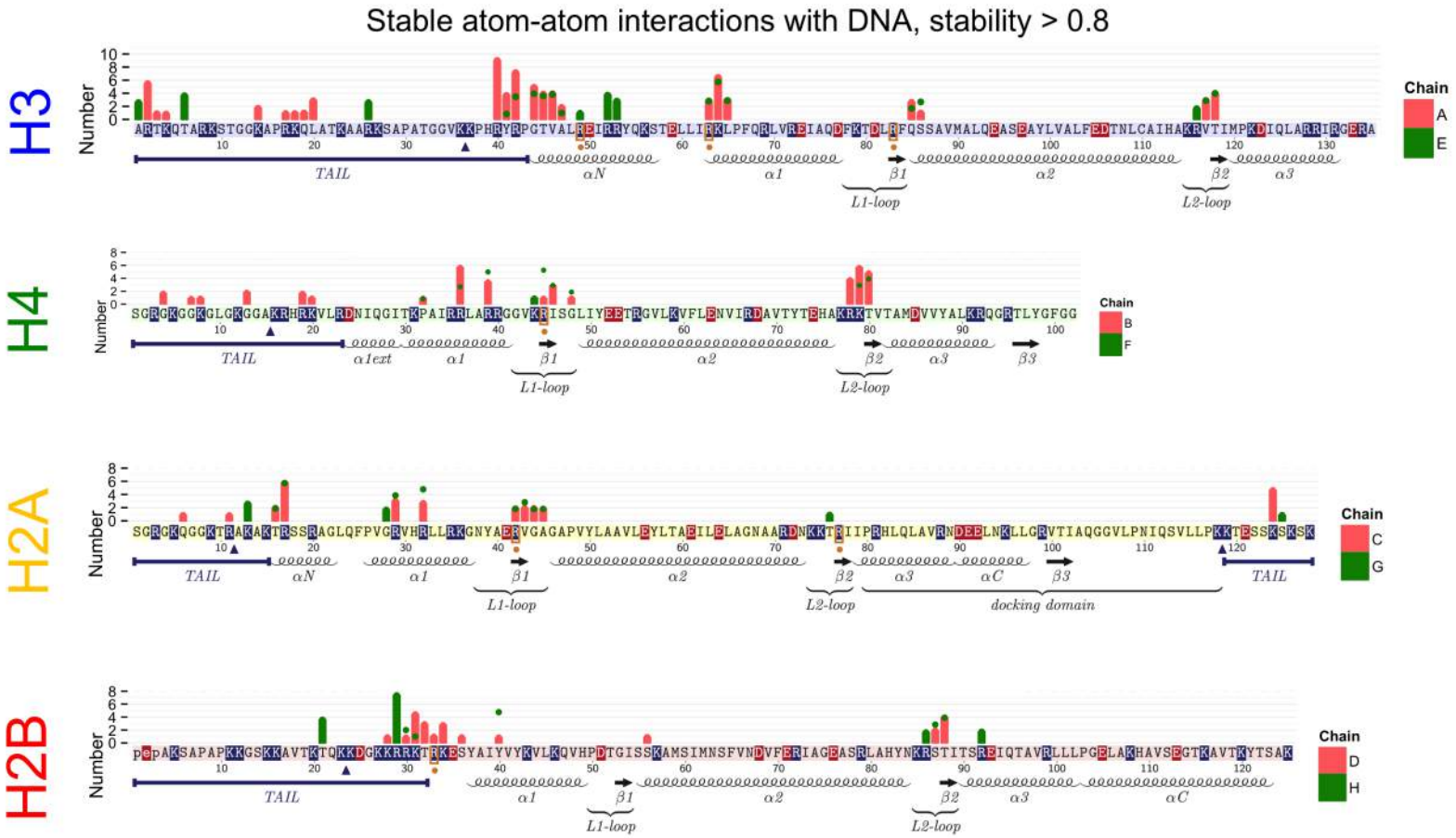

Figure S24. Profiles of histone residues making stable contacts with DNA in FN model simulation. Stable protein/DNA contacts, defined as the individual atom-atom contacts that were present in at least 80% of trajectory frames in the last 750 ns of simulation used for analysis.

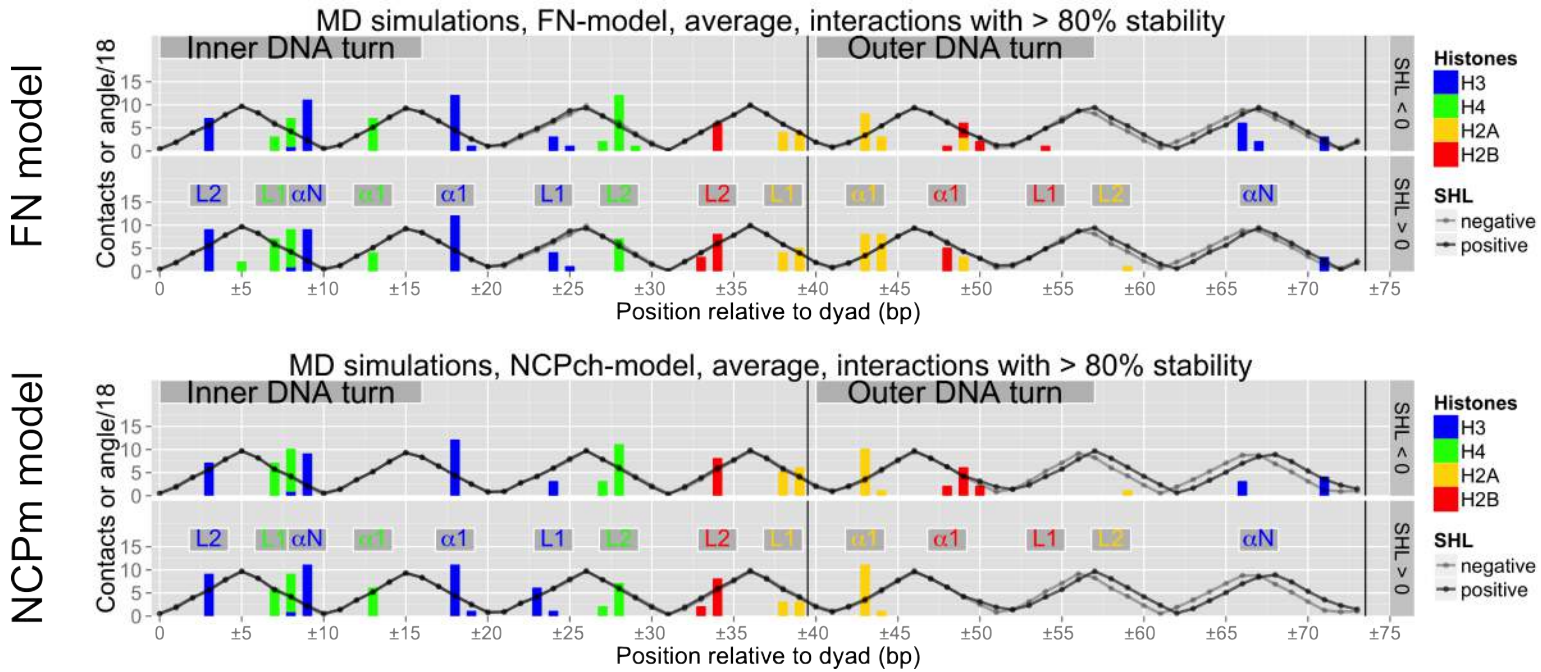

Figure S25. FN and NCPm simulations: plots of stable atom-atom contacts between histone globular core and DNA along the DNA sequence for each side of nucleosome.

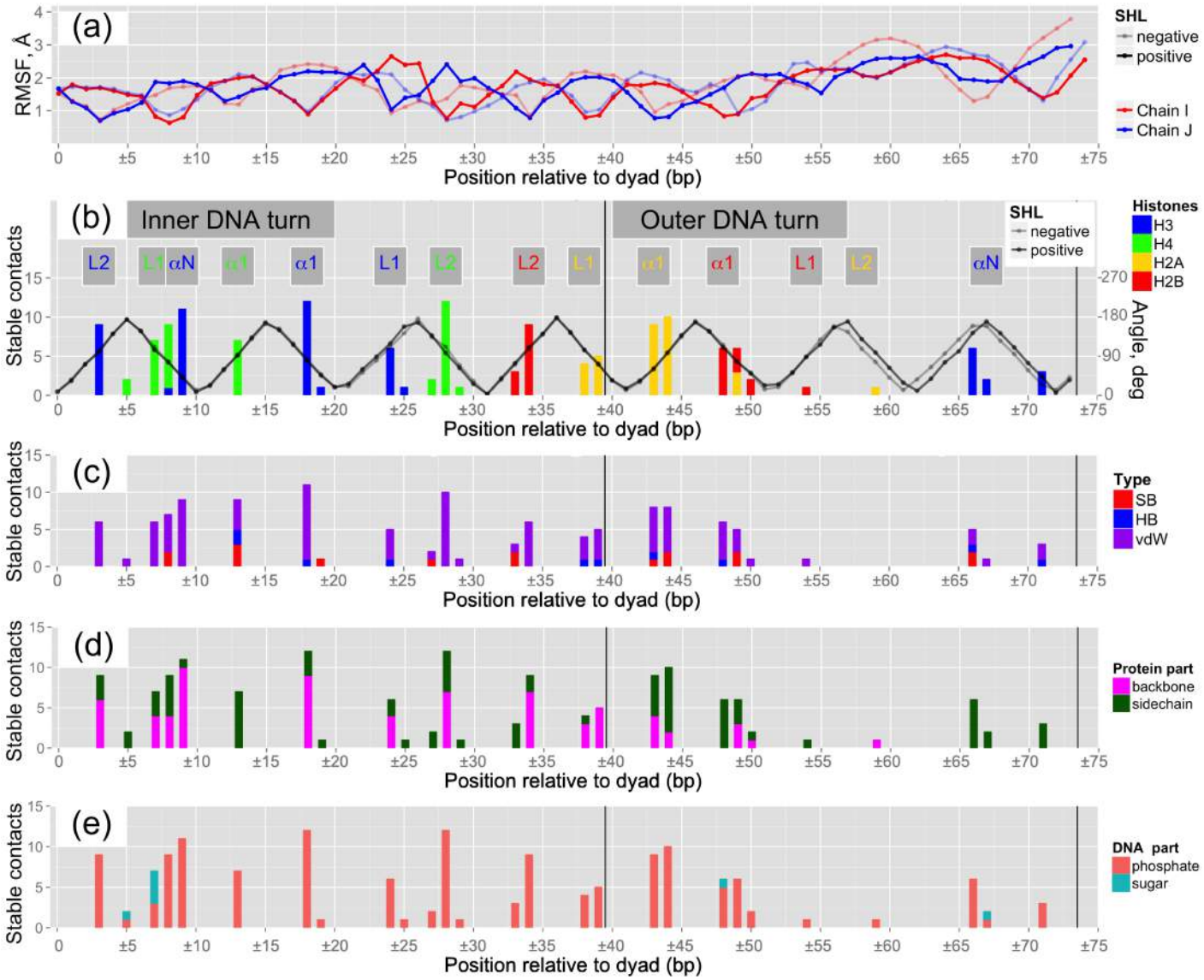

Figure S26. FN model: detailed view and classification of stable interactions between histone globular core and DNA. An expanded version of **Figure 8** in main text. (a) Root mean square fluctuations (RMSF) of phosphorus atoms in the DNA backbone for both DNA chains plotted simultaneously for symmetric positions along nucleosomal DNA. (b) Distribution of stable contacts along nucleosomal DNA for different types of histones. Stable contacts are defined as atom-atoms contacts that where present in more than 80% of MD simulation trajectory frames for either symmetric halves of nucleosome. The black and grey curves show the periodicity of DNA in nucleosome by showing the angle between the base pair vector (connecting N1 and N9 atoms of bases in the base pair) and the nucleosomal superhelical axis. Note the shift of periodicity from 10 to 11 bp in certain cases. (c-e) plot further characterize the obtained set of stable interactions through interaction type, protein part and DNA part.

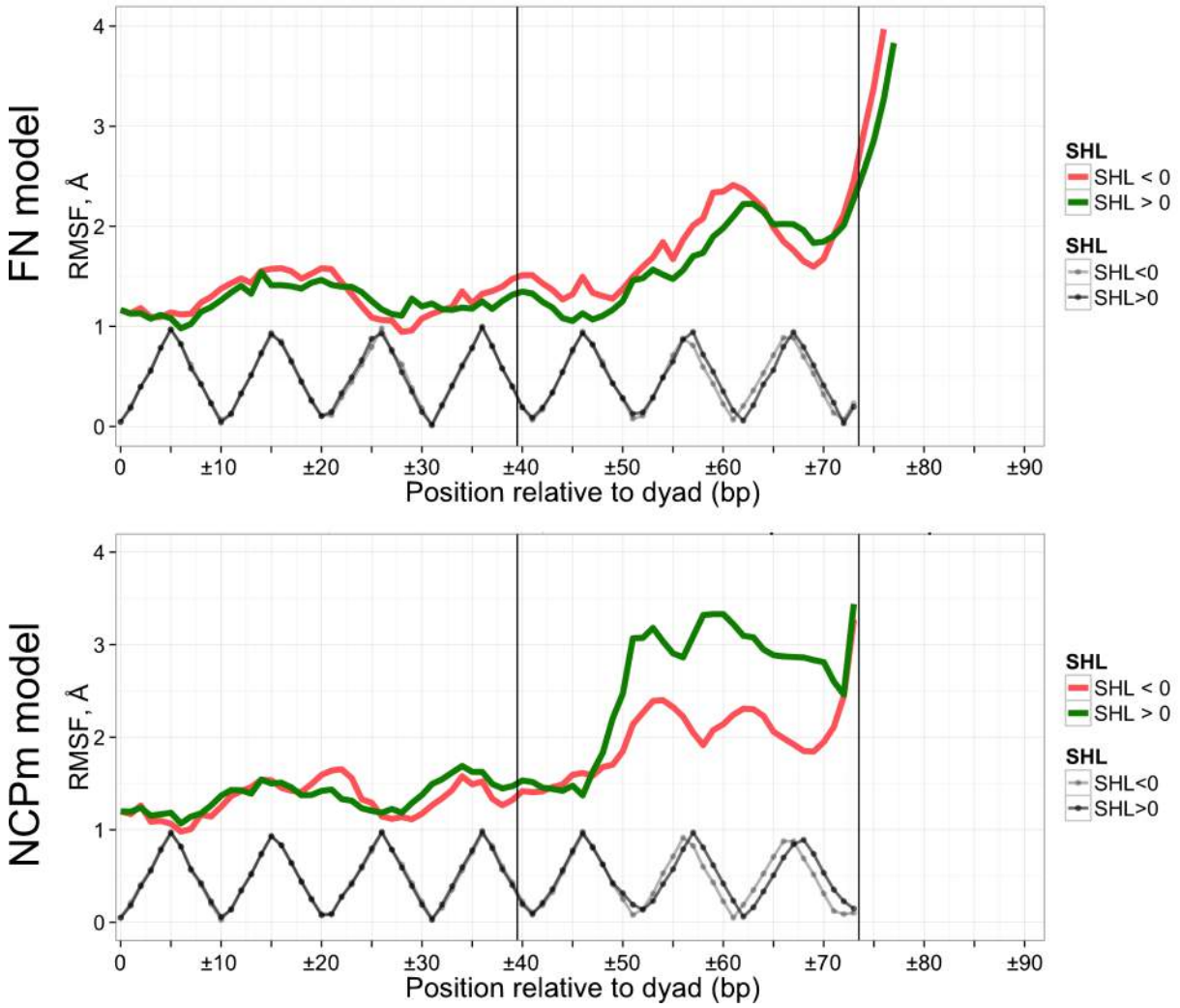

Figure S27. Comparison of nucleosomal DNA fluctuations between FN and NCPm models. The root-mean-square fluctuation (RMSF) of DNA base pair centers with respect to their average positions during the last 750 ns of MD trajectory (red and green curves). The pseudo symmetry of nucleosome is taken into account to show the difference between the symmetric halves. The black and grey curves show the geometry and periodicity of DNA superhelix in nucleosome core similar to **Figure 8**. The data exceeding RMSF of 4 Å is not shown

References

[1] Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic acids research. 2012;40:D337-42.
[2] Svergun D, Barberato C, Koch MHJ. {{\it CRYSOL} {--} a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates}. Journal of Applied Crystallography. 1995;28:768-73.