TASK QUARTERLY 18 No 3, 281–288

MOLECULAR DYNAMICS SIMULATION OF HISTONE H3 AND H4 N-TERMINAL TAIL CONFORMATION IN THE PRESENCE AND ABSENCE OF NUCLEOSOME CORE

RUIHAN ZHANG, JOCHEN ERLER AND JÖRG LANGOWSKI

Division Biophysics of Macromolecules, German Cancer Research Center INF 580, D-69120 Heidelberg, Germany

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Histone N(C)-terminal tails play an important role in chromatin remodeling and gene regulation. Posttranslational modifications (PTMs) of histone tails are known to be correlated to distinct states of gene activity, but the molecular mechanism of these processes is largely unknown. PTMs alter the electrostatic environment and conformation of histone tails, as well as their interaction with other components. Here we performed extensive Replica Exchange Molecular Dynamics (REMD) simulations for the H4 and H3 tails, isolated and with inclusion of a nucleosome. Our results agree with the predictions of previous theoretical studies for the secondary structure of isolated tails, but show strong dependence on the force field used. In the presence of the nucleosome, the secondary structure of histone tails is destabilized. Furthermore, H4K16 is found to insert into the DNA minor groove, whereas the acetylated H4K16 stays on the surface of DNA.

Keywords: histone tail, REMD, H4K16 acetylation, nucleosome

1. Introduction

Approximately 147 base pairs wrap around an octamer of two copies each of histone H3, H4, H2A and H2B [1] to package DNA in eukaryotic cells so as to form the nucleosome (Figure 1). While the central fold domain of each histone is essential for the compact structure of the nucleosome, the disordered and positively charged N(C)-terminal histone tails also contribute to the stability of the nucleosome by interacting with DNA and histone components [2]. Furthermore, the basic residues of histone tails such as lysines and arginines are subjected to post-translational modifications (PTMs), e.g. methylation, acetylation, phosphorylation. PTMs induce gene activation or repression [3] by influencing the recruitment of chromatin-binding proteins or directly altering the dynamics of the nucleosome or chromatin [4]. However, the molecular mechanism of how PTMs changes the

environment and organization of chromatin remains unclear. Our work has been focused on the H3/H4 tail and H4K16 acetylation, which is reported to influence both the chromatin structure and non-histone protein binding [5]. H4K16 acetylation as well as other lysine acetylations on H3/H4 tails cause transcriptional activation, probably related to the destabilization of the nucleosome [6] or unfolding of chromatin [7] which increases the accessibility of DNA. At the atom level, lysine acetylation causes neutralization of the positive charge and increases the residue volume, and thus influences the conformation and interaction of the H4 tail.

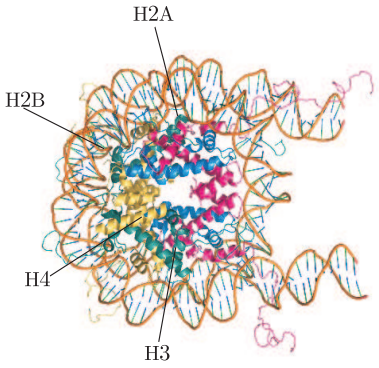

Figure 1. Illustration of the nucleosome model: 147 bp DNA wrapping around two copies of histone H3 (pink), H4 (blue), H2A (green) and H2B (yellow) (PDB ID: 1KX5), plus 10 bp linker DNA on each terminal

Both theoretical and experimental investigations have been performed to reveal the influence of PTMs on the conformation of histone tails. Previous simulation studies have proposed that isolated histone tails adopt a secondary structure. The H4 tail is found to form an α-helix in the region of residues 15–20, K16 acetylation reduces the α-helix forming tendency [8]. However, in another study, β-hairpin is found for the H4 tail when using a different simulation protocol and force field [9]. On the other hand, the electrostatic interaction with nucleosomal DNA influences the conformation of histone tails [10]. The H/D exchange NMR has shown that residues 16–22 of the H4 tail fold onto the nucleosome core, whereas residues 1–15 remain disordered. The K16Q mutation, which mimics the K16 acetylation, induces a structure disorder of the H4 basic patch (K16R17H18R19) [11]. However, the circular dichroism (CD) analysis suggested that the acetylations on histone tails increase the α-helical content of histone tails. Compared to native tails, the acetylated tails form a α-helix regardless of the interaction with DNA [12]. Umbrella sampling simulation has also shown an enhancement of α-helix propensity of the K16 acetylated H4 tail, and stronger binding to DNA than the wildtype tail [13].

Agreement has not been reached about the conformation of disordered histone tails and the influence of PTMs. To what extent does the simulation

protocol, e.g. the time scale and force fields, influence the results? Is there a possibility to improve the model which would more closely resemble the real system? In this study, we employed replica exchange molecular dynamics simulations (REMD) to investigate the conformation of H3 and H4 tails with and without a nucleosome, testing the robustness of the result against different force fields, and exploring the influence of H4K16 acetylation on the binding pattern of the H4 tail on nucleosomal DNA. It is particularly the fact that histone tails are connected to the compact and stable histone core that is taken into account. Under such consideration, the position of histone tails is not completely free but restricted by the heavy nucleosome.

2. Methods

2.1. Model System

The REMD simulations were performed on 1) isolated tails, 2) the nucleosome and 3) tails binding to a DNA fragment with residue 16 outside the DNA groove (H4_OUT). We also prepared 1) and 2) for the H3 tail to verify the results. The H4 and H3 tails are defined as residue 1–26 and 1–44, accordingly. The nucleosome model was built based on the crystal structure containing eight histone proteins and 147 bp DNA (PDB ID: 1KX5) [14]. In addition, 10 bp linker DNA was added on each end so that the tails could fully bind to DNA [15]. The H4_OUT model was designed to investigate the influence of H4K16 acetylation, therefore, we prepared two systems containing the wildtype H4 tail (H4WT_OUT) as well as the K16 acetylated H4 tail (H4K16Ac_OUT), separately. On the basis of the converged structure from the nucleosome REMD simulation, residue 1–45 of H4 and 24 bp DNA around the H4 binding site was extracted. The residue 16, which was found in the DNA minor groove in the converged binding conformation, was pulled out by rotating the side-chain manually in Pymol [16] (Figure 4). The point charges for acetylated lysine were calculated following the protocol used in the development of the Amber force field [17]. We optimized the conformation by the B3LYP Density Functional Theory (DFT) with the 6-31G* basis, then calculated the electrostatic potential surfaces (EPS) by HF/6-31G*. The restrained electrostatic potential (RESP) [18] method was applied to fit the EPS to point charges. The bond parameters were introduced from the general Amber force field (GAFF) [19]. The generalized Born (GB) solvation model was used to simulate an implicit solvent [20]. We used modified Bondi van der Waals radii (mbondi2) from the Amber package for the intrinsic radii. The surface tension was set to 2.25936 kJ/mol/nm^2.

2.2. REMD Simulation

The REMD simulation implemented in GROMACS 4.5.7 [21, 22] was employed to support efficient sampling of the conformation space. In the REMD simulation, multiple replicas are run parallel in different temperatures, and exchanges between replicas are attempted in a given time interval to overcome high energy

barriers for a conformation transition. In this study, the temperature ranged from 300 K to 518 K, and the exchange probability was around 0.2. The step size for simulation was 2 fs and the exchange attempt was every 100 steps (0.2 ps). 8–12 replicas were prepared according to the atom number in the H3 or H4 tail. The initial system was energy minimized by the conjugate gradient method. The number of replicas in REMD largely depends on the degree of freedom in the system, so we defined a freeze group in a big system like the whole nucleosome and H4_OUT. It was only the tails under investigation that were flexible, while the position of other atoms was fixed. The freedom of the tails was constrained since the core region was very stable in the simulation time scale. Moreover, the bonds between heavy atoms and hydrogen atoms were constrained by the LINCS algorithm [23]. The non-bond interactions and GB pairwise summation were set to 1.5 nm. Three different force fields were employed for the simulation of isolated tails: Amber99SB* [24], Amber99SB-ILDN [25] and Amber03 [26]. The simulation time was dependent on the convergence of the system: 160 ns for the isolated H3 tail, 120 ns for the isolated H4 tail and the nucleosome, 100 ns for H4_OUT.

3. Data Analysis

The 300K trajectory was used for data analysis. The definition of the secondary structure is based on the DSSP algorithm [27]. For the H4_OUT simulation, we analyzed the occupancy of the conformations in which the residue 16 was "inside" (IN) or "outside" (OUT) the DNA groove, by measuring the distance (d) between the K16 NZ atom and the DNA T13 O2 atom (Figure 4). The "IN" conformation is defined as $d \leq 0.35$ nm, "OUT" as $d \geq 0.90$ nm, and "ELSE" as in between.

4. Results and Discussion

4.1. The predicted secondary structure for isolated histone tails depends on the force field

Although there are several studies on the secondary structure of isolated histone tails, the inconsistence of the results when using different simulation protocols is rarely discussed. In the present study, we performed the REMD simulation with three different force fields for the isolated histone tails and compared the secondary structures. Figure 2 shows the residues in α-helix and β-sheet structures as a function of time. In the Amber03 force field, the H4 tail converged to an α-helix with 4–10 residues, which agrees with the result in [8]. However, the stable β-sheet presented in [9] when simulated in the Amber99SB force field, only showed up temporarily and switched to a short α-helix content in this study. In Amber99SB-ILDN, the secondary structure of the H4 tail was unstable and showed very little β-sheet and even less α-helix. In summary, the secondary structure prediction for the H4 tail depends largely on the force fields used in the REMD simulation. Does this dependency also exist for other histone tails? We performed the same simulation protocol on an isolated H3 tail.

As illustrated in Figure 2, the H3 tail shows a high content of α-helix in all three force fields as concluded before [9]. But still, the 10–15 residue α-helix was very stable in Amber03, whereas in Amber99SB and Amber99SB-ILDN this long α-helix disappeared after certain simulation time, showing the force field dependence. The preference of specific secondary structures in different force fields has been discussed. For example, the Amber03 force field tends to overstabilize secondary structures, whereas Amber99SB-ILDN tends to understabilize them [28]. Therefore, it is risky to conclude from a simulation with one force field which specific secondary structure is adopted by histone tails in nature. On the other hand, the "converged" structure observed from the REMD simulation also depends on the simulation time. For example, after simulating the H4 tail with Amber99SB for 80 ns, we would probably have concluded a β-sheet conformation. In general, the longer H3 tail preferred to form an α-helix, whereas the H4 tail had a more widely distributed conformation space.

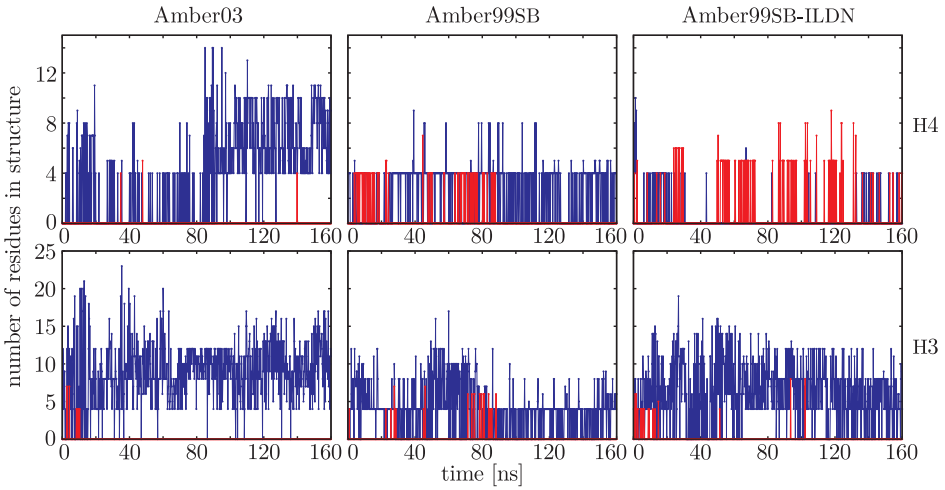

Figure 2. Residues of isolated H3 (bottom) and H4 (top) tail involved in secondary structure as a function of time; during the REMD simulation with three different force fields (Amber03, Amber99SB and Amber99SB-ILDN), the content of α-helix (blue) and β-sheet (red) shows force field dependence

4.2. Histone tails bind to nucleosomal DNA in disordered conformation

In the context of the nucleosome, the positively charged histone tails interact with negatively charged DNA as well as with the structured histone core. This interaction also influences the conformation of the tails. Whether histone tails bind in a structured, as proposed in [12], or disordered conformation to DNA remains to be answered. Figure 3 shows the secondary structure content of H3 and H4 tails in simulations of the whole nucleosome using the Amber99SB force field. Compared to the isolated tails (Figure 2), the number of residues

involved in α-helix and β-sheet is significantly reduced in the context of the nucleosome, which suggests that histone tails bind to DNA in a disordered conformation. This may be a consequence of the fact that the positively charged residues on the tail favor electrostatic interactions with negatively charged DNA, instead of hydrogen-bonds inside the tail which would enable the secondary structure formation. This result has, however, to be taken with some caution, since we investigated the tail conformation only in the context of intra-nucleosome interactions within a mononucleosome. Much more extended simulations on di- or larger oligonucleosomes would be required to take into inter-nucleosome interactions account.

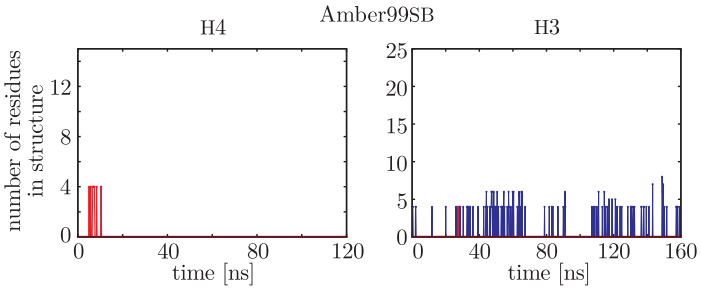

Figure 3. Residues of H3 and H4 tails involved in secondary structure as a function of time, in context of nucleosome; when binding to nucleosomal DNA, the content of α-helix (blue) and β-sheet (red) in histone tails reduced significantly

4.3. H4K16 inserts into DNA minor groove but acetylated H4K16 binds to the DNA surface

During the REMD simulation of the full nucleosome, the side-chain of H4K16 was found to insert into the DNA minor groove, while other lysines and arginines bound to the DNA surface. K16 gave up the strong electrostatic interaction with the phosphate groups, preferring hydrogen-bonding or cation-pi interaction in the minor groove. The obvious question is now how K16 acetylation influences the position of K16 and the binding conformation of the H4 tail. To study this, we performed 100 ns REMD simulations starting from models with the K16 residue outside the groove. By comparing the number of frames in which the K16 was inside or outside the groove (Figure 4), we found that the K16 residue bound back to the DNA groove very quickly, whereas the acetylated K16 preferred to stay on the surface of DNA, as suggested from an umbrella sampling study [13]. When the positive charge is canceled by acetylation, the steric hindrance caused by the acetyl group forces the bulky side-chain to stay outside the DNA groove.

In summary, a strong force field dependence was found for the secondary structure of isolated histone tails. In the context of the nucleosome, the H4 tail binds to nucleosomal DNA in a disordered conformation with K16 inserted into the DNA minor groove. H4K16 acetylation altered the binding conformation such that the acetylated K16 binds to the surface of DNA. Since we have discussed

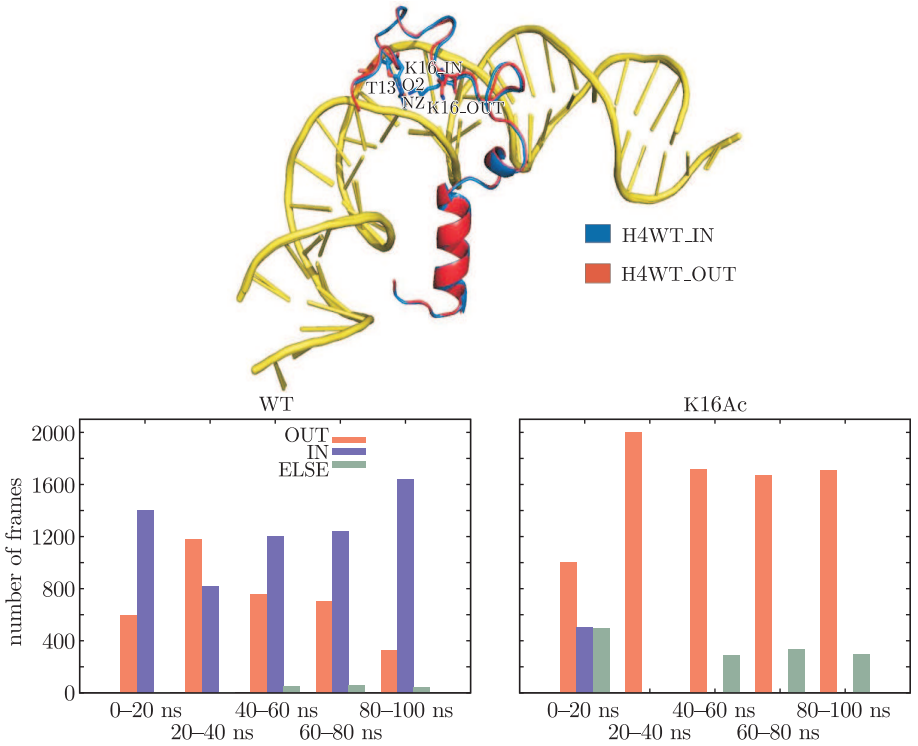

Figure 4. Models of H4 tail binding to DNA fragment with K16 residue inside (H4WT_IN) or outside (H4WT_OUT) the DNA minor groove; the "IN" is defined as the distance between the H4K16 NZ atom and DNA T13 O2 atom shorter than 0.35 nm, "OUT" as larger than 0.90 nm, "ELSE" as in between; the number of frames belong to these three categories indicated that K16 inserted into the groove, whereas the acetylated K16 preferred to stay on the surface

here only the single nucleosome model, the influence of H4K16 acetylation on internucleosomal interactions, on the global chromatin structure, and on the implications for these structural changes on gene activation remains to be explored. An in vitro study suggested that H4K16 acetylation disrupted the interaction between the H4 tail and the H2A/H2B acidic patch of adjacent nucleosomes, thus weakening the nucleosome-nucleosome stacking [29]. However, a recent study in mammalian cells has not found any influence of H4K16 acetylation on the high-order chromatin compaction [30]. Here we provide preliminary results that the H4 acetylation changes the intra-nucleosome binding conformation of the H4 tail. Further work is required to explore the interaction between adjacent nucleosome and the molecular mechanism of gene activation.

References

[1] Luger K, Mader A W, Richmond R K, Sargent D F, Richmond T J 1997 *Nature* **389** 251
[2] Biswas M, Voltz K, Smith J C, Langowski J 2011 *PLoS Comp Biol* **7**
[3] Kouzarides T 2007 *Cell* **128** 693
[4] Luger K, Richmond T J 1998 *Curr Opin Genet Dev* **8** 140
[5] Shogren-Knaak M, Ishii H, Sun J M, Pazin M J, Davie J R, *et al.* 2006 *Science* **311** 844

[6] Gansen A, Toth K, Schwarz N, Langowski J 2009 *J Phys Chem B* **113** 2604
[7] Robinson P J J, An W, Routh A, Martino F, Chapman L, *et al.* 2008 *J Mol Biol* **381** 816
[8] Yang D R, Arya G 2011 *PCCP* **13** 2911
[9] Potoyan D A, Papoian G A 2011 *J Am Chem Soc* **133** 7405
[10] Baneres J L, Martin A, Parello J 1997 *J Mol Biol* **273** 503
[11] Zhou B R, Feng H Q, Ghirlando R, Kato H, Gruschus J, *et al.* 2012 *J Mol Biol* **421** 30
[12] Wang X Y, Moore S C, Laszckzak M, Ausio J 2000 *J Biol Chem* **275** 35013
[13] Potoyan D A, Papoian G A 2012 *P Natl Acad Sci USA* **109** 17857
[14] Davey C A, Sargent D F, Luger K, Maeder A W, Richmond T J 2002 *J Mol Biol* **319** 1097
[15] Voltz K, Trylska J, Calimet N, Smith J C, Langowski J 2012 *Biophys J* **102** 849
[16] Schrodinger L L C 2010 *The PyMOL Molecular Graphics System, Version 1.3r1*
[17] Cornell W D, Cieplak P, Bayly C I, Gould I R, Merz K M, *et al.* 1996 *J Am Chem Soc* **118** 2309
[18] Bayly C I, Cieplak P, Cornell W D, Kollman P A 1993 *J Phys Chem* **97** 10269
[19] Wang J M, Wolf R M, Caldwell J W, Kollman P A, Case D A 2004 *J Comput Chem* **25** 1157
[20] Onufriev A, Bashford D, Case D A 2004 *Proteins-Structure Function and Bioinformatics* **55** 383
[21] Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, *et al.* 2013 *Bioinformatics* **29** 845
[22] Hess B, Kutzner C, van der Spoel D, Lindahl E 2008 *J Chem Theory Comput* **4** 435
[23] Hess B, Bekker H, Berendsen H J C, Fraaije J 1997 *J Comput Chem* **18** 1463
[24] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, *et al.* 2006 *Proteins-Structure Function and Bioinformatics* **65** 712
[25] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J L, *et al.* 2010 *Proteins-Structure Function and Bioinformatics* **78** 1950
[26] Duan Y, Wu C, Chowdhury S, Lee M C, Xiong G M, *et al.* 2003 *J Comput Chem* **24** 1999
[27] Kabsch W, Sander C 1983 *Biopolymers* **22** 2577
[28] Cino E A, Choy W Y, Karttunen M 2012 *J Chem Theory Comput* **8** 2725
[29] Allahverdi A, Yang R L, Korolev N, Fan Y P, Davey C A, *et al.* 2011 *Nucleic Acids Res* **39** 1680
[30] Taylor G C A, Eskeland R, Hekimoglu-Balkan B, Pradeepa M M, Bickmore W A 2013 *Genome Res* **23** 2053