Histone H3 and H4 N-Terminal Tails in Nucleosome Arrays at Cellular Concentrations Probed by Magic Angle Spinning NMR Spectroscopy

Min Gao,‡ Philippe S. Nadaud,‡ Morgan W. Bernier,‡ Justin A. North,‡ P. Chris Hammel,‡ Michael G. Poirier,*‡ and Christopher P. Jaroniec*†

‡Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States
§Department of Physics, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Chromatin is a supramolecular assembly of DNA and histone proteins, organized into nucleosome repeat units. The dynamics of chromatin organization regulates DNA accessibility to eukaryotic transcription and DNA repair complexes. Yet, the structural and dynamic properties of chromatin at high concentrations characteristic of the cellular environment (>~200 mg/mL) are largely unexplored at the molecular level. Here, we apply MAS NMR to directly probe the dynamic histone protein regions in 13C,15N-enriched recombinant nucleosome arrays at cellular chromatin concentrations and conditions designed to emulate distinct states of DNA condensation, with focus on the flexible H3 and H4 N-terminal tails which mediate chromatin compaction. 2D 1H−13C and 1H−15N spectra reveal numerous correlations for H3 and H4 backbone and side-chain atoms, enabling identification of specific residues making up the dynamically disordered N-terminal tail domains. Remarkably, we find that both the H3 and H4 N-terminal tails are overall dynamic even in a highly condensed state. This significant conformational flexibility of the histone tails suggests that they remain available for protein binding in compact chromatin states to enable regulation of heterochromatin. Furthermore, our study provides a foundation for quantitative structural and dynamic investigations of chromatin at physiological concentrations.

Eukaryotic DNA is dynamically organized into chromatin fibers, which regulate essential functions of the genome including transcription and DNA repair.1,2 The basic building block of chromatin is the nucleosome core particle, which contains ~146 base pairs (bp) of DNA wrapped 1.65 times around a histone protein octamer containing two copies each of histones H2A, H2B, H3, and H4.3 The nucleosome X-ray structure has been determined to near atomic resolution4 and reveals a compact helical core with ~15−30% of the histone sequences protruding from the core as largely unstructured, and presumably flexible, N-terminal tail domains (Figure 1A). The crystal structure of a tetranucleosome has also been solved,5 but the low (9 Å) resolution of this structure precludes the definition of N-terminal histone tail conformation.

In human cells, single chromatin fibers are chains of ~100 000 nucleosomes located in the nucleus at extremely high concentrations of >200 mg/mL,6 with each fiber organized into distinct chromosome territories.7 In vitro, longer (>~10−12-mer) nucleosome arrays form a variety of higher order structures in the presence of Mg2+, ranging from an extended beads-on-a-string type "10-nm" fiber in the absence of Mg2+ to a folded "30-nm" fiber at an intermediate (~1 mM) Mg2+ concentration to highly condensed aggregates at high Mg2+ concentrations.8,9 The 30-nm chromatin fiber has been observed in a few distinct cell types.10,11 Interestingly, however, recent small-angle X-ray scattering studies indicate that the 30-nm fiber is not the dominant structural form of chromatin in mitotic chromosomes,12 suggesting that the high cellular concentrations of chromatin may impact its higher order structure.

It is well-established that the positively charged N-terminal tails of histones H3 and H4 mediate the compaction of chromatin into 30 nm diameter fibers and interfiber condensation in vitro, apparently through interactions with DNA and/or acidic regions on the histone octamer surface of neighboring nucleosomes.13−15 Remarkably, peptides with sequences corresponding to the H4 N-terminal tail are also able to mediate self-association of H4 tail-less nucleosome arrays,16 suggesting that charge neutralization plays an important role in chromatin compaction. The mobile N-terminal histone tails have been previously probed by solution nuclear magnetic resonance (NMR) in ~20 mg/mL samples of single nucleosomes17−19 and oligonucleosomes20 in the absence of Mg2+. Of particular note is the recent elegant study of mononucleosomes containing 13C,15N-enriched histones by Bai et al.,19 which found that flexible H3 and H4 domains encompass residues 1−36 and 1−15, respectively, based on sets of backbone amide signals detected in 15N−1H heteronuclear single quantum coherence spectra. On the other hand, the structure and dynamics of H3 and H4 tails in compacted chromatin fibers remain largely unknown, with the analysis of these domains to date being limited to modeling17−23 and relatively indirect biochemical and biophysical approaches such as mutagenesis combined with analytical centrifugation or chemical cross-linking14,24 and hydrogen−deuterium (H/D) exchange coupled with solution NMR.25 Interestingly, the recent H/D exchange NMR study of 12-mer nucleosome arrays25 concluded that the H3 tail forms stable folded structures in highly condensed chromatin fibers. This is in contrast with the observation of a dynamically disordered H3 N-terminus in soluble single
nucleosomes\(^9\) and suggests possible differences in the conformational flexibilities of histone N-terminal domains between uncompacted and compacted chromatin.

Here, we apply 1D and 2D magic-angle spinning (MAS) NMR to probe histones H3 and H4 within isotopically \(^{13}\text{C},^{15}\text{N}\)-enriched recombinant 17-mer nucleosome arrays at very high densities of ~200–400 mg/mL of the cellular environment. The arrays were incubated with 0, 1, or 5 mM Mg\(^{2+}\), conditions that respectively yield extended chromatin, compact 30 nm chromatin fibers, and a highly condensed chromatin state with self-association between multiple fibers.\(^9\) The major advantage of the MAS NMR methodology in the present context is that it enables the amino acid residues comprising the rigid and flexible histone protein segments to be monitored directly under physiologically relevant conditions, in a manner similar to that previously demonstrated for a variety of large macromolecular complexes including membrane and fibrillar protein assemblies.\(^{16–28}\)

The nucleosome arrays used in this study were reconstituted from a 3046 bp DNA template containing 17 tandem repeats of a Widom 601 nucleosome positioning sequence variant with 30 bp of linker DNA\(^{29}\) (Figure S1) and histone octamer containing \(^{13}\text{C},^{15}\text{N}\)-labeled H3 or H4 (see Supporting Information (SI) for details). Sucrose gradient purified 17-mer arrays were analyzed by composite gel electrophoresis, BamHI and AvaI restriction enzyme digestions, and atomic force microscopy (AFM) (Figure 1B) to assess their level of saturation with a histone octamer. These analyses (Figures S2–S3 and accompanying SI text) show that the array samples are highly homogeneous, effectively (>95%) saturated with 17 histone octamers, and stable during the course of the MAS NMR experiments.

As noted above, different levels of chromatin condensation were induced by increasing amounts of Mg\(^{2+}\).\(^{2}\) The extent of self-association of the different array samples was readily evident in their physical appearance and was quantitatively confirmed by using a UV absorbance based sedimentation assay\(^{30}\) (Figure S4). For the NMR measurements the array samples were concentrated to ~200–400 mg/mL by ultracentrifugation, which is typical of chromatin densities found in cells.

The presence of flexible and rigid segments of histones H3 and H4 in the context of the highly concentrated 17-mer nucleosome arrays was first assessed using a set of 1D \(^{13}\text{C}\)-refocused INEPT and cross-polarization MAS NMR spectra recorded as a function of temperature between −20 and 30 °C and described in the SI (Figures S5–S7). The key result from this set of experiments is that, at temperatures of ca. −10 °C and above, both H3 and H4 contain highly mobile domains spanning ~25% of the protein sequences, irrespective of the degree of array compaction. Next, we performed a more detailed analysis of the flexible H3 and H4 domains by 2D MAS NMR methods. Figure 1 shows a series of \(^{1}\text{H}−^{13}\text{C}\) and \(^{1}\text{H}−^{15}\text{N}\) correlation spectra recorded at 30 °C using a 2D refocused INEPT pulse scheme\(^{26,27}\) as a function of

\[ \text{H3: } \text{ARTQRATARKSTGGKAPKQKLATKAARSKAPATGTVKPRTRPG} \text{VALRE}----\text{IRGERA} \]

\[ \text{H4: } \text{SGRKGGKGIQGKGAAPHRKVRLR} \text{N1QG}----\text{YFGG} \]
Increasing Mg$^{2+}$ concentration for arrays containing $^{13}$C,$^{15}$N–H3 (panels D and E) or H4 (panels F and G). As expected based on the 1D INEPT $^{13}$C spectra, the 2D $^1$H–$^{13}$C and $^1$H–$^{15}$N data sets display multiple intense resonances for both histones H3 and H4. A qualitative inspection of the $^1$H–$^{15}$N spectra reveals that all detectable backbone amide protons resonate in a narrow range between 8.0 and 8.5 ppm, characteristic of unstructured proteins. Indeed, by assuming that the flexible H3 and H4 tails are located at the N-termini and behave as ensembles of random-coil-like states we have been able to account for the vast majority of the observed protein signals by mapping onto the spectra the average residue- and site-specific chemical shifts obtained from the BioMagResBank database. While the spectral sensitivity and resolution achievable for these array samples were not sufficient to establish sequence specific resonance assignments de novo, analysis of the 2D NMR spectra in conjunction with the H3 and H4 amino acid sequences nevertheless permitted a relatively precise identification of the residues encompassing the flexible H3 and H4 N-terminal domains in the 17-mer arrays as follows. For both histones, resonances characteristic of the various amino acids were accounted for starting from the N-terminus, until signals from one or more amino acids in the sequence that should be clearly discernible in the spectra if sufficiently mobile were no longer detectable. It is important to note that while identification of amino acid types was fairly straightforward, the number of residues contributing to the individual resonances could not be readily determined due to the complicating effect of local protein dynamics on signal intensities.

For H3 a number of unique resonances were observed corresponding to residues in the range 1–38, including for example Ga, Af, Sk, Tβ, Vγ, Lo, and Ke signals in the $^1$H–$^{13}$C spectra as well as Qβ, Re, and Ry signals in the $^1$H–$^{15}$N spectra. Particularly remarkable is the observation of the Vγ signals. Combined with the fact that V35 is the first valine encountered in the sequence, this finding indicates that the flexible H3 tail extends to at least position 35. Equally crucial is the fact that no side-chain $^1$H–$^{13}$C signals characteristic of His or Tyr were observed, which indicates that H39 and Y41 are relatively immobile and defines the H3 tail boundary. In the absence of sequential resonance assignments it is impossible to unambiguously establish from these data whether residues K36–P38 are sufficiently mobile to be detectable, because multiple other residues of the same type that yield observable signals are located between positions 1 and 35. It is also unlikely that any of the H3 signals originate from protein regions other than the flexible N-terminal tail—in particular, it is improbable that the Vγ signal could be attributed to any of the other five valines located in the core domain of H3. The reason is that, in addition to the absence of His and Tyr resonances, we also do not observe any characteristic Ile side-chain signals (there are 7 isoleucines in H3 located between positions S1 and 130). Altogether, we conclude that flexible N-terminal H3 tails in concentrated 17-mer nucleosome arrays span amino acids 1–35 and possibly several additional residues in the 36–38 regime (Figure 1C), while the remaining H3 residues (certainly 39–130) are largely immobile.

Analogous analysis of the H4 spectra identified unique signals arising from multiple residues in the range 1–21, including for example Ga, Af, Vγ, Ho, He, and Ke signals in the $^1$H–$^{13}$C spectra as well as Re and Ry signals in the $^1$H–$^{15}$N spectra. Key is the observation of the Af, Vγ, Ho, and He signals because A15, H18, and V21 are the first Ala, His, and Val residues in the H4 sequence. Also critical is the finding that characteristic side-chain signals from Asp, Asn, Ile, Gln, and Thr, located in the range 24–30, are missing, which clearly defines the flexible tail boundary. Given that no Tyr and Phe resonances are detected either (there are two Phe and four Tyr residues in H4, with Y98 and F100 located near the C-terminus), we conclude that flexible N-terminal H4 tails in the 17-mer arrays span residues 1–21 while most of the remaining residues are relatively rigid (Figure 1C) (note: H4 residues L22 and R23 may also be somewhat flexible, but this cannot be unambiguously established due to the other leucines and arginines in the 1–21 segment).

In addition to the identification of specific amino acids making up the flexible N-terminal domains of histones H3 and H4 in concentrated chromatin, a central finding of our study is that by and large the same residues remain dynamic in uncompacted, folded, and highly condensed nucleosome arrays as judged by the similarity of the 2D $^1$H–$^{13}$C and $^1$H–$^{15}$N NMR spectra in Figure 1. To compare the extent of protein dynamics between the different array samples, we performed $^{15}$N INEPT based NMR experiments to monitor the transverse relaxation rates of the amide protons exclusively for residues encompassing the flexible histone tails (Figure S8). Although rather qualitative, these studies reveal that collectively the amide $^1$H coherences relax somewhat more rapidly for arrays incubated with 1 mM and 5 mM Mg$^{2+}$ relative to those prepared in the absence of Mg$^{2+}$, particularly for H3. This finding is suggestive of a subtle overall reduction in the flexibility of the histone tail domains in the most highly condensed and concentrated arrays, which would lead to less efficient motional averaging of the $^1$H–$^2$H dipolar couplings and consequently faster $^1$H coherence decay. We also recorded complementary electron paramagnetic resonance (EPR) spectra of the 17-mer arrays containing nitroxide spin labeled analogs of histone H3 at two locations including the N-terminus and residue 35 near the flexible tail boundary. The EPR spectra (Figure S9) display nearly identical, relatively narrow line shapes for all Mg$^{2+}$ concentrations indicating significant flexibility of the H3 N-terminal tail irrespective of the degree of compaction, and, importantly, no detectable broad components characteristic of immobilized protein segments. The moderate line broadening observed for position 35 relative to the N-terminus is consistent with the expected reduction in mobility of the nitroxide spin probe in the vicinity of the nucleosomal DNA.

In summary, we have demonstrated that MAS NMR spectroscopy permits the direct, unequivocal analysis of flexible histone tails in large nucleosome arrays at cellular concentrations. A strong correlation exists between the most dynamic histone residues identified in the current study, amino acids 1–35 in H3 and 1–21 in H4, and the crystal structure of the nucleosome core particle which reveals the same residues as being largely unstructured. Our data are also generally in agreement with the NMR study of single nucleosomes in solution, which reports that flexible H3 and H4 tails encompass residues 1–36 and 1–15, respectively. The finding that H4 residues 16–21 display significant mobility in the array samples could be indicative of somewhat different properties of these domains in the context of single nucleosomes at relatively low concentration versus highly concentrated arrays. Nevertheless, we also cannot discount the possibility that this difference stems from the fact that the solution NMR study focused solely on the protein backbone, while MAS NMR techniques monitor both the backbone and side-chain signals. Conversely, our results do not support the main conclusion of a recent H/D exchange NMR study that the H3 N-terminal tail forms stable folded structures in highly condensed nucleosome arrays. The origin of the discrepancy between the H/D exchange study and MAS NMR...
data is unclear. Yet, given that the former method does not visualize mobile residues directly but rather infers information about protein dynamics from amide proton occupancies, it is plausible that the histone tails in the highly condensed arrays could remain flexible yet concurrently somewhat protected from solvent exchange on the relatively short exchange time scales investigated. It is also noteworthy that the significant flexibility of histone tails in condensed nucleosome arrays revealed by MAS NMR is compatible with the EPR data for H3 presented in this study as well as the previous observation of mobile histone protein domains in oligonucleosomes by natural abundance $^{13}$C solution NMR at low and moderate ionic strengths, with the caveat that those experiments were done on relatively dilute nucleosome array samples.

The finding that H3 and H4 N-terminal tail domains are flexible even in highly condensed nucleosome arrays strongly suggests that chromatin compaction does not involve specific, high-affinity protein–protein or protein–DNA contacts that would lead to their immobilization. However, given the well-established importance of H3 and H4 histone tails in the folding of chromatin fibers, it is feasible that these flexible domains participate in the chromatin condensation process via multiple weak, transient interactions that shield the electrostatic repulsion between DNA moieties associated with different nucleosome units. The overall conformational flexibility of the H3 and H4 tails indicates that they are accessible to the numerous regulatory proteins that bind histone tails to regulate transcription even in compact heterochromatin at extremely high cellular densities. This provides a mechanism by which transcription activating complexes could directly gain access to histone tails to initiate conversions from heterochromatin to euchromatin. Factors in addition to histone tails that compact chromatin are also dynamic. Heterochromatin Protein 1 (HP1), which binds trimethylated lysine 9 within H3, and linker histone H1, which binds near the nucleosome DNA entry–exit region, facilitate heterochromatin formation. Both HP1 and H1 rapidly exchange in vivo, suggesting that multiple factors, which compact chromatin, function through dynamic mechanisms. Further studies of histone tails and the proteins that bind them are required to determine the role of histone tail dynamics in the conversion between heterochromatin and euchromatin.

**REFERENCES**