JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

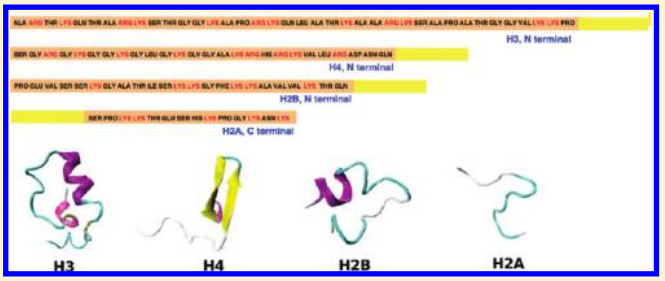
Energy Landscape Analyses of Disordered Histone Tails Reveal Special Organization of Their Conformational Dynamics

Davit A. Potoyan[†,‡] and Garegin A. Papoian*[,†,‡,§]

[†]Institute for Physical Science and Technology, [‡]Chemical Physics Program, [§]Department of Chemistry and Biochemistry University of Maryland, College Park, Maryland 20742, United States

Supporting Information

ABSTRACT: Histone tails are highly flexible N- or C-terminal protrusions of histone proteins which facilitate the compaction of DNA into dense superstructures known as chromatin. On a molecular scale histone tails are polyelectrolytes with high degree of conformational disorder which allows them to function as biomolecular "switches", regulating various genetic processes. Unfortunately, their intrinsically disordered nature creates obstacles for comprehensive experimental investigation of both the structural and dynamical aspects of histone tails, because of which their conformational behaviors are still not well understood. In this work we have carried out ∼3 microsecond long all atom replica exchange molecular dynamics (REMD) simulations for each of four histone tails, H4, H3, H2B, and H2A, and probed their intrinsic conformational preferences. Our subsequent free energy landscape analysis demonstrated that most tails are not fully disordered, but show distinct conformational organization, containing specific flickering secondary structural elements. In particular, H4 forms β-hairpins, H3 and H2B adopt α-helical elements, while H2A is fully disordered. We rationalized observed patterns of conformational dynamics of various histone tails using ideas from physics of polyelectrolytes and disordered systems. We also discovered an intriguing re-entrant contraction−expansion of the tails upon heating, which is caused by subtle interplay between ionic screening and chain entropy.



1. INTRODUCTION

All eukaryotic cells face the dilemma of tightly packaging their genomes inside a small nucleus, while also providing timely access to individual genes for transcription and replication, upon receiving internal and external signals. This difficult physical problem is elegantly solved by formation of a chromatin, a nucleoprotein complex consisting of DNA and histone proteins, densely packed together into regular repeating arrays. The fundamental structural unit of the chromatin is the nucleosome,[1−5] a ∼146 base-pair-long DNA segment tightly wrapped around histone octamer, comprised of four pairs of histone proteins: H4, H3, H2B, and H2A. Although the structure of the nucleosome has been resolved at a near atomic resolution,[6−8] and the nucleosome core particle was investigated using molecular simulations,[2,9] the structural information about higher-order polynucleosomal arrays is lacking. Nevertheless, it is clear that the histone proteins play a prominent role in determining chromatin structure and dynamics, where the latter, in turn, influences many cellular processes such as gene expression, silencing, and replication. In particular, histone terminal tails mediate internucleosomal attraction and control chromatin conformation through site-specific covalent modifications. The latter mechanism is the basis of the so-called histone code hypothesis,[1,10−14] according to which a specific combination of post-translational covalent modifications creates different biochemical responses by switching on or off various gene transcription and other signaling events. Despite their biological significance, molecular details of how histone tails carry out many of these tasks still remain insufficiently clarified, largely because of their intrinsic disorder. To elucidate molecular mechanisms of histone tail functioning, it is necessary to gain deeper understanding of their internal dynamics and conformational preferences. In the present study, we shed light on the nature of the natively disordered ensembles of various histone tails, focusing in particular on the role of transiently populated secondary structure elements. We also explore the role of mobile counterions in modulating histone tail conformational dynamics.

In addition to having pivotal roles in determining chromatin structure and dynamics, histone tails are also known for belonging to a special class of proteins that lack stable and densely packed 3D structure in vivo. This class of proteins that explores their unstructured nature to achieve functional promiscuity is known as intrinsically disordered proteins (IDP).[15−22] These proteins are distinguished by several features including low hydrophobicity, high net charge, and low sequence complexity, which have a combined effect of impeding the formation of densely packed globular structures. Recently, several experimental[23−25] and theoretical[26−30] studies on various ID proteins demonstrated that, despite the lack of major hydrophobic

Received: December 12, 2010
Published: April 25, 2011

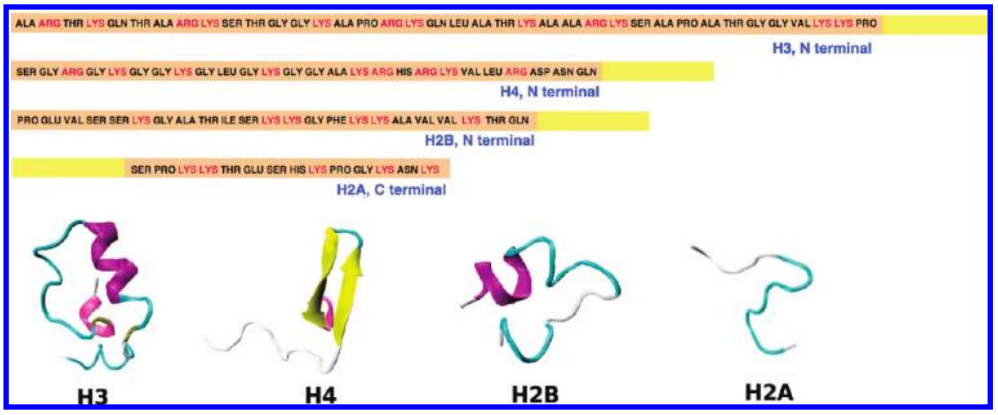


Figure 1. Representative conformations of histone tails are shown, obtained from simulations presented in this work. Sequences of histone tails are shown in the upper panel, where charged residues are indicated with red and neutral ones with black letters. The solid yellow bars indicate the remaining portions of the histone proteins that are not part of histone tails.

interactions, these proteins often do not show random coil statistics as one might have anticipated. Instead, most populated states may be rather compact, with a partial presence of local secondary order, fluctuating in the absence of strong stabilizing forces. The intrinsically disordered nature of histone tails became apparent from the X-ray structures of nucleosomes, where tail domains appear to sample multiple conformations.[6,7] This high conformational flexibility stems from the amino acid sequences which contain a high number of hydrophobic, charged (one-third of all residues) and structure-breaking (GLY) residues (Figure 1). For example ∼30% of residues in H3 and H2B tails are hydrophobic, whereas in H4 and H2A tails only ∼15% of residues are hydrophobic. Small globular proteins, on the other hand, are found much higher on the hydrophobicity scale, having sequences which are at least ∼50% hydrophobic, with some ∼30% of residues consisting of bulky hydrophobic groups (PHE, ILE, LEU).[31,32] Histone tail conformational disorder might accelerate binding-on rates through the fly-casting mechanism,[21,33,34] while their high net positive charges should enhance the binding affinity toward a negatively charged DNA surface. The accumulating wealth of experimental data showing specific binding propensities of histone tails eventually led to a suggestion that some tails may adopt specific secondary structures while bound to a linker DNA or acidic patches of core histones.[35−37]

Experiments on nucleosomes using circular dichroism (CD) and a combination of hydrogen exchange with NMR showed that H4/H3 tails acquire structured conformations as part of nucleosome core particles, while H2A and H2B were found to be essentially in random coil-like states.[37−39] The results of CD experiments suggested that some α-helical structure is present in isolated H4/H3 tails; however, due to the impossibility of selectively cleaving the H3 and H4 tails, distribution of α-helical amino acids among them had not been assigned. Hence, the possible interpretations allow either helical conformation for most of the H3 tail residues or equal distribution among residues of H4 and H3 tails.

There are only a few atomistic computational studies of histone tails. Of particular relevance to our work is the study of the wild and covalently modified forms of H3 histone tail by Yang, et al.[40] In the mentioned work implicit solvent replica exchange molecular dynamics (REMD) simulations were performed, finding that wild-type H3 tail populates α-helical conformations in qualitative agreement with our simulations. In another recent study, Arya et al.[41] found short α-helical elements in the conformations of the H4 histone tail, which is inconsistent with our results. We attribute the discrepancy to the fact that the authors used helically biased forcefields along with the generalized Born implicit solvent model, which has a combined effect of strongly favoring α-helices (see refs 42 and 43 and Methods in this paper). Many coarse-grained studies investigated the role of histone tails in chromatin folding by using low-resolution models of histone tails attached to uniformly charged spheres.[44−49] However, for these models to be realistic, should one treat histone tails as random coils, which is a common practice, or as chains with flickering secondary structure elements and potentially complex internal dynamics? The answer to this question should pave a way for large-scale computational modeling of chromatin dynamics. In the present work we are providing the first comprehensive overview of histone tail conformational preferences at high structural resolution.

We carried out all atom REMD simulation of all four histone tails with the aim to clear up the ambiguity related to their structural behavior, to find the driving force behind observed conformational preferences, and ultimately to devise a suitable qualitative framework for understanding histone tails that can also be utilized in lower-resolution studies. The main finding of our study is that three histone tails, H4, H3,and H2B, adopt persistent secondary structural elements and show behavior strongly deviating from random coil statistics. In particular, our results suggest that the H4 tail forms a β-strand in the well-known binding region, while H3 and H2B form α-helices, consistent with prior experimental findings. In contrast, H2A may be characterized as a random coil. Additionally, we used ideas from the energy landscapes theory[50−53] and polymer physics[30,54,55] to analyze the behavior of histone tails on the coarser scale. In particular, on the basis of this analysis we discovered an intriguing re-entrant contraction−expansion of histone tails upon heating, which is caused by a subtle competition between enhanced ionic condensation around charged side chains[2,56−60] and chain entropy.

2. METHODS

2.1. Simulation Protocol. Since there are no reliable structural data on histone tails, we have built initial structures based on the available amino acid sequences.[61] Histone tails have not been uniquely

defined in literature. For instance, histone tails are biochemically isolated by trypsination, e.g. cleaving at the so-called "weak points", dividing disordered and ordered regions. However, the cleavage point does not always coincide with the residues of histone tails that are adjacent to the exit point of nucleosomal DNA.[62] In this work we have constructed histone tails that are slightly longer than their biochemical definitions, thus following a structural viewpoint. The tail lengths are 38 residues for H3, 26 residues for H4, 23 residues for H2B, and 14 residues for H2A. All the simulations were carried out using the AMBER10[63] package suite and the ff99SB protein force field.[64] Finding an appropriate forcefield for simulating IDPs can be a challenging task for several reasons. For one thing, the parametrization of currently available forcefields is fine-tuned to reproduce dynamics of proteins with single, well-defined structures. Hence, a priori expectation that the default parametrization will work for IDPs is not high, and indeed extensive simulations on model unstructured peptides diagnosed serious shortcomings for many forcefields.[65,66] Luckily the comparison of explicit solvent simulations with NMR structural and relaxation data established adequacy of the ff99SB force field for IDPs.[65,67] In a recent report it was suggested that in specific cases ff99SB might slightly underestimate the helical propensity of polyalanine model peptides.[43] Therefore, to validate the robustness of our results, we have carried out additional simulations with modified ff99SB, finding that simulations with both force fields led to virtually identical results.

After constructing the initial structures for histone tails in a fully stretched state, we performed preliminary minimization and equilibration steps in the GBSA implicit solvent[68] to bend straight conformations to some degree in order to save computational resources associated with relaxing conformations in explicit solvent. Afterward, each histone tail was immersed in TIP3P explicit water boxes (H3: $71 \times 61 \times 56$ Å^3, H4: $56 \times 45 \times 44$ Å^3, H2B: $53 \times 44 \times 42$ Å^3, H2A: $48 \times 46 \times 45$ Å^3) with distances between the farthest atoms of histone tails and box edges set to ~12 Å. Particle Mesh Ewald summation technique was used for all electrostatic calculations with 12 Å real space cutoff. Periodic boundary conditions were used in all simulations.

Ions were added to neutralize uncompensated charges, and further salt (NaCl) was added to represent 0.150 M ionic concentration and thus mimic the physiological environment. System preparation included minimization of protein chains with the rest of the system fixed and subsequent minimization of the full system. After minimization steps, leapfrog integration with a 2-fs time step in the NVT ensemble was used to propagate dynamics of all atoms, where the latter were coupled to a Langeven bath with a 2-ps collision frequency and with weak restraining force on protein.[63] After 200 ps of restrained NVT simulation, the system was equilibrated without restraints in the NVT ensemble for 300 ps followed by 1.5 ns density equilibration in the NPT ensemble. SHAKE[69] was used to constrain all bonds containing hydrogen atoms.

Before running replica exchange simulations each system was replicated, and each replica was slowly heated and additionally equilibrated at the predefined target temperatures. Temperatures for REMD were chosen on the basis of the criteria of a good overlap between energy distributions of neighboring replicas, guarantying significant acceptance rates.[70−72] The temperature range was chosen between 300 and 450 K, with 2−3 K spacing resulting in 50−54 replicas and 30−35% expected acceptance rate estimated with the help of T-REMD server.[73] The time interval between the exchange attempts was set to 5−10 ps. Each replica was simulated at the constant temperature and volume (NVT) for 55−60 n, resulting in cumulative 3 μs of sampling time for each histone tail. The first 10−15 ns of all the trajectories were discarded, to allow for initial equilibration, while the rest of the trajectories were used for the subsequent analysis.

2.2. Principal Component Analysis (PCA). We performed a dihedral PCA (dPCA) on the single-temperature trajectories sampled by replica exchange molecular dynamics. Since we are interested in detecting states with residual order, using dihedral PCA (dPCA) is a natural choice, because dihedral angles are the main degrees of freedom responsible for backbone flexibility and formation of secondary structural elements. In previous studies dPCA was used to scrutinize native state dynamics of small globular proteins[74,75] and large-scale conformational rearrangements of hydrated proteins.[76−78] Commonly, a few PC modes are used to reduce dimensionality of conformational space and allow investigation of dynamics on simplified landscapes.[50] For instance, the main PC mode served as a good order parameter successfully discerning states which undergo major conformational change between open and closed forms of elastin.[77] In a different work, following splitting of basins in successively higher PC dimensions allowed mapping of the protein's conformational substates into a hierarchical tree.[74,75]

In dPCA, the covariance matrix $\mathbf{C}$, is constructed using sines and cosines of (ϕ,ψ) protein dihedral angles $\eta = \{\sin\phi, \sin\psi, \cos\phi, \cos\psi\}$, in order to avoid problems of discontinuity and multivalues associated with angular variables.[79,80] Each principal component is a basis vector in the high dimensional conformational space of the macromolecule, along which motions occur corresponding to the greatest variance in the data. After diagonalizing the covariance matrix (eqs 3.1 and 3.2), one obtains a set of orthogonal PCs and corresponding eigenvalues (diagonal elements of $\mathbf{\Lambda}$), where the former indicate how atoms are displaced in a particular mode from the time-averaged structure. Thus, most of the interesting dynamics is contained in the PCs with the largest eigenvalues; these capture most of the essential macromolecular motions. After finding the PCs, we used projections of the trajectory along two main PCs to map out the protein's free energy surface in these collective coordinates (eq 3.3).

$$\mathbf{CM} = \mathbf{M\Lambda}, \mathbf{C} = \langle \eta\eta^T \rangle \tag{3.1}$$

$$v_i = \mathbf{T}m_i, m_i m_j = \delta_{ij} \tag{3.2}$$

$$\Delta F(v_1, v_2) = -k_B T \log P(v_1, v_2) - F_{\min} \tag{3.3}$$

Where $\mathbf{T}$ is a trajectory matrix where each column contains sampled values of individual dihedral angles, m_i (the ith PC) is an ith column of eigenvector matrix ($\mathbf{M}$) corresponding to the ith eigenvalue (Λ_{ii}) and v_i is the projection of trajectory onto ith PC. We identified distinct basins on the constructed free energy landscape by enclosing them within squares and estimate their structural heterogeneity by computing the corresponding pairwise-q values (often used as an order parameter in spin glass physics[81]) among conformations within each square.

$$q_{AB} = \frac{1}{N_{\text{pairs}}} \sum_{i>j} \exp(-(r_{ij}^{A} - r_{ij}^{B})^2) \tag{3.4}$$

$$q = \langle q_{AB} \rangle_{\text{basin}} \tag{3.5}$$

where q_{AB} is the structural overlap between the A and B conformations, assuming values from 0 (no structural resemblance) to 1 (identical structures). In eq 3.4, N_{pairs} is the total number of C_α atom pairs and r_{ij}^{A} indicates the pairwise distance between C_α atoms in conformation A which are correspondingly labeled as i and j. Overall, a larger value of q in any specific basin indicates the potential presence of a deep trap state(s), since then conformations sampled within that basin will show high structural resemblance to each other.

2.3. Scaling Relations. To classify the compaction state of various histone tails, we used scaling relations for globular and thermally denatured proteins to compare our computed average radius of gyration values, R_g, with the typical values for native and random coil-like states of proteins having the same number of residues:

$$R_{g\ \text{glob}}(N) = 2.2N^{0.38} \tag{3.6}$$

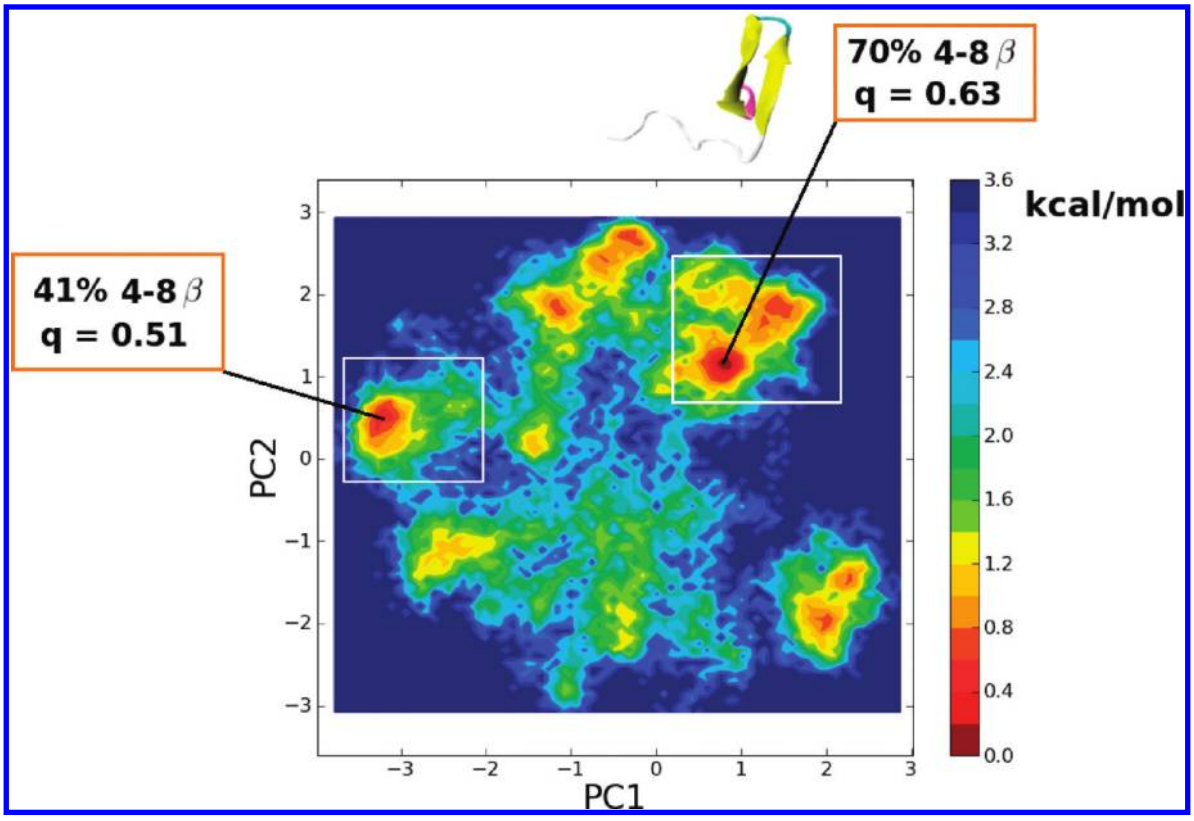


Figure 2. Free energy projection of the H4 tail dynamics at 300 K into its two main principal components is shown, based on eq 3.3.

$$R_{g\ denat}(N) = 2.02N^{0.60} \tag{3.7}$$

The relation for globular proteins is based on a best fit of power law dependence of R_g on sequence length for a subset of proteins in PDB.[82] Similarly, for denatured proteins we use the relation that was found by the best fitting of R_g values obtained through simulation and experiments.[83]

Furthermore, we compared the chain statistics of histone tails with the behavior of an ideal chain with excluded volume interactions, where we used des Cloizeaux equation[55] to estimate the probability density of end-to-end distance of a polymer chain. This equation was derived using renormalization group theory by incorporating fast decay at large distances and low probability of end contacts into the ideal chain probability density function.[55] In the following equation,

$$P(X) = CX^{0.269}\exp(-1.2X^{2.427}) \tag{3.8}$$

C is a constant determined by normalization condition ($\int 4\pi X^2 P(X)\, dX = 1$) and $X = R/\langle R\rangle$. It should be noted that des Cloizeaux equation is exact only in the asymptotic limit of long and uniform chains which have simple pairwise interactions that decay with sequence-wise separation.

3. RESULTS AND DISCUSSION

3.1. "Order in Disorder": Secondary Structure-Forming Propensities of Histone Tails at Physiological Conditions. To produce realistic conformational ensembles of histone tails we have carried out long-time REMD simulations in a wide temperature range (300−450 K), taking into account explicit solvent environment and ions. Afterward, we employed various physically motivated means to classify and catalog the sampled conformational space. The results outlined in this section clearly demonstrate that, despite possessing sequences atypical for native globular proteins, histone tails do show strong propensities for forming secondary structural elements, at specific local spots.

In our simulations we found residual secondary structural elements in three out of four histone tails. The H4 tail was enriched in β-hairpin conformations, the H3 and H2B tails had helical content, and the H2A tail showed no structural features throughout the whole simulation (see Figure 1 for representative snapshots). In the H3 chain, there are 2−3 regions that strongly favor formation of α-helices, implying that upon binding an extensive helix formation could follow. In the H4 tail, half of the chain in the C-terminal segment (Res 12−26) forms a β-hairpin, while the N-terminal segment (Res 1−12) remains fully disordered. The H2B histone tail showed weaker propensity to form secondary structures. To disentangle interesting conformational modes from pool of states sampled by single-temperature REMD trajectories we applied PCA in space of (ϕ,ψ) peptide backbone torsional angles (dPCA). Two-dimensional free energy surfaces for all four histone tails were mapped using first two PC modes corresponding to the largest eigenvalues, since these modes move on the slowest time scales.

Obtained two-dimensional free energy landscapes (Figures 2−4) demonstrate that the conformational spaces of three histone tails (H4, H3, and H2B) are well-defined by a handful of distinct basins. In each basin, we have computed the percentage of conformations that possess residual secondary structure, finding a noticeable variation in the degree of secondary structural content among different basins. To quantify the conformational heterogeneity

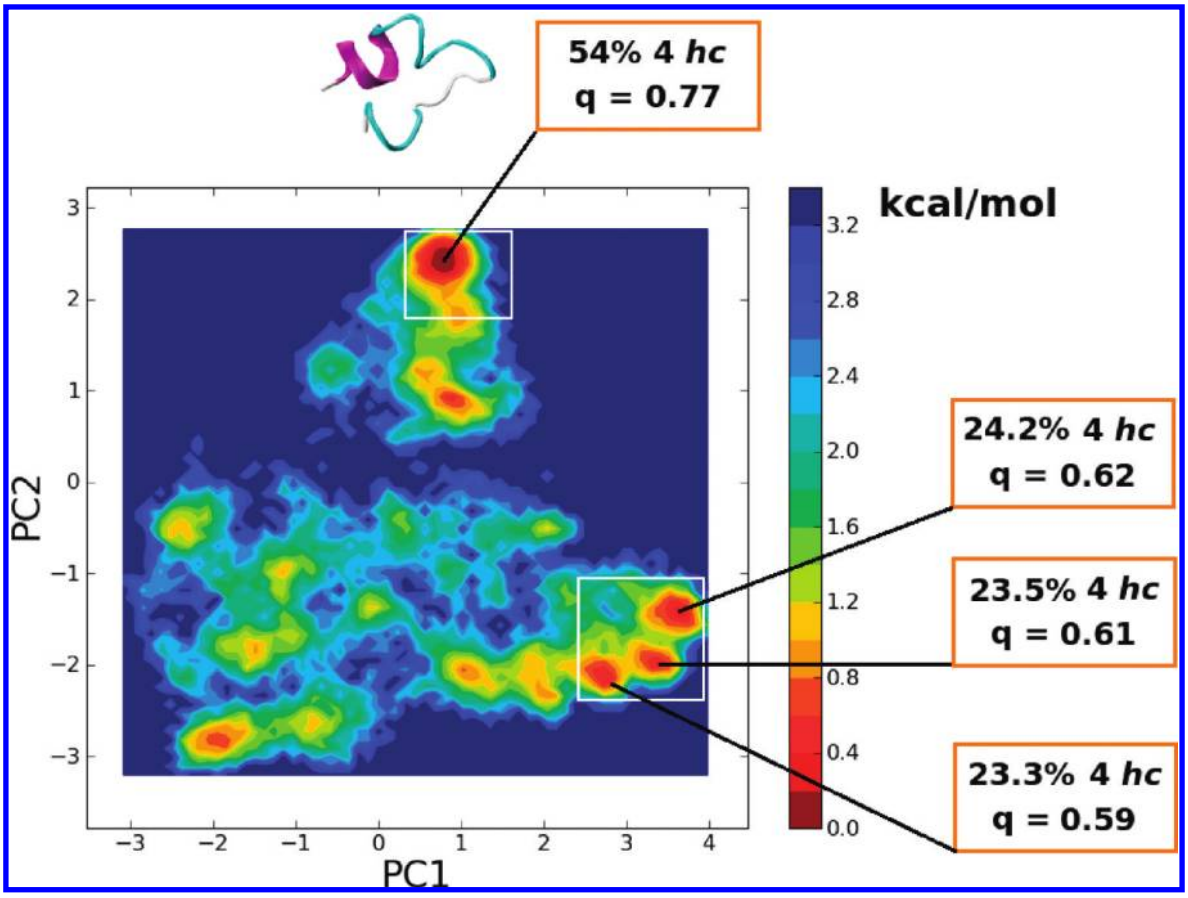


Figure 3. Free energy projection of the H2B tail dynamics at 300 K into its two main principal components is shown, based on eq 3.3.

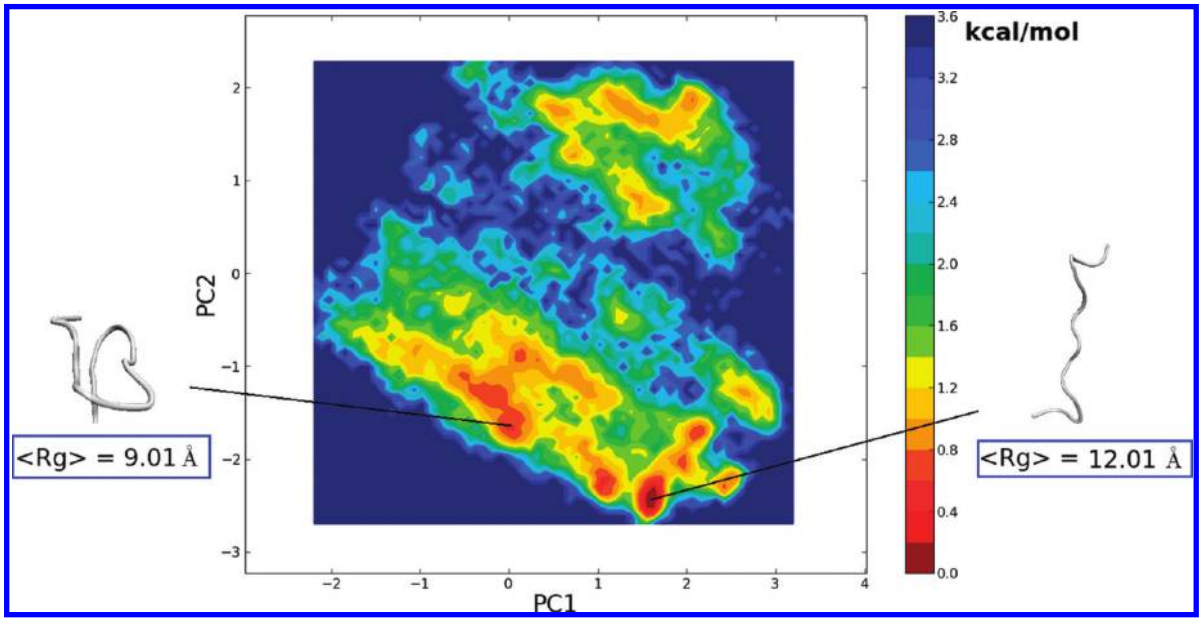


Figure 4. Free energy projection of the H2A tail dynamics at 300 K into its two main principal components is shown, based on eq 3.3.

inside basins, we have also computed the average mutual structural overlap q between all conformations within each basin. When a conformational trap(s) with deep free energy is present in a given basin, it tends to attract many visits during peptide dynamics, resulting in similarity of many conformational snapshots to each other and, as a result, high q values. Our subsequent analysis indicated that only a weak correlation exists between residual secondary structural content and a basin's q, which is due to the fact that

dx.doi.org/10.1021/ja1111964 |*J. Am. Chem. Soc.* 2011, 133, 7405–7415

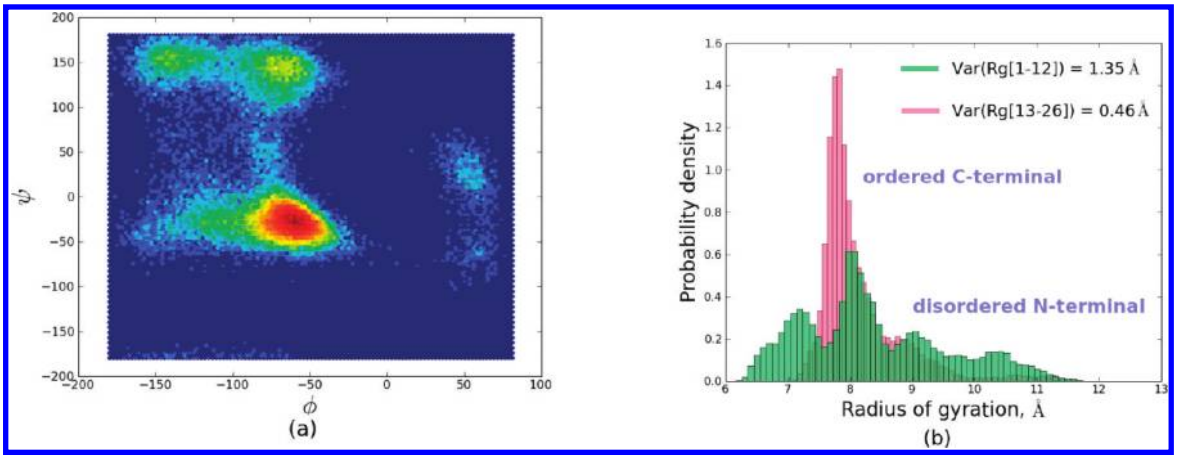


Figure 5. (a) Ramachandran plot of Lys-16 in the H4 tail is shown. (b) Distributions of radii of gyration for the N (residues 1−12) and C (residues 13−26) terminals of the H4 tail are plotted.

significant portions of all chains are in disordered conformations. Here we should also emphasize that the proximal basins on the PC free energy surface are kinetically accessible,[84] e.g. the states in neighboring basins are structurally closer to each other compared to states in more distant basins.

Landscape topographies of the H4, H3, and H2B tails show a signature of well-defined multiple basins of various depths, whereas the basins are more shallow for the H2A tail. We have indicated the thermodynamically dominant basins with white boxes in Figures 2 and 3. On average, the basin depth difference between dominant and other basins is $\sim 2-3k_BT$; hence, we expect that the dynamics will mostly be determined by moderate lingering in the dominant basins and frequent transitions among dominant and other basins. When examining the specific features of each tail's free energy landscape, we found partially structured conformations along with fully disordered ones in all basins of the H4, H3, and H2B tails, where in all cases the dominant basin had the highest content of partially structured states (see Figures 2−4). The structural overlap parameter, q, was also highest in these basins, indicating the presence of thermodynamically dominant, specific conformations. In particular, the landscapes of the H2B tail (Figure 3) and the H3 tail (Supporting Information) contain basins with varying content of α-helical segments, while the H4 tail basins were enriched in β-hairpins (Figure 2). On the other hand, the free energy landscape of the H2A tail (Figure 4) shows a more connected web of basins with the absence of conformations possessing any secondary structure, which is primarily attributed to the fact of the H2A tail having the shortest sequence with the highest charge per residue (Figure 1). The broad and connected basins on the landscape of the H2A tail (see Figure 4) correspond to random coil-like states, and the deepest basin consists of mostly stretched conformations with relative R_g values significantly exceeding R_g of other histone tails, as discussed below. Our present results are consistent with prior experiments[39] and secondary structure prediction results,[85] which support the view of the H2A tail being a random coil and the other tails possessing residual structure.

Next, we closely examined the distributions of (ϕ,ψ) dihedral angles for several key residues in various histone tails. In the H4 tail, for instance, we found an interesting structural feature that adds an extra layer of stability to β-hairpin state relative to its disordered conformation. There are two pairs of LYS-ARG residues in a β-turn that reduce the conformational flexibility of the turn, restricting it to a small number of states. The resultant states favor β-hairpin over the other conformations where, if the latter were realized, that would have inevitably led to a steric clash between positively charged side chains of LYS and ARG. On the contrary, in globular proteins GLY is the residue that frequently resides in the β-turns making them more flexible and thereby facilitating formation of compact secondary structures.

The LYS-16 residue, which in our simulations is frequently locked into a specific rotameric state in the β-turn, is a well-known hot spot for post-translational covalent modifications.[86,87] Hence, on the basis of the crucial role that LYS-16 plays in stabilizing H4 tail's β-hairpin as observed in the current work, we speculate that covalent modifications may potentially shift the H4 tail conformational equilibrium, providing an additional mechanism for signaling through post-translational modifications. Additionally, in vitro experiments have shown that homogeneous monoacetylation of the H4 tail undermines the stability of chromatin fiber to an extent that removal of the whole H4 tail is equal in its effect.[88] Our current results suggest a possible molecular level explanation, namely that acetylation of the H4 tail triggers partial or full disruption of the β-hairpin, leading to modulation of binding affinity toward linker DNA . Ramachandran plots (Supporting Information) of several other key H4 residues showed that LYS and ARG residues sample diverse conformational states in the disordered N-terminal region, while the analogous residues in the C-terminal region are highly constrained (see Figure 5a). This conformational dissimilarity of two halves of the H4 chain was further explored by computing radii of gyration of two terminal segments. Obtained distributions of R_g values demonstrated that the part with enhanced secondary structural content (C-terminal) on average has less variability in its size distribution (Figure 5b), thus supporting the global view of H4 tail as "half-ordered, half-disordered". In the H3 tail, there are three LYS-ARG regions which, as in the H4 tail, form β-turns aiding in overall compaction. However, contrary to the H4 tail, local sequences of the H3 tail around β-turns favor α-helices, producing an overall enrichment in α-helical transient states. In the H2A tail, all residues were unconstrained and spanned almost all allowed regions on the Ramachandran diagram.

Next, we explore the size distributions of histone tails. We computed distributions of radii of gyration for the conformations

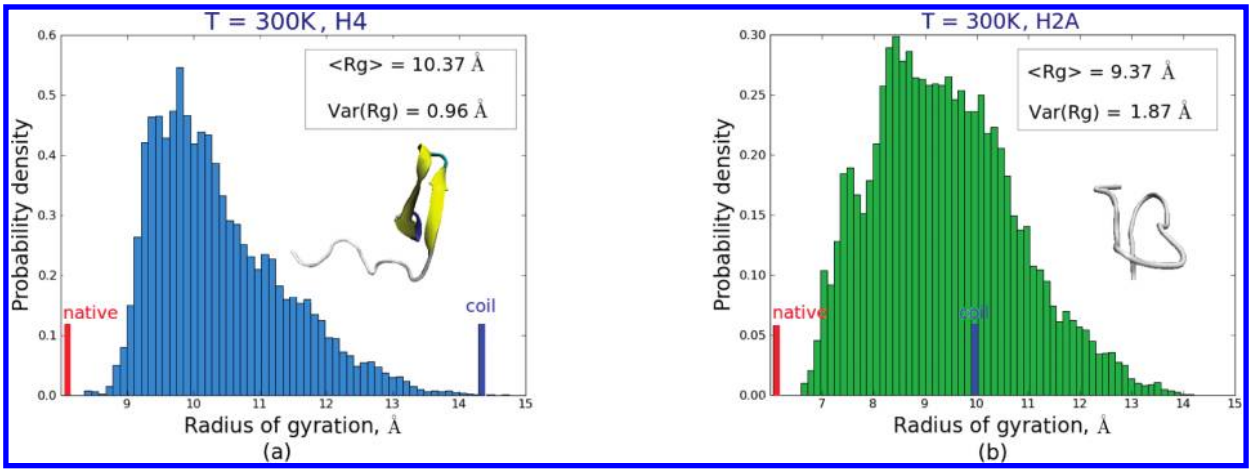


Figure 6. Distributions of radii of gyration for the (a) H4 and (b) H2A tails are shown at 300 K. Red and blue bars respectively indicate expected sizes for a globular folded protein and a random coil for a hypothetical peptide of the same length.

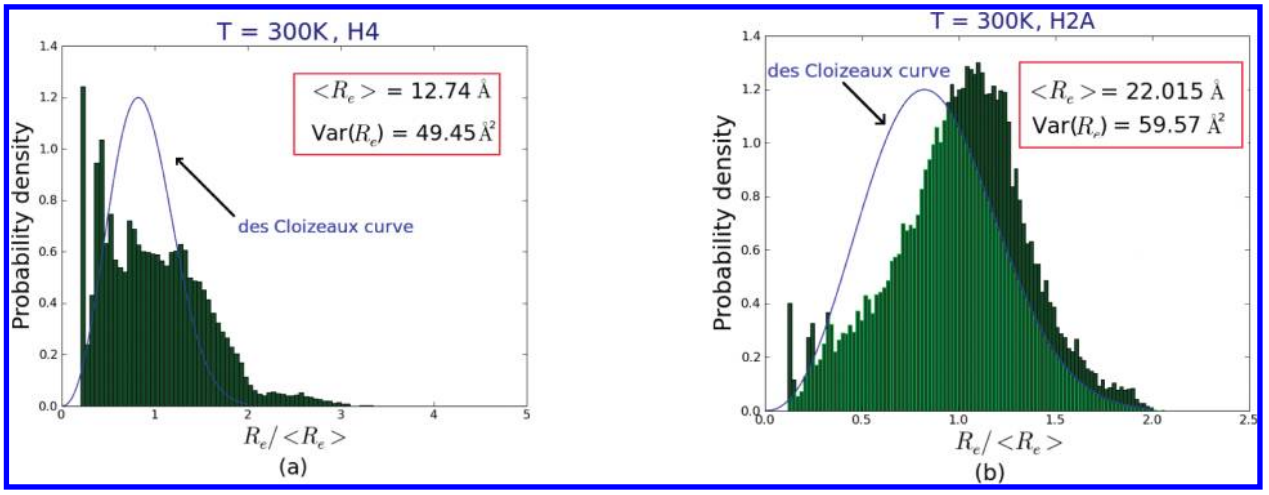


Figure 7. End-to-end distance histograms of the (a) H4 and (b) H2A tails are shown at T = 300 K and compared with the des Cloizeaux equation predictions.

sampled from the equilibrium ensembles of histone tails and used mean R_g values for comparison with well-investigated classes of globular and thermally denatured proteins (Figure 6). From this comparison one immediate finds that all histone tails (except H2A) on average have sizes that are closer to native globular states than random coils. In particular, despite their disordered nature, conformational ensembles of H4/H3/H2B histone tails significantly deviate from random coil statistics, and instead are better described as collections of relatively compact "molten globular" (MG) states. This is shown by the mean values of their radii of gyration, which for the H4 tail differs by ∼30% from the corresponding hypothetical native state (see eqs 3.6 and 3.7) of a globular protein with the same length ($\langle R_{g\ H4}\rangle$ = 10.4 Å, $R_{g\ glob}$ = 7.6 Å, $R_{g\ denat}$ = 14.3Å) The H3 tail's degree of compaction ($\langle R_{g\ H3}\rangle$ = 11.6 Å, $R_{g\ glob}$ = 8.8 Å, $R_{g\ denat}$ = 17.91 Å) is qualitatively similar to the H4 tail, although the spread of R_g distribution is much higher. The H2B tail, which is only two residues shorter than H4, showed a higher degree of compaction, with $\langle R_{g\ H2B}\rangle$ being only ∼20% higher than the corresponding native globular form ($\langle R_{g\ H2B}\rangle$ = 9.12 Å, $R_{g\ glob}$ = 7.36 Å, $R_{g\ denat}$ = 13.59Å) This enhanced compactness compared with the H4 tail may be rationalized by the lower ratio of H2B tail's net charge to its sequence length (H2B: 0.25, H4: 0.35) and also a higher ratio of hydrophobic structure-forming residues. For the H2A tail, the situation is qualitatively different as indicated by its R_g distribution (Figure 6b), which puts the H2A conformational ensemble closer to random coil-like states ($\langle R_{g\ H2A}\rangle$ = 9.37 Å, $R_{g\ glob}$ = 6.0 Å, $R_{g\ denat}$ = 9.86 Å).

From a polymer physics perspective, histone tails are polyelectrolytes with intricate conformational behavior stemming from propensity to form secondary structures, having nonuniform charge distribution and hydrogen-bonding patterns. Because of the presence of such interactions, one would expect a significant deviation from classical polymer theories that are mainly based on simple coarse-grained pairwise potentials between uniformly shaped monomers. Hence, we also computed the end-to-end distance distributions for all histone tails and compared them with the corresponding distributions obtained from Gaussian chain in the excluded volume limit as described by des Cloizeaux equation (eq 3.8). Indeed, from comparisons with des Cloizeaux curve (Figure 7), we see a drastic difference with the low-temperature end-to-end distance distributions of the H4,

dx.doi.org/10.1021/ja1111964 |J. Am. Chem. Soc. 2011, 133, 7405–7415

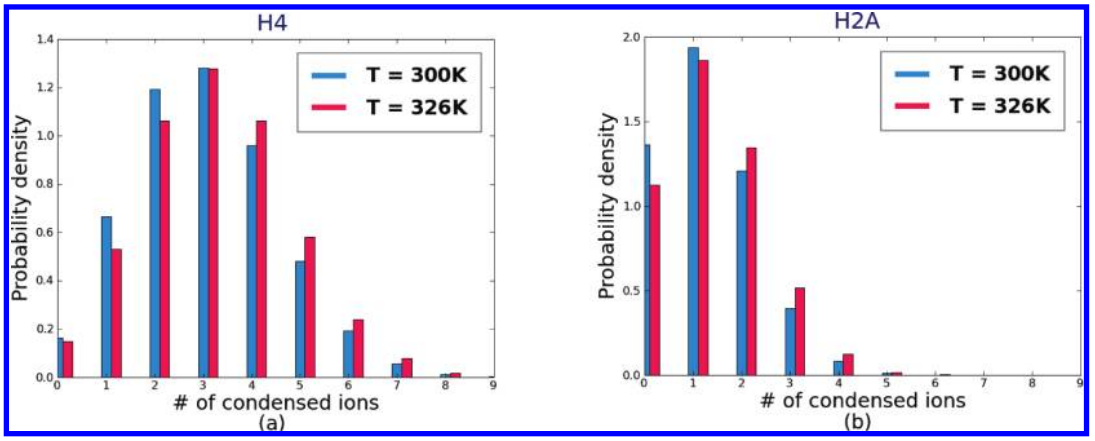


Figure 8. Temperature dependencies of ionic condensation around backbones of the (a) H4 and (b) H2A histone tails are shown at 300 and 326 K.

H3, and H2B tails (data for H3 and H2B not shown). The multipeak nature of these distributions is most likely caused by the presence of secondary structural elements, where the latter introduce additional complexity in monomer−monomer interactions that is not captured by des Cloizeaux equation. For instance, in the H4 chain there is a significant population of both states with on average close approach of monomers and states with higher degree of separation (Figure 7a). These results clearly point to the limitations of classical polymer physics ideas for characterizing even disordered protein chains, which in general demand inclusion of more structural details. On the contrary, the H2A tail is reasonably well approximated by des Cloizeaux expression (Figure 7b), indicating that it is a random coil. The peak of distribution is shifted relative to des Cloizeaux curve due to the presence of multiple charges. Overall, our results suggest that wormlike chain models are not expected to adequately capture conformational dynamics of histone tails.

3.2. Thermal Denaturation and Ionic Effects in Histone Tails. Histone tail sequences are unusual because they contain a plethora of positively charged residues. This aspect is not surprising since histone tails regulate chromatin structure and dynamics by interacting with negatively charged DNA and acidic patches of histones cores. Hence, electrostatics plays a key role in their functionality. In particular, the distribution of counterions around charged surfaces of histone tails is expected to play a crucial role in interactions of tails with DNA, through release and reassociation of ions around biomolecular surfaces, as well as influencing the equilibrium ensemble of histone tail conformations.[59,89] To gain further insight into the histone tail electrostatics, we investigated the nature of the ionic environment around protein backbones, including temperature dependencies. One remarkable aspect of polyelectrolytes is their ability to undergo rapid contraction into compact forms upon increase of the counterion concentration.[90−92] The main driving force for this transition is screening of the electrostatic repulsion among the charges of the chain which, in turn, diminishes swelling of the conformations that occurs due to these repulsive interactions. In our case, however, concentration of the salt is held constant, and the parameter whose change induces chain contraction is the temperature. This temperature-induced chain collapse has not been studied in prior theoretical works on flexible polyelectrolytes,[91−97] and here we give qualitative explanation about the cause and explain the specifics of the transition for different histone tails.

First, to characterize the low-resolution details of ionic association around histone tails, we computed the total number of anions (Cl^-) around a positively charged backbone using as a threshold a distance between nitrogens of ARG and LYS and chloride ions being less than or equal to the Bjerrum length (∼7.5 Å) at ambient temperature. Interestingly, upon heating we observed consistently increasing accumulation of ions around both histone tails (Figure 8). This type of increase of ionic condensation with temperature was already observed in a number of MD simulations of other proteins[56,59] and in some cases was shown to be sequence specific.[60] One way to understand the observed elevation of ionic adsorption at high temperatures is to take into account the temperature dependence of the dielectric constant. The experimentally found expression for the dielectric constant of water[98] is: $\varepsilon(T) = (T^*/T)^{1.4} \approx T^{-1.4}$; hence, the Bjerrum length ($l_B = (e^2)/(\varepsilon(T)k_BT)$) of an aqueous medium is effectively a weakly increasing function of temperature: $l_B \approx T^{0.4}$. Therefore, qualitatively, at high temperatures ionic interactions are stronger, which under favorable conditions can lead to enhanced ionic condensation.

Furthermore, we observed consistently greater accumulation of Cl^- ions around residues in disordered N-terminal regions compared to ordered C-terminal regions. This trend seems to arise because of larger steric accessibility of disordered regions for mobile ions. Accumulation of ions around chains increases the ionic screening, resulting in diminution of electrostatic repulsion among charged side-chain groups. This enhanced screening in turn explains why histone chains instead of expanding with moderate increase of temperature, somewhat contract. For the H4/H3/H2B tails, the effects of ionic condensation and subsequent chain contraction were more pronounced than for the H2A tail. This difference in ion adsorption at elevated temperatures is caused by the disruption of secondary structural elements in the H4/H3/H2B tails which ultimately leads to more surface-exposed conformations, with charged side chains having fuller access to ions floating around. Hence, at all temperatures above T = 300 K the H2A tail on average attracts less mobile ions per chain charge compared with the other tails.

The main principal modes (PC 1−5) at moderately high temperatures ($T \approx 320-326$ K) showed the emergence of new compact states, as evidenced by the appearance of additional deep minima (see Supporting Information), which vanish upon further heating. Additionally at temperatures corresponding to collapse states, we have found sharper decay of eigenvalues (e.g.,

Λ_{ii} from eq 3.1) when arranged in decreasing order (see Supporting Information). This interesting observation is explained by the idea that, in collapsed chains dihedral angles are fluctuating in a more restricted range of values and motions are relatively damped because of tighter monomer packing. These restrictions naturally lead to a greater correlation between various dihedral angles, which as a result increases the contribution of first two PCs into total atomistic motions. On the other hand, when the chain is disordered and samples a variety of weakly correlated conformations, dihedral angles make frequent transitions between allowable Ramachandran regions and cause increased PC degeneracy. For instance, first two PCs were able to capture ~50% of motions for a globular native protein eglin c,[74] wheras for IDPs like histone tails the contribution of two main PCs are below 20% (see Supporting Information).

Finally, the Gaussian chain model with volume interactions seems to be more adequate for histone tails at very high temperatures (see Supporting Information), which can be explained by both unraveling of secondary structure elements, as well as better screening of histone tail charges by mobile ions, leading to increased relative contribution of excluded volume interactions. In a notable contrast with regular globular proteins, which in general expand abruptly in an "all or none" fashion near the melting temperature, histone tails first contract when temperature is increased and then expand at higher temperatures. The dPCA free energy landscapes clearly show how, upon further heating, these collapsed states gradually and slowly swell, with fragmented PC basins increasing in number and coalescing into a large, smooth single basin at high temperatures (see Supporting Information). This collapse/expansion re-entrant transition results from the competition between the electrostatic interactions mediated by mobile ions, discussed above, and chain entropy, which eventually wins at high enough temperatures.

4. CONCLUDING REMARKS

Our analysis of 3-μs-long REMD simulation trajectories showed that equilibrium conformational ensembles of histone tails are composed of states with various degrees of residual order. We further characterized the ensemble behaviors by mapping out simplified free landscapes using the first two PCs. On every landscape we identified highly populated states and classified them by using secondary structural content and average pairwise structural overlap values. We found that the H4 tail is enriched in β-hairpins, H3 and H2B tails contain flickering α-helical segments, while the H2A tail is disordered, with partially stretched conformations.

On the basis of our findings we propose that states with high content of secondary structural order may be important for binding to linker DNA and acidic patches since rigidity of these elements provides higher affinity and lower entropic penalty for binding. From comparisons of chain size distributions we conclude that H4, H3, and H2B tails show behavior consistent with compact molten globular states such as those appearing on protein folding pathways, while the H2A tail samples predominantly random coil-like states.

In summary, our analysis demonstrated that the presence of secondary structural elements in histone tails creates a population of conformations that are drastically different from the random coil like states described by Gaussian chain with excluded volume interactions or worm like chain analogues. These results invalidate the naive picture of histone tails which are viewed as a collection of mostly disordered and structurally uncorrelated states. In addition, our investigation of histone tails' ionic environment and their temperature dependence revealed the remarkable nature of denaturation of histone tails which occurs via re-entrant mechanism, when the chain at first contracts and then slowly swells as temperature is increased. Finally, in the H4 and H3 tails, we found that a specific sequence motive, LYS-ARG, locally favors β-turn formation and kinks the chain making it more compact and favorable for secondary structure formation. Post-translation modifications of the corresponding LYS residues may potentially unlock these local structural motifs inducing large-scale conformational changes, leading to significant changes in tail-DNA and tail-histone core binding free energies, which can potentially account for the corresponding disruption of chromatin fiber observed in in vitro experiments . The main results of this paper should be helpful in interpreting future experiments on histone tails and better understanding of their functional roles.

■ ASSOCIATED CONTENT

Supporting Information. Seven figures in the Supporting Information document provide additional analysis of histone tail conformations and dynamics. Full citations for references 22 and 63 are given. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
gpapoian@umd.edu

■ ACKNOWLEDGMENT

We are grateful to Dr. Michael Rubinstein for insightful comments. This work was supported by the Beckman Young Investigator Award and Camille Dreyfus Teacher Scholar Award. This manuscript is dedicated to Dr. Michael Goldfeld on the occasion of his 70th birthday.

■ REFERENCES

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.;Walter, P. *Molecular Biology of the Cell*, 5th ed.; Garland Science, Taylor Francis Group, LLC: New York, 2007.
(2) Materese, C. K.; Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2009**, *131*, 15005–15013.
(3) Luger, K.; Hansen, J. C. *Curr. Opin. Struct. Biol.* **2005**, *15*, 188–196.
(4) Schiessel, H. *J. Phys.: Condens. Matter* **2003**, *15*, R699–R774.
(5) Widom, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 285–327.
(6) Luger, K.; Mder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251–260.
(7) Luger, K.; Richmond, T. J. *Curr. Opin. Genet. Dev.* **1998**, *8*, 140–146.
(8) Richmond, T. J.; Davey, C. A. *Nature* **2003**, *423*, 145–150.
(9) Sharma, S.; Ding, F.; Dokholyan, N. V. *Biophys. J.* **2007**, *92*, 1457–1470.
(10) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41–45.
(11) Nightingale, K. P.; O'Neill, L. P.; Turner, B. M. *Curr. Opin. Genet. Dev.* **2006**, *6*, 125–136.
(12) Margueron, R.; Trojer, P.; Reinberg, D. *Curr. Opin. Genet. Dev.* **2005**, *15*, 163–176.
(13) Rice, J. C.; Allis, C. D. *Nature* **2001**, *414*, 258–261.
(14) Dutnall, R. N. *Mol. Cell* **2003**, *12*, 3–4.

(15) Uversky, V. N.; Dunker, A. K. *Science* **2008**, *322*, 1340–1341.
(16) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys. Biomol. Struct.* **2008**, *37*, 215–246.
(17) Permyakov, S. E.; Bakunts, A. G.; Denesyuk, A. I.; Knyazeva, E. L.; Uversky, V. N.; Permyakov, E. A. *Proteins* **2008**, *72*, 822–836.
(18) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.
(19) Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14237–14238.
(20) Latzer, J.; Papoian, G. A.; Prentiss, M. C.; Komives, E. A.; Wolynes, P. G. *J. Mol. Biol.* **2007**, *367*, 262–274.
(21) Papoian, G. A.; Wolynes, P. G. *Biopolymers* **2003**, *68*, 333–349.
(22) Dunker, A. K.; *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.
(23) Soranno, A.; Longhi, R.; Bellini, T.; Buscaglia, M. *Biophys. J.* **2009**, *96*, 1515–1528.
(24) Mukhopadhyay, S.; Krishnan, R.; Lemke, E. A.; Lindquist, S.; Deniz, A. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2649–2654.
(25) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16764–16769.
(26) Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 5214–5223.
(27) Tran, H. T.; Mao, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2008**, *130*, 7380–7392.
(28) Vitalis, A.; Wang, X.; Pappu, R. V. *J. Mol. Biol.* **2008**, *384*, 279–297.
(29) Vitalis, A.; Wang, X.; Pappu, R. V. *Biophys. J.* **2007**, *93*, 1923–1937.
(30) Tran, H. T.; Pappu, R. V. *Biophys. J.* **2006**, *91*, 1868–1886.
(31) Sandelin, E. *Biophys. J.* **2004**, *86*, 23–30.
(32) Jayaraj, V.; Suhanya, R.; Vijayasarathy, M.; Anandagopu, P.; Rajasekaran, E. *Bioinformation* **2009**, *3*, 409–412.
(33) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
(34) Toth-Petroczy, A.; Simon, I.; Fuxreiter, M.; Levy, Y. *J. Am. Chem. Soc.* **2009**, *131*, 15084–15085.
(35) Zheng, C.; Hayes, J. J. *Biopolymers* **2003**, *68*, 539–546.
(36) Hansen, J. C. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 361–392.
(37) Kato, H.; Gruschus, J.; Ghirlando, R.; Tjandra, N.; Bai, Y. *J. Am. Chem. Soc.* **2009**, *131*, 15104–15105.
(38) Wang, X.; Moore, S. C.; Laszckzak, M.; Ausi, J. *J. Biol. Chem.* **2000**, *275*, 35013–35020.
(39) Banres, J. L.; Martin, A.; Parello, J. *J. Mol. Biol.* **1997**, *273*, 503–508.
(40) Liu, H.; Duan, Y. *Biophys. J.* **2008**, *94*, 4579–4585.
(41) Yang, D.; Arya, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2911–21.
(42) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846–1857.
(43) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07–L09.
(44) Arya, G.; Zhang, Q.; Schlick, T. *Biophys. J.* **2006**, *91*, 133–50.
(45) Arya, G.; Schlick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16236–41.
(46) Arya, G.; Schlick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16236–16241.
(47) Korolev, N.; Lyubartsev, A. P.; Nordenskild, L. *Biophys. J.* **2006**, *90*, 4305–4316.
(48) Sharma, S.; Ding, F.; Dokholyan, N. V. *Biophys. J.* **2007**, *92*, 1457–1470.
(49) Arya, G.; Schlick, T. *J. Phys. Chem. A* **2009**, *113*, 4045–4059.
(50) Zhuravlev, P. I.; Papoian, G. A. Q *Rev. Biophys.* **2010**, *43*, 295–332.
(51) Zhuravlev, P. I.; Papoian, G. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 16–22.
(52) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
(53) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
(54) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York, 2003.
(55) des Cloizeaux, J. *Phys. Rev. A* **1974**, *10*, 1665–1669.
(56) Fedorov, M. V.; Goodman, J. M.; Schumm, S. *J. Am. Chem. Soc.* **2009**, *131*, 10854–10856.
(57) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2008**, *112*, 9135–9145.
(58) Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2007**, *129*, 6060–6061.
(59) Formaneck, M. S.; Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 9506–9517.
(60) Ma, L.; Cui, Q. *Biochemistry* **2006**, *45*, 14466–14472.
(61) Mario-Ramrez, L.; Hsu, B.; Baxevanis, A. D.; Landsman, D. *Proteins* **2006**, *62*, 838–842.
(62) Cubizolles, F.; Gasser, S. M. *Genome Biol.* **2001**, *2*, REPORTS4023.
(63) Case, D.A.; , *AMBER 10*; University of California: San Francisco, 2008.
(64) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(65) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* **2008**, *130*, 6145–6158.
(66) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009**, *97*, 853–856.
(67) Showalter Scott, A..; B., R. *J. Chem. Theory Comput.* **2007**, *3*, 961–975.
(68) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
(69) Jean-Paul Ryckaert, G. C.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(70) Koji Hukushima, K. N. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
(71) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(72) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
(73)
(74) Materese, C. K.; Goldmon, C. C.; Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10659–10664.
(75) Zhuravlev, P. I.; Materese, C. K.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 8800–8812.
(76) Rousseau, R.; Schreiner, E.; Kohlmeyer, A.; Marx, D. *Biophys. J.* **2004**, *86*, 1393–1407.
(77) Schreiner, E.; Nicolini, C.; Ludolph, B.; Ravindra, R.; Otte, N.; Kohlmeyer, A.; Rousseau, R.; Winter, R.; Marx, D. *Phys. Rev. Lett.* **2004**, *92*, 148101.
(78) Garca *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.
(79) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. *J. Chem. Phys.* **2007**, *126*, 244111.
(80) Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005**, *58*, 45–52.
(81) Parisi, G. *Phys. Rev. Lett.* **1983**, *50*, 1946–1948.
(82) Skolnick, A. K. J. *J. Chem. Phys.* **1997**, *107*, 953–964.
(83) Ding, F.; Jha, R. K.; Dokholyan, N. V. *Structure* **2005**, *13*, 1047–1054.
(84) Altis, A.; Otten, M.; Nguyen, P. H.; Hegger, R.; Stock, G. *J. Chem. Phys.* **2008**, *128*, 245102.
(85) Hansen, J. C.; Tse, C.; Wolffe, A. P. *Biochemistry* **1998**, *37*, 17637–17641.
(86) Dang, W.; Steffen, K. K.; Perry, R.; Dorsey, J. A.; Johnson, F. B.; Shilatifard, A.; Kaeberlein, M.; Kennedy, B. K.; Berger, S. L. *Nature* **2009**, *459*, 802–807.
(87) Dion, M. F.; Altschuler, S. J.; Wu, L. F.; Rando, O. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5501–5506.
(88) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. *Science* **2006**, *311*, 844–847.
(89) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. *Biochemistry* **2006**, *45*, 2761–2766.
(90) Lee N, T. D. *Macromolecules* **2001**, *34*, 3446.
(91) H. Schiessel, P. P. *Macromolecules* **1998**, *31*, 7953–7959.
(92) Ha; Thirumalai *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *46*, R3012–R3015.

(93) Chepelianskii, A.; Mohammad-Rafiee, F.; Trizac, E.; Raphal, E. *J. Phys. Chem. B* **2009**, *113*, 3743–3749.
(94) Lo, T. S.; Khusid, B.; Koplik, J. *Phys. Rev. Lett.* **2008**, *100*, 128301.
(95) Hsiao, P.-Y.; Luijten, E. *Phys. Rev. Lett.* **2006**, *97*, 148301.
(96) Chitanvis, S. M. *Phys. Rev. E: Stat., Nonlinear Soft Matter Phys.* **2003**, *68*, 061802.
(97) Dobrynin, A. V. ; Colby, R.H.; Rubinstein, M. *Macromolecules* **1995**, *28*, 1859–1871.
(98) Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **1999**, *77*, 3242–3251.