JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

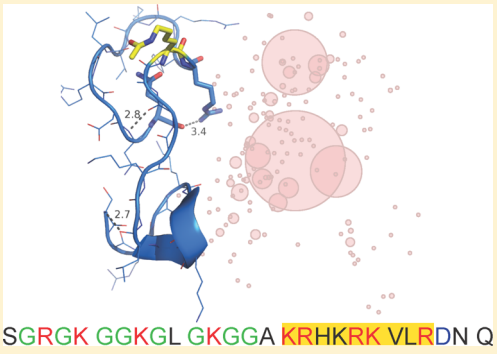
The Acetylation Landscape of the H4 Histone Tail: Disentangling the Interplay between the Specific and Cumulative Effects

David Winogradoff,[†,∥] Ignacia Echeverria,[‡,∥] Davit A. Potoyan,[‡,§] and Garegin A. Papoian*[,‡]

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

Supporting Information

ABSTRACT: Histone tails, the intrinsically disordered terminal regions of histone proteins, are key modulators of the structure and dynamics of chromatin and, consequently, are central to many DNA template-directed processes including replication, repair, and transcription. Acetylation of histone tails is a major post-translational modification (PTM) involved in regulating chromatin, yet it remains unclear how acetylation modifies the disordered state of histone tails and affects their function. We investigated the consequences of increasing acetylation on the isolated H4 histone tail by characterizing the conformational ensembles of unacetylated, mono-, di-, tri-, and tetra-acetylated H4 histone tails using Replica Exchange Molecular Dynamics (REMD) simulations. We found that progressive acetylation has a cumulative effect on the H4 tail, decreasing conformational heterogeneity, increasing helical propensity, and increasing hydrogen bond occupancies. The monoacetylation of lysine 16, however, has unique and specific effects: drastically decreasing the conformational heterogeneity of the H4 tail and leading to highly localized helical secondary structure and elongated conformations. We describe how the cumulative effects of acetylation arise from the charge reduction and increased hydrophobicity associated with adding acetyl groups, while the specific effects are a consequence of steric interactions that are sequence specific. Additionally, we found that increasing the level of acetylation results in the formation of spatially clustered lysines that could serve as recognition patches for binding of chromatin regulating proteins. Hence, we explore the mechanisms by which different acetylation patterns may result in specific recognition of the H4 histone tails by protein or DNA binding partners.



INTRODUCTION

Many proteins do not form well-defined three-dimensional structures in cells of higher organisms, yet they are biologically active and involved in a variety of biological processes.[1−9] These intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are characterized via heterogeneous structural ensembles, where the potentially complex conformational landscapes are regulated by tuning the nonspecific interactions, the overall chain entropy and the specific inter-residue interactions.[5,10] Disordered proteins play a key role in signaling and transcription regulation by interacting with each other or with more structured proteins. To accomplish their activities, some IDPs or IDRs undergo disorder-to-order transitions or bind to their biological partners by conformational selection,[11−17] though some are known to function without ever becoming structured.[18−20] Furthermore, IDPs and IDRs are richly regulated by a combinatorial variety of post-translational modifications (PTMs) that can significantly change their conformational and binding preferences.[21,22] The structural flexibility and conformational heterogeneity of IDPs and IDRs are considered advantageous and even essential to the biological complexity found in higher organisms, where the potential for highly intricate biological regulation emerges from the sophisticated and flexible interaction networks formed by the inclusion of these proteins.[9]

Histone tails, the terminal segments of histone proteins, are key IDRs that regulate the structure and dynamics of the genomic DNA-protein fibers, called chromatin, where the latter are central to many template-directed processes, including DNA replication, repair, and transcription.[23−25] Histone tails are highly flexible, highly positively charged and low in hydrophobicity peptides that feature multiple sites for potential PTMs, such as acetylation, methylation, and phosphorylation.[26] It was initially thought that histone tails acted mainly as unstructured electrostatic mediators and that PTMs simply function as modulators of these interactions. This implies, for example, that lysine acetylation, by neutralizing the positive charge of the lysine amino, reduced the electrostatic interactions between the histones and the DNA phosphates, making the DNA more accessible for active processes such as transcription. However, the realization that PTMs are highly diverse, acting individually or in various combinations, led to the hypothesis that the PTMs could form a histone code where highly specific PTM combinations specify different chromatin states.[24,27,28]

Received: January 8, 2015
Published: April 23, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

The combinatorial effect of different PTMs of histone tails can be interpreted by the distinct, yet overlapping, direct and effector-mediated mechanisms. In the direct mechanism, histone tails interact with the neighboring nucleosome, such that these interactions regulate internucleosomal structure. For example, histone tails are known to mediate intra- and internucleosome interactions with both protein and DNA in condensed chromatin structures.[29,30] In contrast, the effector-mediated mechanism postulates that PTMs serve as recognition sites for macromolecular complexes involved in chromatin remodeling activities,[26,31,32] which, in turn, can alter the chromatin architecture. A salient example of these effectors are bromodomains, a large family of proteins which recognize acetyl-lysine motifs, found in chromatin remodeling complexes.[26,33]

The H4 histone tail (Figure 1) has been identified to be of great importance for chromatin structure formation and stability.

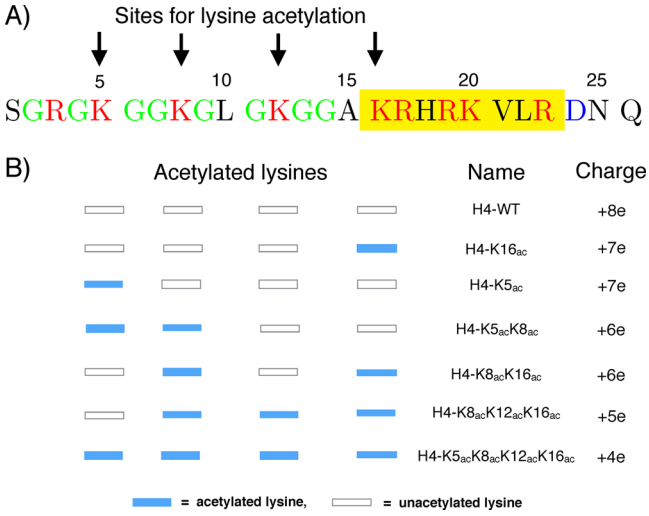


Figure 1. The H4 N-terminal histone tail sequence and acetylation sites. (A) The H4 histone tail sequence, including residue numbers, possible sites for acetylation, and residue types: glysines (green), positive residues (red), and negative residues (blue). A region of high positive charge density (i.e. basic patch), residues 16–23, is highlighted in yellow. (B) Studied models with different levels of acetylation. Solid blue boxes highlight the specific sites of lysine acetylation. The given name and net charge for each of the studied models is provided as well.

For example, early *in vitro* experiments of nucleosomal arrays determined that histone tails are necessary for the stability of higher order chromatin structure,[34−36] and, furthermore, that the acetylation of histone tails plays a key role in regulating chromatin structure by, for example, disrupting the formation of the 30 nm fiber.[37−39] In a landmark *in vitro* study, the homogeneous monoacetylation of the H4 histone tail at lysine 16 alone was enough to inhibit the formation of higher order chromatin structures and impede the interactions between chromatin and nonhistone proteins.[40] *In vitro* analysis of the cation-induced nucleosome−nucleosome association determined that the nucleosome stacking is mainly governed by electrostatics interactions that are modulated by ion−ion correlations and histone-tail bridging.[41] The latter is mediated by a region of high charge density of the H4 tail, which includes lysine 16 (Figure 1), that interacts with the acidic patch on the H2A-H2B dimer of an adjacent nucleosome.[29,42−44] Consequently, acetylation of lysine 16 directly alters these interactions, disrupting the nucleosome−nucleosome stacking.[41] Besides the charge reduction associated with acetylation, recent studies have found that the monoacetylation of lysine 16 induces the partial ordering of the H4 tail, increases the affinity to DNA,[45] and is associated with the formation of transient elements of secondary structure.[10,45−48]

The unique role of lysine 16 acetylation has been further confirmed by systematic genetic studies which show that single lysine mutations of the H4 tail do not result in defects in chromatin assembly or DNA replication, with the exception of the single mutation of lysine 16. Similarly, the mutation of all four H4 tail lysines is lethal, but not triple mutations. These studies suggest that the different H4 tail lysines are partially redundant and that acetylation of the histone H4 tails may be mediated through two distinct mechanisms: a cumulative and nonspecific effect for lysines 5, 8, and 12 and a specific mechanism for lysine 16.[49,50]

Our understanding of the effects of different levels of H4 acetylation has been obtained, mainly, from functional,[40,51] biochemical,[52] and genetic analysis,[49,50] especially from the proteins that recognize these modifications[26] as well as from the conformational effects in chromatin. The histone-code hypothesis suggests that PTMs, such as lysine acetylations, may yield rich combinatorial outputs. However, it is unclear how different levels of acetylation change the conformational preferences of the H4 tail and if these effects are combinatorial or cumulative, in particular from a structural viewpoint. Further understanding of how acetylation changes the conformational landscape of the H4 tails would allow us to understand, for example, the role of conformational selection or induced fit mechanisms in H4 tail recognition, i.e., binding competent conformations, which are well-visited already in the unbound state are selected (conformational selection) or binding partner induces the required binding conformations (induced fit). For example, recent experimental and computational studies illustrate that histone tails display transient elements of secondary structure,[10,45,53,54] suggesting that histone tails could participate in specific interactions that tune chromatin structure. In this context, PTMs would add another important layer of control by regulating the histone tail conformations.

Here, we explore the effects of different levels of acetylation on the conformational preferences of an isolated H4 histone tail and how these preferences are modulated by different levels of acetylation. Using all-atom replica exchange molecular dynamics (REMD) simulations in explicit solvent, we determined how various combinatorial acetylation patterns affect the peptide's conformational landscapes. We found that progressive acetylation reduces the conformational heterogeneity of the sampled states, altering both short and long-range interactions. Increased acetylation also results in greater helical propensities and hydrogen bond occupancies without significantly changing the overall radii of gyration. These cumulative effects of acetylation highlight how the charge reduction and increased hydrophobicity associated with adding the acetyl groups enhance cohesive interactions within the peptide. Our results also show that the sole acetylation of lysine 16 has structural effects that are unique to this case, including a significant reduction in the number of states sampled and the formation of a specific 3_{10} helix corresponding generally to more elongated structures and specifically to a unique positioning of the lysine 16 residue. At the microscopic level, the specific effects of the sole acetylation of lysine 16 include effectively rigidifying the peptide, setting an entropic constraint on the accessible conformations, and leading

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

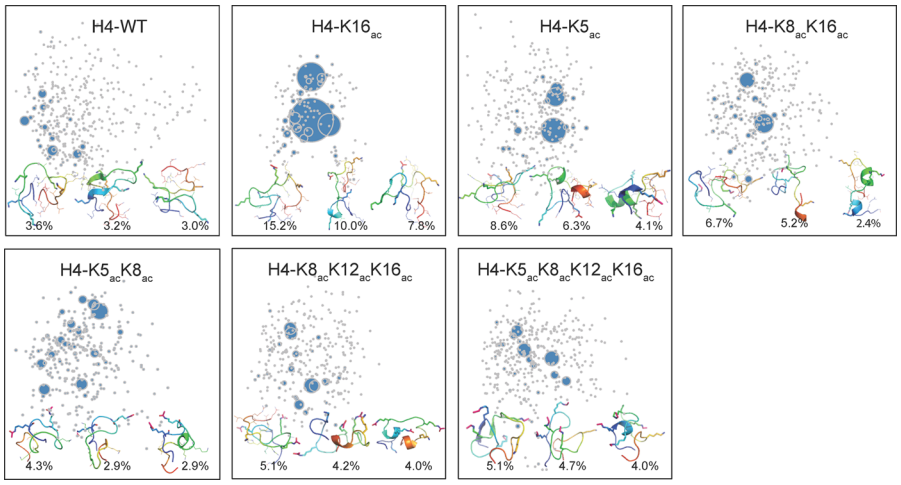


Figure 2. Conformational clustering analysis of each of the H4 models. In each panel we present the results of the clustering analysis. Each cluster is represented by a point, with its size proportional to the number of trajectory frames in it. The distances between points are approximately equal to their dissimilarities. The three most representative structures (i.e., largest clusters) are shown in each panel along with the percentage of the sampled ensemble that that cluster represents. Structures are shown colored in rainbow mode from blue (N-terminal) to red (C-terminal) and with elements of secondary structure shown in cartoon representation. Lysines are shown in sticks representation.

to the formation of elongated structures. Results presented here provide, therefore, a structural characterization baseline of the conformations of the H4 tail. These results provide important information to understand the effects of the histone cores and/or nucleosomal DNA on the conformations of the H4 tail under physiological conditions.

METHODS

Molecular Dynamics Simulations. We performed all-atom molecular dynamics (MD) simulations using the Amber12 MD software,[55] the amber99SB*[56] force field for proteins, the ions94[57] force field for ions, and the TIP3P water model. Starting from the wild-type (H4-WT) or unacetylated H4 N-terminal histone tail model from an earlier work,[10] we used the *xleap* tool in AmberTools12 to prepare WT, mono-, di-, tri-, and tetra-acetylated H4 N-terminal histone tails. We modeled two monoacetylated tails with acetyations at lysine 16 (H4−$K16_{ac}$) and lysine 5 (H4−$K5_{ac}$); for diacetylated, we acetylated lysines 8 and 16 (H4−$K8_{ac}K16_{ac}$) and lysines 5 and 8 (H4−$K5_{ac}K8_{ac}$); for triacetylated, we acetylated lysines 8, 12, and 16 (H4−$K8_{ac}K12_{ac}K16_{ac}$); for tetra-acetylated, we acetylated lysines 5, 8, 12, and 16 (H4−$K5_{ac}K8_{ac}K12_{ac}K16_{ac}$), every possible site. We did not consider lysine 20 given that it is mostly found methylated.[58] We determined the specific sites for lysine acetylation by their physiological abundance.[59] For further details about the simulations setup and convergence tests, see the Supporting Information and Figure S1.

To characterize the conformational ensemble of the H4 tail at different levels of acetylation, we performed replica exchange molecular dynamic (REMD) simulations.[60] Exchanges between replicas at different temperatures enhance the conformational sampling relative to standard MD simulation, creating an ensemble that includes both high and low energy configurations. First, each system was copied to generate a total of ∼60 replicas. The temperatures used in REMD simulations, ranging from 300 to 450 K, were determined by T-REMD,[61] an REMD temperature online server, with a target exchange probability of 30%. Then, each replica was heated to the desired temperature over 500 ps in the NVT ensemble. REMD production runs were performed in the NVT ensemble, attempting exchanges every 5 ps with a 2 fs time-step, saving coordinates and energies every picosecond for further analysis. 100 ns of REMD simulations were performed for each system using the Langevin thermostat with a 2 ps time constant, totaling 6 μs of simulation for each system. The exchange probability observed for each system was ∼50%. For analysis we only considered only the final 90 ns of trajectories set to 300 K. This allowed us to account for further thermal equilibration.

Analysis of the Trajectories. We determined the radius of gyration (R_g) along each trajectory for all the studied systems. The obtained values were first compared to the prediction for a globular protein of the same length ($R_{g,\,globular} = 2.2N^{0.38}$), a relation based on a power law best fit of R_g as a function of sequence length for a subset of proteins in the PDB.[62] Similarly, the R_g values were compared to the prediction for a thermally denatured random coil of the same length ($R_{g,denatured} = 2.02N^{0.60}$), a relation proposed by Flory's theory and confirmed by power fitting R_g values determined by computation and experiment.[63]

The secondary structure present for each simulation snapshot was determined with the Amber 12 *secstruct* tool, which uses the DSSP program[64] to identify hydrogen bond motifs through backbone amide (N−H) and carbonyl (C═O) group positions. By definition, a 3_{10} helix spans at least three consecutive residues requiring two hydrogen bonds between residues (i, $i + 3$), and an α-helix spans at least four consecutive residues requiring two hydrogen bonds between residues (i, $i + 4$). For each residue, we determined the percentage of simulation snapshots where each residue is part of a 3_{10} or α-helix, which we will refer to as the helix propensity per residue. Also, we identified all of the protein's hydrogen bonds. A geometric definition of a hydrogen bond was used: two heavy atoms are considered to be bonded if (1) their donor−acceptor distance is <3.5 Å and (2) the acceptor−donor-hydrogen angle is <30°. Furthermore, we analyzed inter-residue contact preferences. We identified contacts between residues, excluding (i, $i \pm 1$) residue pairs, for all levels of acetylation. A contact was determined to exist when the distance between two non-hydrogen atoms from different residues was <3.6 Å. Contacts were first calculated as percentages of their respective entire trajectories. Then, we divided the H4 tail into five segments: residues 1−5, residues 6−10, residues 11−15, residues 16−23 (the aforementioned basic patch), and residues 24−26. Then, contact pairs between, and within, segments were considered together as sums. We present contacts between and within segments for all acetylated tails as ratios relative to the corresponding intersegmental contact sums for the unacetylated WT H4 tail. Lastly, we used Ramachandran plots of key residues to graphically display specific backbone dihedral angle preferences.

Clustering Analysis. We performed clustering analysis to characterize the conformational ensemble sampled during the REMD simulations. For this purpose we defined the dissimilarity metric as the pairwise RMSD, after proper alignment, between the backbone atoms of simulation snapshots selected every 1 ps ($N \sim$ 88,000 structures per trajectory). Following a bottom-up approach,[65] clustering was performed as follows: we computed the RMSD between structures i and j, if the RMSD was smaller that a given cutoff ($RMSD_{cutoff}$), the

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

structure j was added to cluster i, and the center of cluster i was defined as the average between structures i and j.

Following, we computed the RMSD between the average structures of clusters i and $j + 1$ and merged the clusters if the RMSD < $RMSD_{cutoff}$. Conversely, if RMSD > $RMSD_{cutoff}$, $j + 1$ is defined as its own cluster. We repeated this procedure until we compared the structure $i = N$ to all clusters. In cases where a structure could be assigned to two or more distinct clusters, we assigned it to the cluster to which its RMSD with respect to the cluster's center was the smallest. The $RMSD_{cutoff}$ was set to 2.8 Å.

After the first round of clustering, we redefined the center of each cluster by identifying the reference structure. To determine the reference structure, we computed the average position of the backbone atoms over all the simulation snapshots in the cluster and then identified the structure with the lowest RMSD with respect to the average backbone. A second round of clustering followed considering the obtained reference structures. We found that after two rounds of clustering we obtained a converged set of clusters. Among the advantages of this clustering approach is that the $RMSD_{cutoff}$ is the only free parameter and that it is not necessary to define the desired number of clusters beforehand. The distance between the clusters was defined as the RMSD between the representative structures (i.e., structures at the center of the cluster).

To visualize the results obtained from the clustering analysis, we computed the principal components of the dissimilarity matrix $\mathbf{R} = R_{ij}$, where R_{ij} is the pairwise RMSD between the reference structures of clusters i and j. By computing the first two principal components, we obtained a set of points in two-dimensional space such that the distances between the points are approximately equal to their dissimilarities (Figure 2). The size of each cluster is proportional to the number of structures in each cluster.

To assess the heterogeneity of conformational ensemble sampled by the WT and acetylated H4 histone tails, we determined the cumulative percentage of structures represented in the clusters, after ordering clusters by size. Additionally, we determined the distances of all clusters to the center-of-mass. The center-of-mass was determined as $COM = \sum_{i=1}^{M} S_i X_i / M$, where S_i is the size of cluster i, X_i is the position of clusters i, defined by the principal components, and M is the number of clusters. We analyzed the representative structures of the most populated clusters to determine the stabilizing interactions and the molecular determinants of prominent structural features.

RESULTS AND DISCUSSION

Acetylation of the H4 Tail Reduces the Conformational Heterogeneity of the Sampled Ensemble. To elucidate the effects of acetylation on the conformational preferences of the H4 histone tail, we performed conformational clustering analysis on each of the studied models. This analysis allowed us to represent the peptides' conformational space by a set of discrete microstates, one corresponding to each cluster, where the most prominent structures are identified as those belonging to the largest clusters. Figure 2 illustrates that the acetylation of the H4 tails has a significant effect on the number of sampled microstates and their dissimilarity. For example, the WT H4 explores a sparse set of conformations, which are shown as small clusters (i.e., with only few simulation snapshots per cluster) that are far apart structurally from each other, indicating high dissimilarity. In contrast, the sole acetylation of K16 has a dramatic effect on the conformational preferences of the peptide, significantly reducing the number of clusters. For H4–$K16_{ac}$ we identified three prominent structures which represent 15%, 10%, and 8% of all the sampled structures.

To explore how acetylation affects the heterogeneity of the conformational ensemble, we determined the cumulative number of clusters necessary to account for a fraction of the sampled structures (Figure 3A). For example, the largest cluster identified for the WT represents 3.5% of all the sampled

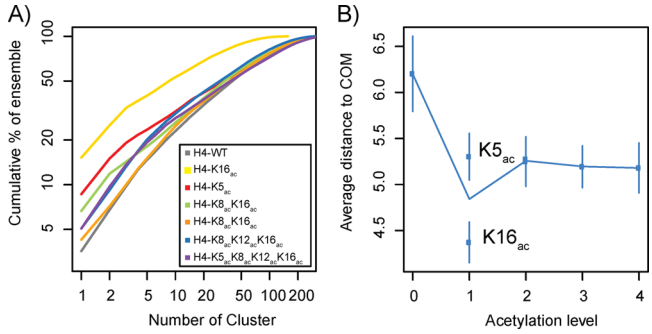


Figure 3. Heterogeneity of the conformational space of WT and acetylated H4 tails. (A) Analysis of the number of clusters necessary to account for the sampled conformations, after ordering clusters by size. (B) Average distance of all clusters to the center-of-mass.

structures. Analogously, for the acetylated H4–$K16_{ac}$, H4–$K5_{ac}$, H4–$K8_{ac}K16_{ac}$, H4–$K5_{ac}K8_{ac}$, H4–$K8_{ac}K12_{ac}K16_{ac}$, and H4–$K5_{ac}K8_{ac}K12_{ac}K16_{ac}$ models, the largest clusters represent 15%, 8.6%, 6.7%, 4.3%, 5%, and 5% of all the sampled structures, respectively. In general, the monoacetylation of K16 has the largest effect in changing the conformational preferences of H4, as evidenced by the presence of large clusters. The effects of di-, tri-, and tetra-acetylations appear to, in part, to counteract this effect. However, for all levels of acetylation, we observe that some conformations are favored. We will discuss below how these conformational preferences may play a role in the context of binding and recognition of histone tails.

We further characterized the heterogeneity of the conformational ensemble by measuring the average distance from all clusters to the center-of-mass (see Methods section). Using this metric, we determined that the WT exhibits the most heterogeneous ensemble, while H4–$K16_{ac}$ has the least heterogeneous ensemble. For all other models, we observed the general trend of decreasing structural heterogeneity with increasing the extent of acetylation (Figure 3B). These results again indicate that the monoacetylation of K16 stands out from the overall trend, crucially altering the conformational landscape of the H4 histone tail. An alternative approach to quantify the conformational heterogeneity is to determine the distribution of the pairwise RMSD or Q between all sampled structure, as discussed in the Supporting Information (Figure S2).[45,66] Below, we further elaborate on the way the various levels of acetylation structurally affect the conformational preferences of the tails.

Acetylation of K16, but Not Other Lysines, Leads to More Extended Conformations. We used the radius of gyration (R_g) as a coarse metric to assess the type of conformations sampled by the different models. For all levels of acetylation, the average radii of gyration of the H4 histone tail are between the predicted values for globular (7.6 Å) and thermally denatured random coil (14.3 Å) proteins of the same length, indicating that the H4 tail adopts molten globule-type conformations (Figure 4A). Furthermore, the average R_g values for all the systems studied (Table S1) are closer to the predicted value for globular proteins than to the prediction for a thermally denatured random coil, suggesting that the H4 tail could include elements of secondary structure. However, the H4–$K16_{ac}$ tail exhibits a slightly greater average R_g, and its R_g probability distribution features two distinct peaks, as opposed to one. We observe that, to a lesser extent, the H4–$K8_{ac}K16_{ac}$ model also has a bimodal R_g distribution. These results indicate that, even though the H4 tails are highly charged peptides, the global

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245–6253

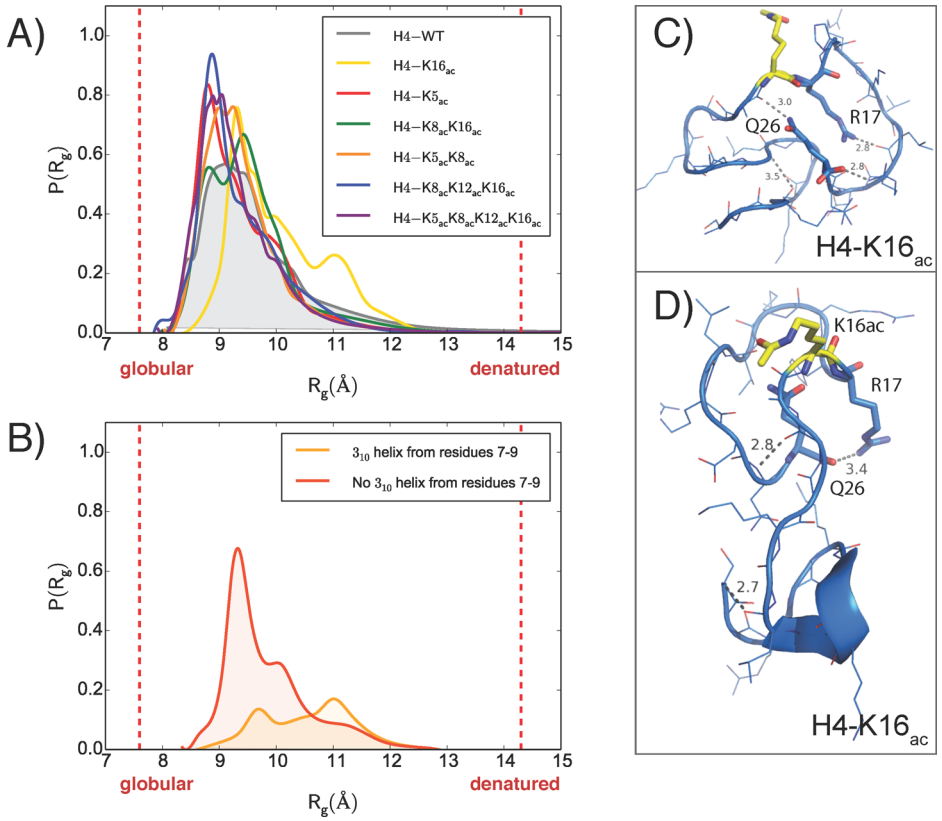


Figure 4. Probability distributions of R_g. (A) R_g at different levels of acetylation. The vertical red lines represent the predicted R_g values for globular (7.6 Å) and thermally denatured random coil (14.3 Å) peptides of the same length ($N = 26$ residues). (B) R_g of the H4 tail monoacetylated at K16 divided into two groups: (1) simulation frames with a 3_{10} helix from residues 7−9 and (2) simulation frames without this specific helix. (C) Characteristic structure obtained from the most populated cluster of the H4−K16$_{ac}$ showing the stabilizing interactions between R17 and Q26 with the backbone. The acetylated K16 is shown in yellow. (D) Sample structure from the second most populated cluster exhibiting the H4−K16$_{ac}$ characteristic 3_{10} helix. This structure is stabilized mainly by backbone−backbone hydrogen bonds. The acetylated K16 is shown in yellow.

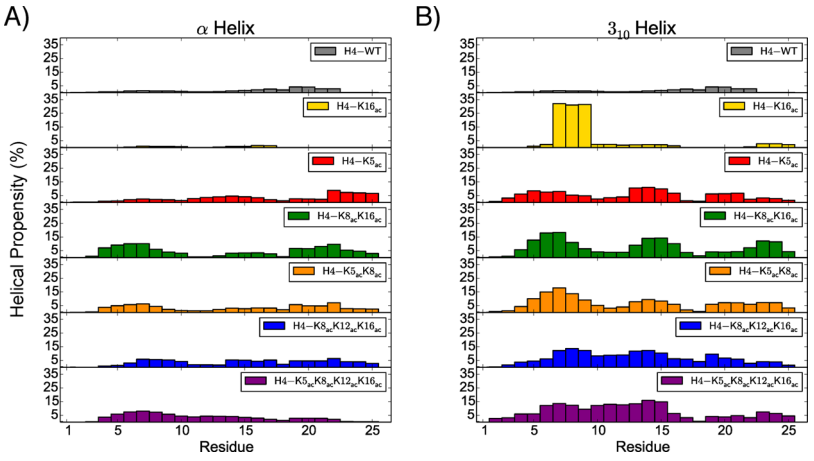


Figure 5. Helical propensity per residue. (A) α-helical propensity per residue as a fraction of time, across all levels of acetylation. (B) 3_{10} helical propensity per residue.

dimensions are not strictly dictated by the repulsive electrostatic interactions.[67]

Acetylation of H4 Tails Induces Increased Helical Propensities. To better understand the preferences of the H4 tail to form secondary structure, we investigated the helical propensities, per residue, at different levels of acetylation. Our results demonstrate that the WT H4 histone tail has a small amount of α and 3_{10} helical propensity spanning the entire sequence and, furthermore, that the helical propensity spanning the whole sequence increases with acetylation, leveling off beyond diacetylation (Figure 5 and Table S1). The profile of helical propensity for the H4 tail monoacetylated at lysine 16 (H4−K16$_{ac}$), however, is unique; upon K16 monoacetylation, the H4 histone tail undergoes a significant structural rearrangement whereby the helical propensity becomes highly localized, featuring a specific 3_{10} helix from residues 7−9, which is formed ∼30% of the time. The formation of this helix is intermittent throughout the simulation (Figure S3). This individual helix is

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

the single most prominent element of secondary structure among all levels of acetylation (Figure 5). In contrast to the monoacetylation of K16, the H4 tail monoacetylated at lysine 5 (H4−K5$_{ac}$) induces a slight increase in the 3_{10} and α helical propensities relative to the WT tail that continues to span the entire sequence, the average helical propensity per residue increasing from ∼5% for the H4-WT tail to ∼7% for the H4−K5$_{ac}$ tail (Table S1). The significant difference between the structural effects of K16 monoacetylation and K5 monoacetylation could be explained, in part, by the location of these specific residues in the sequence of the H4 histone tail; K16 is found within a region of high charge density close to the C-terminus of the tail, and, in contrast, K5 is located in a region of relatively low charge density close to the N-terminus (Figure 1). Di-, tri-, and tetra-acetylated H4 tails exhibit a greater increase in the 3_{10} and α helical propensities than K5, spanning the entire sequence, further increasing the average helical propensities per residue to ∼10% (Table S1, Figure S4). Furthermore, the two patterns of diacetylation have similar effects on the helical propensity of the H4 tail (Figure 5, Table S1, Figure S4). For all levels of acetylation, the helical propensity is primarily 3_{10} helix, and the majority of helices formed are short, either three or four residues long. The overarching trend of helical propensity increasing with acetylation (Figure S4) is consistent with a previous circular dichroism (CD) experiment, which reported the α helical content of the H4 tail increases monotonically with progressive acetylation,[54] and several recent computational MD studies performed with explicit solvent,[10,45,46,48] but not with a computational study with implicit solvent.[47] However, quantitative comparison with experiments is currently not feasible, given that helicities were experimentally measured in the context of the nucleosomal core particles. Histone cores and nucleosomal DNA can alter the structural preferences of the H4 histone tail and may contribute to the higher helicities observed by Wang et al.[54] (Figure S4).

The unique effect of the sole acetylation of K16 on helical propensity is consistent with our clustering analysis, where an H4−K16$_{ac}$ structure with a 3_{10} helix from residues 7−9 is representative of the second most populated cluster (Figures 2 and 4D). The presence of the 3_{10} helix from residues 7−9 also explains why the H4−K16$_{ac}$ model exhibits a higher R_g. For example, by dividing the H4−K16$_{ac}$ sampled structures into two groups, one with the 3_{10} helix from residues 7−9 and the other without this helix, we discovered that structures characterized by the presence of this specific helix correspond to the more extended conformations (Figure 4D) and contribute to the second peak in the probability distribution (Figure 4B). By performing a similar division of the H4−K8$_{ac}$K16$_{ac}$ sampled structures based on helical structure, we found specific 3_{10} helices contribute to slightly more extended conformations as well (Figure S6).

Acetylation of H4 Tails Increases the Long-Range Contact Occupancies. The general features of conformational ensembles depend, in part, on the formation of secondary structure and other stabilizing interactions, such as hydrogen bonds, hydrophobic contacts, and salt bridges. As already described, acetylation is associated with an increase in helix propensities (Figure 5). To explore the role of nonlocal hydrogen bonds, we characterized the hydrogen bond occupancies of the peptide side chains. We found that acetylated lysines have a higher probability of participating in hydrogen bonds (Figure 6A). These hydrogen bonds include mainly lysine-backbone contacts (Figure 6B). Furthermore, by analyzing the

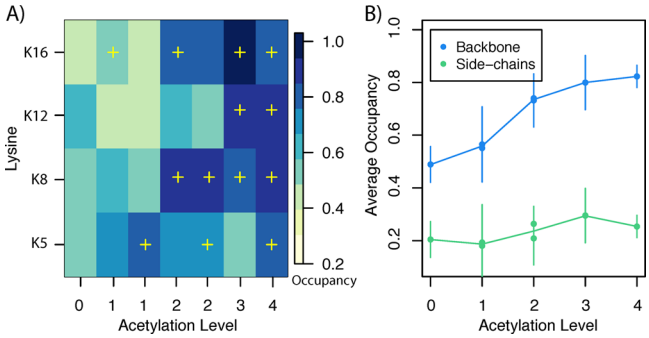


Figure 6. Hydrogen-bond occupancies of lysines. (A) Hydrogen bond occupancies measured as the percentage of time that every lysine is forming a hydrogen bond. Yellow crosses indicate acetylated lysines. (B) Average hydrogen bond occupancies at different levels of acetylation. Hydrogen bonds have been divided between lysine-backbone (blue) and lysine-side-chain (green).

hydrogen occupancies of other residues in the peptide we determined that the increased hydrogen-bond occupancies is specific to acetylated-lysines and that hydrogen bond occupancies of other side-chain do not increase with acetylation level (Figure S7). We also determined that salt bridges play a prime role in stabilizing the H4 tail conformations for all models, except H4−K16$_{ac}$ (Figure S8).

To further understand the role of acetylation in the conformational ensemble, we analyzed the inter-residue contact preferences. By dividing the peptide in five segments, one of which is the basic patch, we computed the relative segment contact occupancies with respect to the WT. Figure 7 shows that the sole acetylation of K16 disrupts the contacts (i.e., lower occupancies) between the basic patch and the first half of the peptide, which is in agreement with the results showing that the H4−K16$_{ac}$ model samples more elongated conformations. For all other studied levels of acetylation, we observe an increase of contact occupancies of the basic patch with other segments of the peptide. Consequently, in general, acetylation has a cumulative effect of making the peptide more cohesive, favoring the formation of contacts between different parts of the peptide, as shown by the hydrogen bonds (Figure 6) and the contact maps (Figure 7) analyses.

At a molecular level, the unique effects of K16 acetylation can be traced to the local conformational preferences. Our previous work demonstrated that, in the WT H4 tail, K16 and other residues in the region of high charge density are conformationally constrained, sampling only a fraction of the sterically allowed conformations described by the Ramachandran plot.[10] Figures 8 and S9 reveal that the acetylation of K16 further reduces the backbone's conformational flexibility, especially for K16 and R17. These changes in the Ramachandran plots reflect the fact that residues K16 and R17 adopt mostly a trans conformation, while in all other models, these residues sample an equilibrium between the cis and trans conformations. The corresponding entropy reduction can be explained, in part, by the steric constraints of adding an acetyl group to the basic patch. However, this effect is reversed by further acetylating the H4 tail, where additional acetylated lysines promote intrachain contact formation, overcoming the steric constraints in the segment flanking the K16$_{ac}$ residue. Consequently, the formation or breaking of the contacts between the basic patch and the rest of the peptide sensitively depends on the interplay between the electrostatic and steric interactions and the entropic effects.

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

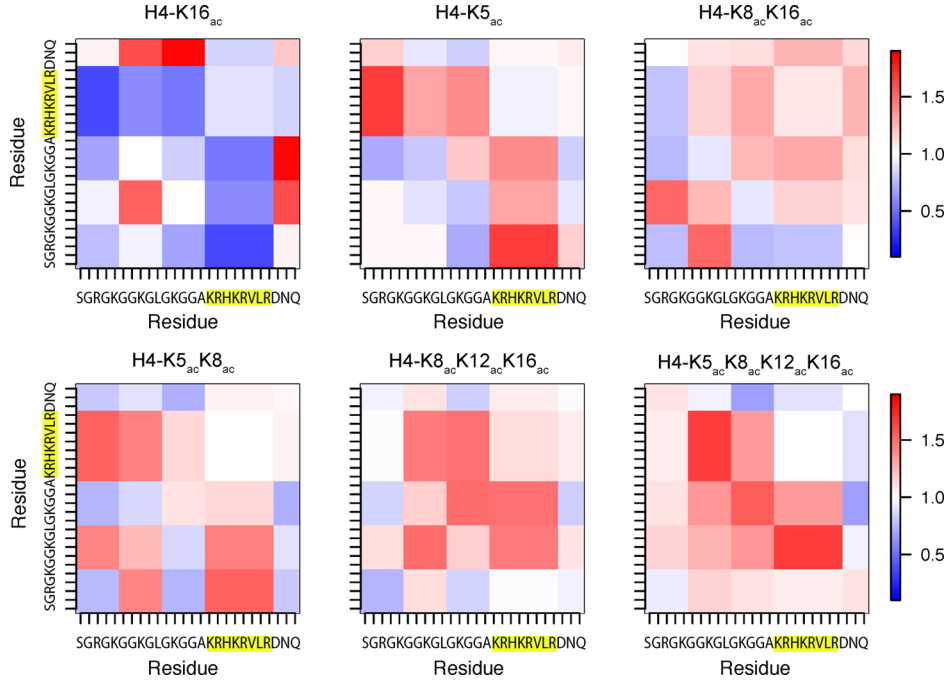


Figure 7. Contact maps between different peptide segments. Contact maps were computed by, first, determining the occupancy of the segment contacts and then considering the ratio between the acetylated model and the WT. The ratios presented range from blue (a contact occupancy decrease with respect to the WT) to red (a contact occupancy increase with respect to the WT). The region of high positive charge density, residues 16−23, is highlighted in yellow.

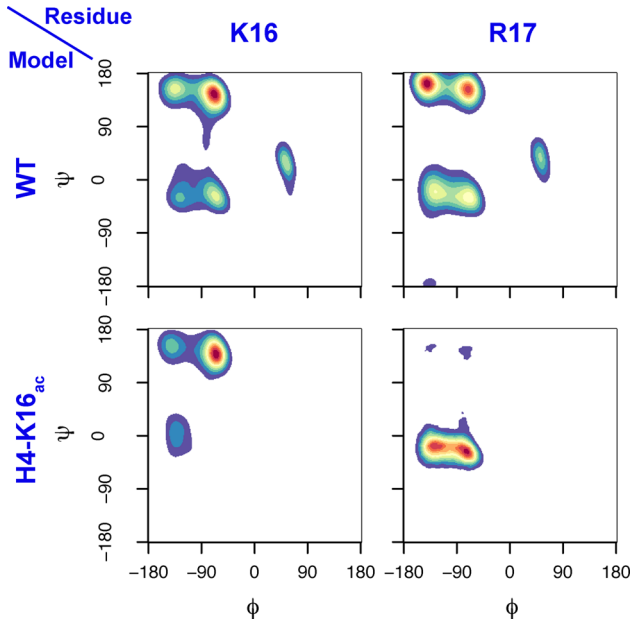


Figure 8. Ramachandran plots of residues K16 and R17.

Finally, to rationalize the unique features of the H4−K16$_{ac}$ we propose that the sole acetylation of K16 results in a cooperative transition, where the acetylation of K16 fixes the stereochemistry of residues in the basic patch and favors the trans conformation of residues K16 and R17 (Figure 8), resulting, subsequently, in the formation of the stable 3_{10} helix in the chain segment from residues 7 to 9. Furthermore, this 3_{10} helix structurally divides the tail into two domains, where the interactions between the residues before the secondary structural element and the residues after are highly disrupted (Figures 7 and 4C). This molecular level explanation highlights both local and global effects of K16 acetylation, whose uniqueness stems from K16's location in the region of high positive charge density and sterically constrained side chains.

Proposed Model for the Recognition of Acetylated H4 Tails. We determined that even though the average radii of gyration of H4 tails remain mostly constant, progressive lysine acetylation significantly changes the conformational preferences and long-range contacts of H4 tails,[68] thus modifying the sampled ensemble. To understand how these changes might affect the recognition of H4 tails by other nucleosomes or proteins involved in chromatin structure regulation, we analyzed how the lysine residues are positioned at different levels of acetylation. By measuring the average distance between the NZ atoms of all lysines, we found that increasing the level of acetylation correlates with a smaller average distance between lysine side chains (Figure 9), which can be explained, in part, by the overall charge reduction upon acetylation. By favoring

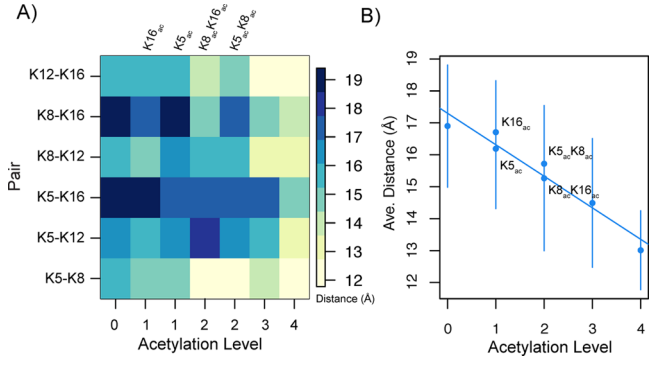


Figure 9. Average distance between lysine side-chains. (A) Average distance between all lysine pairs. (B) Average distance between all lysines at different levels of acetylation.

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

conformations that bring acetylated lysines close to each other, H4 tails can create spatial clusters that act as recognition patches or docking sites for acetylation-dependent histone tail-binding proteins, via conformational selection, induced fit or a mixed mechanism.[52,69,70] This is in agreement with structural studies showing that, for proper binding, acetylation sites have to be closely spaced, such that a single acetyl-lysine binding protein can recognize more that one acetylation mark.[52,69]

As already discussed, the acetylation of K16 has effects that are unique in many of the considered metrics, exhibiting, for example, the lowest conformational heterogeneity and favoring structures with a specific 3_{10} helix spanning residues 7−9. This helix effectively promotes elongated structures with higher R_g that position the acetylated K16 residue opposite to the N-terminal of the peptide, hence, exposing the lysine 16 residue, making it available for a specific recognition by various binding partners, and leading, in turn, to unique biological consequences for this particular modification, in agreement with various experimental observations.

CONCLUSION

Our analysis of all-atom REMD simulations of the H4 N-terminal histone tail highlighted the effects of different levels of acetylation on its conformational preferences. We demonstrated that, with the exception of the monoacetylation of K16, progressive acetylation has a cumulative effect on both global and specific features of the conformational ensemble of the H4 tail. For example, our clustering analysis revealed that conformational heterogeneity decreases with acetylation. We also find that progressive acetylation results in higher helical propensities, both 3_{10} and α-helices.

Acetylation influences specific interactions between amino acids by increasing the hydrogen bond occupancy of acetylated lysines. The described structural changes occur mostly without significantly changing the average radius of gyration, suggesting that acetylation results in local perturbations that modify the structural preferences and heterogeneity of the H4 tail conformational ensemble at the local level. Overall, our investigation suggests that the electrostatic charge reduction and increased hydrophobicity upon acetylation are responsible for the cumulative effects of this post-translational modification on the H4 histone tail. Furthermore, by using a variety of metrics to characterize the disordered state of the H4 tail, we demonstrate that local changes, such a formation of secondary structure, do not necessarily change some of the global or average properties of the polypeptide (i.e., radius of gyration, lysine hydrogen bonding, and interlysine distances).

While we identified many of the effects of acetylation to be cumulative in nature, our analysis illustrates that the effect of K16 monoacetylation is unique. K16 is found within a region of high positive charge density of the H4 tail, which plays an important role in binding and recognition interactions.[41] We illustrate how the monoacetylation of K16 has unique global effects in the corresponding conformational ensemble, which is the least heterogeneous and exhibits a larger radius of gyration. We determined that the acetylation of lysine 16 effectively introduces a soft entropic penalty, rigidifying the chain in the vicinity of lysine 16. These local effects on one segment of the H4 tail induce the formation of highly localized, specific helix in the H4−$K16_{ac}$ system, leading to more elongated chain conformations. Furthermore, these elongated conformations may play a key role in exposing the acetylated K16 residue for participating in direct and/or effector-mediated interactions with various chromatin regulatory proteins and DNA, which have a special role in chromatin structure[40] and transcription regulations.[49]

We propose that spatial clustering of the acetyl-lysines will create recognition patches that could facilitate the recruitment of effector proteins via conformational selection, induced fit, or a mixed mechanism. This result is consistent with structural studies that show that the binding of H4 tails to bromodomains, or other acetylation-dependent histone tail-binding proteins, often require patterns of acetylation marks which, in turn, are highly sensitive to modifications flanking the acetylation site.[52,69] These studies also show that these proteins bind more strongly to H4 tails with higher levels of acetylation.[69] Overall, our study suggests that the acetylation code for an isolated H4 histone tail has largely cumulative effects in the conformational preferences of the peptide. However, highly specific effects were seen for one of the acetylation patterns. Future research should address how the presence of histone cores and/or nucleosomal DNA further modifies these conformational preferences.

ASSOCIATED CONTENT

Supporting Information

Further details of the H4 sampled conformations are available in the Supporting Information section. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b00235.

AUTHOR INFORMATION

Corresponding Author

*gpapoian@umd.edu

Present Address

[§]Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, United States.

Author Contributions

[‖]These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the University of Maryland and by the National Science Foundation (NSF grants CHE-1363081 and CHE-0846701 to G.A.P.). Computational resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (grant TG-MCB120051 to I.E., G.A.P. and D.W.).

REFERENCES

(1) Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11504−11509.

(2) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321−331.

(3) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573−6582.

(4) Tompa, P. *Bioessays* **2003**, *25*, 847−855.

(5) Papoian, G. A.; Wolynes, P. G. *Biopolymers* **2003**, *68*, 333−349.

(6) Papoian, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14237−14238.

(7) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* **2010**, *1804*, 1231−1264.

(8) Dyson, H. J. *Q. Rev. Biophys.* **2011**, *44*, 467−518.

(9) van der Lee, R.; et al. *Chem. Rev.* **2014**, *114*, 6589−6631.

(10) Potoyan, D. A.; Papoian, G. A. *J. Am. Chem. Soc.* **2011**, *133*, 7405−15.

(11) Kohler, J. J.; Metallo, S. J.; Schneider, T. L.; Schepartz, A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11735−11739.

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253

(12) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873−6888.
(13) Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. *Nat. Chem. Biol.* **2008**, *4*, 728−737.
(14) Spolar, R. S.; Record, M. T., Jr *Science* **1994**, *263*, 777−784.
(15) von Hippel, P. H. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 79.
(16) Vuzman, D.; Levy, Y. *Mol. BioSyst.* **2012**, *8*, 47−57.
(17) Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Rep. Prog. Phys.* **2012**, *75*, 076601.
(18) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2−8.
(19) Fuxreiter, M. *Mol. BioSyst.* **2012**, *8*, 168−177.
(20) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. *J. Mol. Recognit.* **2010**, *23*, 105−116.
(21) Bah, A.; Vernon, R. M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. *Nature* **2015**, *519*, 106−109.
(22) Niklas, K. J.; Bondos, S. E.; Dunker, A. K.; Newman, S. A. *Front. Cell Dev. Biol.* **2015**, *3*, 8.
(23) Suganuma, T.; Workman, J. L. *Annu. Rev. Biochem.* **2011**, *80*, 473−499.
(24) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381−395.
(25) Zentner, G. E.; Henikoff, S. *Nat. Struct. Mol. Biol.* **2013**, *20*, 259−266.
(26) Patel, D. J.; Wang, Z. *Annu. Rev. Biochem.* **2013**, *82*, 81−118.
(27) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41−45.
(28) Rothbart, S. B.; Strahl, B. D. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2014**, *1839*, 627−643.
(29) Kan, P.-Y.; Caterino, T. L.; Hayes, J. J. *Mol. Cell. Biol.* **2009**, *29*, 538−546.
(30) Pepenella, S.; Murphy, K. J.; Hayes, J. J. *J. Biol. Chem.* **2014**, *289*, 27342−27351.
(31) Narlikar, G. J.; Sundaramoorthy, R.; Owen-Hughes, T. *Cell* **2013**, *154*, 490−503.
(32) Hwang, W. L.; Deindl, S.; Harada, B. T.; Zhuang, X. *Nature* **2014**, *512*, 213−217.
(33) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025−1040.
(34) Allan, J.; Harborne, N.; Rau, D. C.; Gould, H. *J. Cell Biol.* **1982**, *93*, 285−297.
(35) Garcia-Ramirez, M.; Dong, F.; Ausio, J. *J. Biol. Chem.* **1992**, *267*, 19587−19595.
(36) Schwarz, P. M.; Felthauser, A.; Fletcher, T. M.; Hansen, J. C. *Biochemistry* **1996**, *35*, 4009−15.
(37) Garcia-Ramirez, M.; Rocchini, C.; Ausio, J. *J. Biol. Chem.* **1995**, *270*, 17923−17928.
(38) Tse, C.; Sera, T.; Wolffe, A. P.; Hansen, J. C. *Mol. Cell. Biol.* **1998**, *18*, 4629−4638.
(39) Robinson, P. J. J.; An, W.; Routh, A.; Martino, F.; Chapman, L.; Roeder, R. G.; Rhodes, D. *J. Mol. Biol.* **2008**, *381*, 816−25.
(40) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. *Sci. Signaling* **2006**, *311*, 844−847.
(41) Allahverdi, A.; Yang, R.; Korolev, N.; Fan, Y.; Davey, C. a.; Liu, C.-F.; Nordenskiöld, L. *Nucleic Acids Res.* **2011**, *39*, 1680−91.
(42) Richmond, T. J.; Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F. *Nature* **1997**, *389*, 251−260.
(43) Yang, D.; Arya, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2911−2921.
(44) Kalashnikova, A. A.; Porter-Goff, M. E.; Muthurajan, U. M.; Luger, K.; Hansen, J. C. *J. R. Soc., Interface* **2013**, *10*, 20121022−20121022.
(45) Potoyan, D. A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U. S. A* **2012**, *109*, 17857−62.
(46) Lins, R. D.; Röthlisberger, U. *J. Chem. Theory Comput.* **2006**, *2*, 246−250.
(47) Yang, D.; Arya, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2911−21.
(48) Korolev, N.; Yu, H.; Lyubartsev, A. P.; Nordenskiöld, L. *Biopolymers* **2014**, *101*, 1051−64.
(49) Dion, M. F.; Altschuler, S. J.; Wu, L. F.; Rando, O. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5501−5506.
(50) Dai, J.; Hyland, E. M.; Yuan, D. S.; Huang, H.; Bader, J. S.; Boeke, J. D. *Cell* **2008**, *134*, 1066−1078.
(51) Dorigo, B.; Schalch, T.; Bystricky, K.; Richmond, T. J. *J. Mol. Biol.* **2003**, *327*, 85−96.
(52) Morinière, J.; Rousseaux, S.; Steuerwald, U.; Soler-López, M.; Curtet, S.; Vitte, A.-L.; Govin, J.; Gaucher, J.; Sadoul, K.; Hart, D. J.; Krijgsveld, J.; Khochbin, S.; Müller, C. W.; Petosa, C. *Nature* **2009**, *461*, 664−668.
(53) Banères, J.-L.; Martin, A.; Parello, J. *J. Mol. Biol.* **1997**, *273*, 503−508.
(54) Wang, X.; Moore, S. C.; Laszckzak, M.; Ausió, J. *J. Biol. Chem.* **2000**, *275*, 35013−20.
(55) Case, D., Darden, T., Cheatham, T.III, Simmerling, C., Wang, J., Duke, R., Luo, R., Walker, R., Zhang, W., Merz, T. *Amber12 MD*; University of California: San Francisco, 2012.
(56) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(58) van Holde, K. *Chromatin*; Springer-Verlag; New York, 1988.
(59) Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. a.; Coon, J. J. *Proc. Natl. Acad. Sci. U. S. A* **2008**, *105*, 4093−4098.
(60) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(61) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−7.
(62) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Mol. Biol.* **1997**, *265*, 217−241.
(63) Ding, F.; Jha, R. K.; Dokholyan, N. V. *Structure* **2005**, *13*, 1047−1054.
(64) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(65) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312−2334.
(66) Echeverria, I.; Papoian, G. A. *Isr. J. Chem.* **2014**, *54*, 1293−1301.
(67) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8183−8188.
(68) Luan, B.; Lyle, N.; Pappu, R. V.; Raleigh, D. P. *Biochemistry* **2014**, *53*, 39−47.
(69) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. *Cell* **2012**, *149*, 214−231.
(70) Zhuravlev, P. I.; Papoian, G. A. *Q. Rev. Biophys.* **2010**, *43*, 295−332.

DOI: 10.1021/jacs.5b00235
J. Am. Chem. Soc. 2015, 137, 6245−6253