PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 2911–2921

www.rsc.org/pccp

PAPER

Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation†

Darren Yang and Gaurav Arya*

Received 13th August 2010, Accepted 15th November 2010
DOI: 10.1039/c0cp01487g

The H4 histone tail plays a critical role in chromatin folding and regulation—it mediates strong interactions with the acidic patch of proximal nucleosomes and its acetylation at lysine 16 (K16) leads to partial unfolding of chromatin. The molecular mechanism associated with the H4 tail/acidic patch interactions and its modulation *via* K16 acetylation remains unknown. Here we employ a combination of molecular dynamics simulations, molecular docking calculations, and free energy computations to investigate the structure of the H4 tail in solution, the binding of the H4 tail with the acidic patch, and the effects of K16 acetylation. The H4 tail exhibits a disordered configuration except in the region Ala15–Lys20, where it exhibits a strong propensity for an α-helical structure. This α-helical region is found to dock very favorably into the acidic patch groove of a nucleosome with a binding free energy of approximately -7 kcal mol^{-1}. We have identified the specific interactions that stabilize this binding as well as the associated energetics. The acetylation of K16 is found to reduce the α-helix forming propensity of the H4 tail and K16's accessibility for mediating external interactions. More importantly, K16 acetylation destabilizes the binding of the H4 tail at the acidic patch by mitigating specific salt bridges and longer-ranged electrostatic interactions mediated by K16. Our study thus provides new microscopic insights into the compaction of chromatin and its regulation *via* posttranslational modifications of histone tails, which could be of interest to chromatin biology, cancer, epigenetics, and drug design.

Introduction

The chromatin fiber represents the most important level of DNA organization in eukaryotic organisms.[1,2] It consists of a linear array of nucleosomes—146 bp of DNA wrapped in 1.7 turns around a histone octamer—separated by stretches of naked DNA called linker DNA. The histone octamer is formed from the assembly of two copies each of the four histones H2A, H2B, H3 and H4.[3,4] While a large portion of the histone chains are involved in the formation of the cylindrical core of the octamer, their terminal portions called histone tails project outwards into the solution. At physiological conditions, the linear array of nucleosomes is generally folded into a compact ~30 nm thick chromatin fiber[5,6] due to electrostatic screening from physiological salt, attractive interactions between nucleosomes mediated by the histone tails, and binding of the linker histone at the nucleosome dyad.[7] Apart from its role in compacting DNA, the chromatin fiber plays a central role in the regulation of DNA-related processes such as transcription, repair, replication, and recombination because of its direct control over the accessibility of DNA sequences.

Histone posttranslational modifications (PTMs) constitute a key mechanism by which the structure and biochemical activity of chromatin is regulated.[8–10] Specifically, residues on the histone core and tail regions are subjected to chemical modifications such as acetylation, methylation, phosphorylation, and ubiquitination that regulate various DNA-related processes. For example, acetylation of lysine residues on the histone tails correlates strongly with gene transcription while methylation of some lysine residues on the H3 tail correlates with gene repression. PTMs are believed to modulate chromatin through either direct or indirect mechanisms.[11,12] In the direct mechanism, the PTMs directly alter the protein/protein and protein/DNA interactions to modulate chromatin compaction and structure. In the indirect mechanism, the PTMs act as sites for the recruitment of enzymes like chromatin remodelers to alter chromatin morphology. It is also possible that some PTMs modulate chromatin through a combination of direct and indirect mechanisms.

Department of NanoEngineering, University of California at San Diego, 9500 Gilman Drive, MC 0448, La Jolla, CA 92093, USA.
E-mail: garya@ucsd.edu; Fax: +1 858-534-9553;
Tel: +1 858-822-5542

† Electronic supplementary information (ESI) available: Figures showing exhaustive sampling of the H4 tail by REMD simulations, appearance of the α-helical configuration of the H4 tail from an extended starting configuration, intermittent intramolecular H-bonding within the H4 tail, electrostatic map of the acidic patch, distribution of binding free energies, and free energy contributions of the acidic patch residues. See DOI: 10.1039/c0cp01487g

The focus of this study is on one very important histone PTM, the acetylation of lysine 16 on the H4 histone tail (H4-K16). H4-K16 acetylation, along with acetylation of other lysines on the H3 and H4 tails, is strongly linked to transcriptional activation[13] and is also known to play important roles in dosage compensation,[13] maintenance of euchromatin–heterochromatin barriers,[14] regulation of cellular lifespan,[15] and cancer progression.[16]

The connection between lysine acetylation and gene transcription has been known for decades from observations that the chromatin surrounding transcriptionally active genes tends to be enriched in acetylated lysines.[17–19] One expects that acetylation of lysine, which abolishes its positive charge, would weaken the electrostatic screening of DNA/DNA repulsion mediated by the histone tails,[20,21] leading to partial unfolding of chromatin.[22,23] However, that singly-acetylated H4-K16 could lead to drastic unfolding of chromatin was not expected. Specifically, Shogren-Knaak *et al.*[24] used a chemical ligation strategy to produce H4 histones with a singly acetylated K16 residue. Sedimentation analyses indicated drastic unfolding of 12-nucleosome arrays reconstituted with singly acetylated H4-K16 compared to native nucleosome arrays (the sedimentation coefficients dropped from 54S to 44S). Similar unfolding was also observed by Robinson *et al.* in longer arrays with >60 nucleosomes.[25] Interestingly, acetylation of 30% of the K16 residues inhibited chromatin folding to a larger degree than the complete removal of the H4 tail. That these studies were conducted *in vitro* suggests that H4-K16 acetylation might regulate gene transcription through a direct mechanism involving disruption of chromatin folding, though the experiments cannot rule out the possibility of other mechanisms also contributing to H4-K16 triggered gene transcription. Indeed, acetylation has been known to decrease the stability of the nucleosome[26] and to lead to increased fluctuations of the nucleosomal DNA near its entry/exit,[27] all of which could contribute to an overall increase in the accessibility of DNA for transcription.

How does acetylation of H4-K16 lead to such drastic unfolding of chromatin even though it results in a minor reduction in the H4 tail charge? Clearly, electrostatic screening effects, *i.e.*, reduced tendency of the tails to distribute in regions of negative electrostatic potential alone cannot cause the observed unfolding. Instead, we expect that changes in *specific* interactions mediated by the H4 tail result in the observed unfolding. Indeed, our low-resolution mesoscopic model of chromatin that accounts for electrostatic screening by the tails but not for atomic-level specific interactions between residues predicts that a $+1e$ reduction in the net charge of the H4 tail, mimicking the effect of H4-K16 acetylation, has a marginal effect on the folding of nucleosome arrays.[28] Further support comes from *in vitro* reconstituted nucleosome arrays that fail to fold to their fully compact state when the "native" H4 tails are chemically replaced by equivalent-length polypeptides having the same net charge but different residue sequence.[29]

Luger *et al.*[3] suggested that the H4 tails could interact strongly with the "acidic patch" of adjacent nucleosomes, which is composed of a cluster of seven negatively charged residues of the H2A/H2B dimer. Computational modeling[28,30] suggests that the H4 tails, due to their location on the nucleosome surface, are the most likely amongst all tails to mediate strong interactions with the DNA and acidic patch of *other* nucleosomes. Several studies have demonstrated the importance of this acidic patch in compacting chromatin. Zhou *et al.*[31] reconstituted nucleosome arrays with the histone variant H2A.Bbd,[32] which differs from native H2A in terms of three residues that constitute the acidic patch. The H2A.Bbd arrays were unable to fold to their maximally compact state obtained with H2A, suggesting the importance of the acidic patch in compaction. Other studies have implicated the acidic patch in the oligomerization of nucleosomal arrays.[31,33] The molecular details of the interactions between the H4 tail and the acidic patch remain unknown. Luger and Richmond[34] suggested a model for the configuration of the H4 tail bound to the acidic patch deduced from the crystalline arrangement of stacked nucleosome core particles. However, the feasibility as well as the detailed energetics and stability of this configuration have not been investigated so far.

In summary, the molecular mechanisms by which the H4 tail mediates strong internucleosomal interactions, through the putative nucleosomal acidic patch region, are not well understood. The mechanism by which H4-K16 acetylation triggers dramatic unfolding of chromatin, arguably through disruption of H4-mediated internucleosomal interactions, is even less understood. Here, we employ computational approaches to provide new, molecular-level insights into the solution-state structure of the H4 tail and its interactions with the nucleosomal acidic patch. We also examine how the structure and the interactions of the H4 tail get modulated with H4-K16 acetylation. Our computations reveal several intriguing findings. First, the H4 tails are not completely disordered as suggested previously, but have a strong propensity to form a short α-helical conformation about K16. Second, this α-helical configuration of the H4 tail can surprisingly bind very strongly at the acidic patch groove, more favorably than an extended configuration. Third, K16 acetylation reduces the stability of the α-helical configuration of the H4 tail and the accessibility of the K16 sidechain. Fourth, K16 acetylation disrupts the interactions of the H4 tail with the acidic patch. The above findings along with detailed analyses of residue-level interactions and energetics provide a plausible mechanism for H4 tail mediated chromatin compaction and K16 acetylation triggered unfolding of chromatin that could have important implications in chromatin biology, epigenetics, and drug discovery efforts.[35]

Computational methodology

Replica-exchange molecular dynamics (REMD) simulations of the H4 tail

We use AMBER 10[36] to carry out implicit solvent REMD simulations of (i) wild type H4 histone tail (H4-WT) and (ii) H4 tail containing the acetylated K16 (H4-AcK). The H4 tail fragment consists of residues 1–25[4] with the amino acid sequence SGRGK GGKGL GKGGA **X**RHRK VLRDN, where **X** is a regular lysine or an acetylated lysine. The AMBER ff03 force field[37] is employed. Since this force field

does not provide parameters for acetylated lysines, we adopt the force field parameters for acetylated lysine derived recently by Liu and Duan[38] using a similar approach to that used in the development of AMBER ff03. Solvation effects are treated using the generalized Born model of Onufriev *et al.*[39] in conjunction with the surface area model of Weiser *et al.*,[40] where the dielectric constants of the protein and the solvent are set to 1.0 and 78.5, respectively. This particular combination of force field and implicit solvent model with the above dielectric constants has been applied to many proteins and shown to capture well their folding transitions and configurations.[41,42] Furthermore, the ionic radii are set to mbondi2 and surface tension was set to 0.005 kcal mol^{-1} $Å^{-2}$ in all simulations. The monovalent salt concentration is fixed at 0.2 M to represent physiological conditions. Both H4-WT and H4-AcK simulations are started from fully extended conformations of the H4 tail prepared using the tleap module of AMBER 10. The extended peptides are energy minimized through 500 steps of steepest descent followed by 500 steps of conjugate gradient.

We choose REMD simulations over MD simulations to predict the structure of H4-WT and H4-AcK because of their superior capability to sample the configurational space. In REMD, multiple copies of the H4 tail system are simulated at different temperatures and attempts are made routinely to swap the configurations of two systems at different temperatures *via* a Monte-Carlo procedure. We employ the multi-sander module of the AMBER 10 package to perform REMD simulations. To achieve an exchange ratio of about 0.2, each simulation system consists of 12 replicas simulated at the following temperatures: 250.0, 265.7, 282.3, 300.0, 318.8, 338.7, 360.0, 382.4, 406.3, 431.8, 458.8, and 487.5 K. Each replica is first independently equilibrated for 2 ns at the target temperature without any system swaps and then configuration swaps are routinely attempted every 2 ps. Fig. S1 (ESI†) shows rapid equilibration of the total energies of the 12 replicas. The temperature is maintained constant using Berendsen's weak coupling algorithm[43] with a coupling constant of 2.0 ps. To avoid instability at high temperatures, the time step is fixed at 1 fs and the SHAKE algorithm[44] is employed to constrain the bonds connecting hydrogen atoms. The cutoff for non-bonded interactions and GB pairwise summation is set to 24 Å. The energies and conformations are saved every 2 and 4 ps, respectively. The REMD simulations are performed for 60 ns and only the energies and trajectories from the final 50 ns are used for the analyses. Fig. S2 (ESI†) shows the potential energy distributions computed from the first 10 ns (equilibration) of the REMD simulations. The overlapping energy distributions, especially between neighboring replicas, suggest that configuration swap is indeed possible across neighboring replicas.

Molecular dynamics (MD) simulations of the H4 tail

Explicit-solvent MD simulations of H4-WT and H4-AcK are carried out using AMBER 10 in conjunction with AMBER ff03 force field to confirm the stability of the H4 tail conformations obtained from implicit-solvent REMD simulations. A fully folded α-helical configuration is chosen as the initial configuration of the H4 tail for MD simulations. A simulation box of dimensions 7 × 7 × 7 nm is employed with periodic boundary conditions. The tails are first solvated with TIP3P water molecules using the tleap module of AMBER 10. Na^+ and Cl^- are added to maintain electroneutrality and a physiological salt concentration of 0.2 M. The particle mesh Ewald approach[45] is used to model electrostatic interactions. The resulting configuration is energy minimized using 500 steps of steepest descent followed by slow heating from 50 K to 300 K over 50 ps. The solvent is allowed to equilibrate for an additional 950 ps in the NPT ensemble. During energy minimization and heating, the entire tail is restrained using a soft harmonic potential with a spring constant of 2.0 kcal mol^{-1} A^{-2}. The same restrained potential applied only to the H4 tail backbone is also employed during the equilibration process. After heating and equilibration, we perform a 25 ns long production run under NPT conditions without the backbone restraint. Note that because explicit-solvent MD simulations are more computationally demanding than implicit-solvent REMD simulations, the MD simulations could not be performed for as long a time as the REMD simulations. The temperature is maintained constant at 300 K during equilibration and production MD using a Berendsen thermostat with a coupling time constant of 2 ps. The pressure is maintained at 1 bar using the position scaling thermostat with a relaxation time of 3 ps. SHAKE constraints are applied to all bonds involving hydrogen atoms and the cutoff for non-bonded interactions is set to 12 Å. Energies and conformations are saved every 2 ps and 4 ps, respectively.

Surface area and secondary structure determination of the H4 tail

The energies and trajectories of the H4-WT and H4-AcK are collected from the REMD simulation for each replica and analyzed for the accessible surface area of K16 and tail's secondary structure at physiological temperature 300 K only. The solvent-accessible surface area (SASA) is calculated using VMD 1.8.6,[46] which uses the SURF module developed by Varshney and Brooks.[47] SASA is defined as the surface of a target molecule accessible to an exterior spherical probe as it is rolled over the target's van der Waals surfaces. The radius of the spherical probe is fixed to 1.4 Å to mimic the rough size of a water molecule. We employ do_dssp and ptraj modules within GROMACS 4.0[48] and AMBER 10, respectively, to carry out secondary structure analyses of the H4 tail trajectories. Both modules employ the defined secondary structure of proteins (DSSP) method developed by Kabsch and Sander.[49] Secondary structure predictions for the H4 tail are carried out using the PredictProtein server (www.predictprotein.org), which uses four different methods (APSSP2, Jpred, PROF, and SABLE2) for the prediction and the AGADIR program.[50]

Molecular docking of the H4 tail fragment on the nucleosome

AutoDock Vina 1.0[51] is used to determine the most favorable configuration of the H4 tail bound (ligand) at the nucleosomal acidic patch (receptor). AutoDock Vina 1.0 is about two orders of magnitude faster in predicting binding sites compared to its predecessor (AutoDock 4.2) and also significantly improves on the accuracy of binding by allowing up to

32 rotatable bonds. As the receptor, we consider the entire 1KX5 nucleosome structure of 1.9 Å resolution.[4] For the ligand, we consider an H4 tail fragment comprising residues 16–23. We do not attempt to dock the complete H4 tail but instead choose a eight residue long segment of the H4 tail for docking because of limitations on the number of torsional angles allowed to be flexible in the docking simulations ($\leq$32). We chose the above particular segment for docking because upstream segments (residues 1–14) seem to be less important for binding to the acidic patch than the remaining tail residues, as shown recently through sedimentation coefficient measurements of nucleosome arrays with varying excised regions of the H4 tails.[52] Also, we are primarily interested in the ability of the α-helical region of the H4 tail to bind within the acidic patch groove. To investigate the effect of K16 acetylation and the α-helical configuration, we carry out four independent dockings: wild type (H4-WT) and acetylated-K16 H4 tail (H4-AcK) fragments, with and without an α-helical constraint. Note that the α-helical constraint is placed only on the backbone of residues K16-K20, as this region exhibits the largest α-helical propensity in our simulations and secondary structure predictions.

Before docking, the N and C termini of the H4 tail fragment are blocked using ACE and NME groups, respectively, to avoid undue interactions between the charged termini and the nucleosome. The nucleosome structure is also refined by excluding water molecules and adding the missing hydrogen atoms using the tleap module in AMBER 10. Keeping all heavy atoms of the proteins and nucleotides constrained, a short 500-step steepest descent energy minimization is carried out to remove steric clashes resulting from the addition of the missing hydrogen atoms. Next, non-polar hydrogen atoms are removed using AutoDockTools.[53] For each ligand type, we carry out 1000 independent docking calculations with different initial seeds, within a box of dimension 28 × 28 × 28 Å enclosing the acidic patch. All parameters are set to default, except the exhaustiveness value, which is increased to a value of 128 to allow thorough sampling of the docked configurations. The exhaustiveness value is linearly related to the conformation search time; hence increasing this value leads to an exponential increase in the probability of finding the energy minima. After several iterations, the algorithm produces 1000 low-energy configurations, which are further categorized into clusters according to a structure-based criterion, *i.e.*, the root mean square deviation of all configurations within a cluster should be less than 3.0 Å.

Molecular dynamics (MD) simulations of the H4 tail bound to the acidic patch

Explicit-solvent MD simulations of the H4 tail fragment (residues 16–23) bound at the acidic patch of a nucleosome are conducted using AMBER 10 with AMBER ff03 force field to both test the stability of the bound H4 tail fragment and generate configurations of the complex for the binding free energy calculations. The histone tails of the nucleosome are removed to reduce the system size. We examine only the most favorable configurations of bound H4-WT and H4-AcK from docking, *i.e.*, the lowest-energy configuration from the most populated cluster. The MD simulations are performed at similar conditions and using a similar protocol as the explicit-solvent MD simulations of the complete, isolated H4 tail. The only differences are that the simulation box was larger (15 × 15 × 11 nm), slightly different harmonic constraints (3.0 kcal mol^{-1} A^{-2}) are imposed on the heavy atoms for a longer period during density equilibration under NPT conditions (150 ps), and a slightly longer equilibration step without these constraints under NVT conditions is employed (1 ns). We employ a production step of 4 ns and snapshots of the system are recorded every 10 ps.

Binding free energy calculations

We employ the Molecular Mechanics-Poisson–Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) methods to compute the binding free energy (ΔG_{bind}) of the H4-WT and H4-AcK fragments docked at the nucleosomal acidic patch. According to the two methods ΔG_{bind} is computed as $\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$, where ΔE_{MM} is the change in the molecular mechanics energy upon binding that consists of contributions from van der Waals (ΔE_{vdW}), electrostatic (ΔE_{ele}), and internal, bonded interactions (ΔE_{int}). ΔG_{sol} is the change in the solvation free energy upon binding that is computed using mean field approaches as a sum of polar ($\Delta G_{sol\text{-}pol}$) and non-polar contributions ($\Delta G_{sol\text{-}np}$). $\Delta G_{sol\text{-}pol}$ is computed using either the Poisson–Boltzmann or generalized Born approach and $\Delta G_{sol\text{-}np}$ is computed using the solvent accessible surface area (SASA) as $\gamma\Delta$SASA, where γ is the surface tension. ΔS is the change in the configurational entropy of the ligand and receptor upon binding (T is the temperature). The net polar and non-polar contributions to ΔG_{bind} are then given by $\Delta G_{polar} = \Delta G_{sol\text{-}pol} + \Delta E_{ele}$ and $\Delta G_{np} = \Delta G_{sol\text{-}np} + \Delta E_{vdW}$, respectively.

ΔE_{MM} and ΔG_{sol} and its various contributions are computed from 400 snapshots of the H4 tail/acidic patch complex recorded during the 4 ns MD simulation of the complex using the mm_pbsa.pl script within AMBER 10. We also compute contributions from individual H4 tail and acidic patch residues to ΔE_{MM} and ΔG_{sol} using the same script. Note that only MM-GBSA can provide such residue-level contributions, which is not possible using MM-PBSA. The solute and solvent dielectric constants are set to 1 and 80, respectively. The water probe radius and the surface tension are set to 1.4 Å and 0.0072 kcal mol^{-1} Å^{-2}, respectively. We do not compute the entropic contribution ($T\Delta S$) due to its extremely slow convergence for our large system size.

Results

H4 histone tail exhibits propensity for an α-helical structure in the region A15–K20

Our implicit-solvent REMD simulations of the wild type H4 tail (H4-WT) starting from an unfolded configuration provide its equilibrium structure. Fig. 1A shows the computed secondary structure of H4-WT during the 50 ns production run, following the ~12 ns long equilibration step (Fig. S3a, ESI†). While a majority of the histone tail remains unstructured, the region

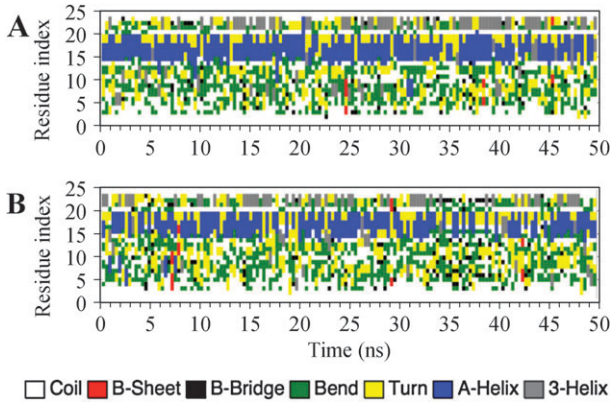

Fig. 1 Time evolution of the secondary structures of the wild type (A) and acetylated-K16 H4 tail (B), computed from 50 ns-long implicit-solvent REMD simulations using the DSSP method with a 300 ps sampling. The blue regions denote observed dihedral angles consistent with α-helical structures.

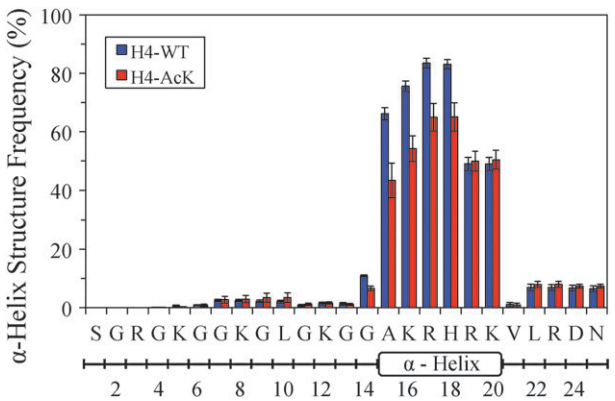

Fig. 2 Frequency of the α-helical structure of each residue in H4-WT and H4-AcK obtained from the implicit-solvent REMD simulations. The error bars show standard error computed from block averages. The blue and red bars indicate wild type and acetylated H4 tail. The rectangular bar shows the most frequently observed location of the α-helix along the H4 tail sequence.

encompassing residues A15 to K20 interestingly exhibits an α-helical structure. We have further quantified this α-helical propensity by computing the fraction of time during the simulation that each residue exhibits dihedral angles ϕ and ψ consistent with that of the α-helix, which are summarized in Fig. 2. All residues in the range A15–K20 exhibit α-helical dihedral angles for a majority of the time during the simulation, with K16, R17, and H18 exhibiting the largest α-helical propensity.

To confirm that the observed α-helical tendency of the H4 tail region is not an artifact of the implicit-solvent model used in the REMD simulations, we have also carried out exhaustive 25 ns explicit-solvent MD simulations of H4-WT. Since it is not possible to probe with explicit solvent MD simulations the entire folding process starting from the extended structure, we start the simulation with a fully folded α-helix. If any α-helical region is unstable, it would dissociate within the simulation time. The secondary structure of H4-WT is analyzed as a function of time (Fig. S4a, ESI†). The α-helical structure

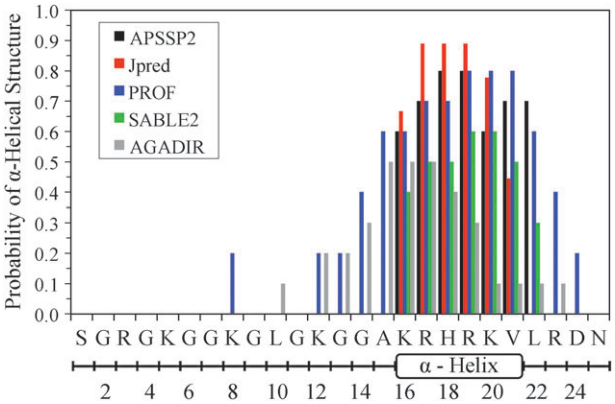

Fig. 3 Predictions for the α-helical propensity of the H4 tail obtained from five sequence-based secondary structure prediction methods. The five methods employed are listed in the figure legend. The rectangular bar shows the predicted location of the α-helix along the H4 tail sequence.

within residues S1-L10 disassociates and becomes disordered within a few ns, while the rest of the residues maintain an α-helical structure throughout the simulation, thus confirming our simulation results.

We have also utilized the sequence-based secondary structure prediction calculations to determine the α-helical propensity of H4-WT. The predictions from five different algorithms are summarized in Fig. 3. All five methods suggest that the region K16–L22 has a high propensity to yield an α-helical conformation, consistent with our simulation results.

K16 acetylation reduces the α-helical propensity of the H4 tail and reduces the side chain's accessibility to mediate external interactions

Fig. 1B shows the observed secondary structure of acetylated H4 tail (H4-AcK) during the 50 ns production step of the REMD simulation following the equilibration step (Fig. S3a, ESI†). Fig. 2 plots the fraction of time that each residue exhibits α-helical dihedral angles. As in the case of H4-WT, the tail remains disordered in all regions except the A15–K20 region where it exhibits an α-helical structure. However, compared to H4-WT, there is a marked reduction in the propensity of the A15–H18 region to exhibit an α-helical structure; the α-helical frequency decreases $>20\%$ for some residues. We have also confirmed the observed destabilization in the α-helical structure using explicit-solvent MD simulations (Fig. S4b, ESI†). The destabilization is in fact more pronounced in explicit solvent simulations. The secondary structure prediction algorithms used earlier for H4-WT cannot be applied to H4-AcK as the algorithms do not apply to modified residues such as acetylated lysines.

Our simulations indicate that the K16 acetylation-induced disruption of the α-helix occurs due to two main factors. First, the acetylation of K16 increases its tendency to interact with other residues and the backbone, contributing to some destabilization of the α-helix. In particular, the C═O group of acetylated K16 side chain is seen forming intermittent hydrogen bonds with the N–H group of R17, R19, and K20

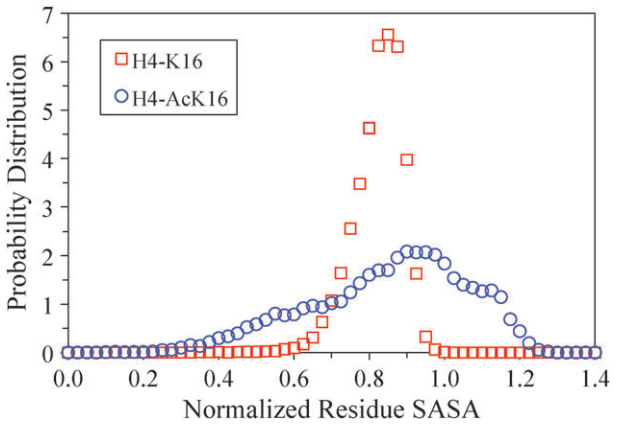

Fig. 4 Probability distribution of the solvent accessible surface area (SASA), normalized by the maximum SASA of the K16 lysine. The blue and red curves represent values computed from wild type and acetylated H4 tails. The probability distribution is normalized such that its integral over the normalized residue SASA is equal to one.

side chains in ~4% of H4-AcK configurations obtained from the REMD simulation. Fig. S5 (ESI†) presents two representative snapshots of the tail showing intermittent hydrogen bonding. The wild-type K16 side chain on the other hand exhibits *no* such interactions. Second, acetylation decreases the solvation of the K16 side chain, or in other words, increases its hydrophobicity. The reason is that the side chain no longer carries the charged amino group, which favors formation of hydration shells.

The above effects are supported by our calculations of the solvent accessible surface area (SASA) of the wild type K16 sidechain and the acetylated K16 sidechain from the H4-WT and H4-AcK REMD simulations, respectively. Fig. 4 plots the probability distribution of SASA values computed for the two types of sidechains. Both distributions have been normalized with respect to the maximum value of SASA obtained for the native K16 sidechain, *i.e.*, the *x*-axis represents $SASA_i/max(SASA_{K16})$, where *i* represents K16 or AcK16. The *y*-axis represents the probability distribution of computed SASA values, normalized in the standard way such that the area under the distribution is unity. The SASA distribution for wild type K16 is sharply peaked with an average value that is 88% of the maximum possible value, suggesting that the K16 sidechain remains highly exposed and rarely makes close contacts with neighboring residues. This is expected given the strongly α-helical nature of the H4 tail backbone between residues A15 to K20 and the strong solvation of the charged amino group of K16. One might expect that the addition of a $CO–CH_3$ group associated with acetylation of K16, which in its native state possesses five groups ($4CH_2$ and NH_3^+), should increase the SASA by ~30%. Instead, the average SASA remains nearly constant, suggesting that K16 becomes less accessible with acetylation. Moreover, the SASA distribution becomes significantly wider, especially towards the smaller SASA values. This decreased accessibility of the K16 side chain could potentially compromise the ability of K16 to interact with external molecules, *e.g.*, the acidic patch of neighboring nucleosomes.

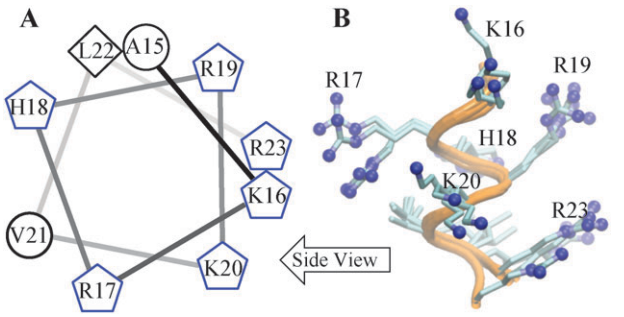

Fig. 5 (A) Helical wheel plot of the H4 tail K16–R23 region illustrating how the α-helical conformation facilitates the orientation of four basic residues in one common direction. The basic residues are shown as blue pentagons. (B) Four snapshots of the H4 tail fragment K16–R23 picked from the REMD simulation superimposed on each other and viewed along the arrow shown in (A) (side view).

α-helical H4 tail configuration uniquely exposes its lysine residues at one face

Several studies have shown that residues K16–R23 of the H4 tail play a crucial role in chromatin compaction. This region of the H4 tail is believed to bind with an acidic patch on the surface of another nucleosome.[34] The proposed binding modes for such an interaction suggest that the side chains of K16, R19, K20, and R23 make multiple hydrogen bonds and salt bridges with the acidic patch residues. Does the formation of the α-helical structure in the region A15–K20, as observed in the current study, preclude the possibility of such interactions?

To first test this at a conceptual level, we have constructed a helical wheel representation of the A15–R23 region to ascertain the orientation of K16, R19, K20 and R23 side chains. Intriguingly, all four residues orient in a common direction (Fig. 5A). Thus, the α-helical configuration of the H4 tail facilitates the orientation of these four residues in directions optimal for potentially interacting with the acidic patch. Next, we have analyzed the actual orientations of the four residues from our H4-WT REMD simulations. Fig. 5B shows four superimposed snapshots of the H4 tail backbone and its residues taken in 12 ns intervals over a 50 ns simulation. Clearly, all four charged groups distribute themselves along one face of the α-helix, thus confirming the results of the helix wheel representation.

H4 tail binds strongly with the acidic patch in the α-helical configuration

Having shown that the α-helical configuration of the H4 tail between residues 16 and 23 exposes its key basic residues along a common direction, we next ask the question: can such a configuration of the H4 tail even accommodate within the acidic patch groove, and if so, can it mediate strong interactions with it?

To test potential binding conformations of the α-helical H4 tail, we have carried out docking calculations of an α-helically constrained H4 tail (H4-WT) fragment onto the acidic patch. Fig. 6A–C shows the lowest-energy configurations corresponding to the three most-populated clusters along with their computed binding free energies (Fig. 6D). That the docked configurations can be categorized into a few, well-defined

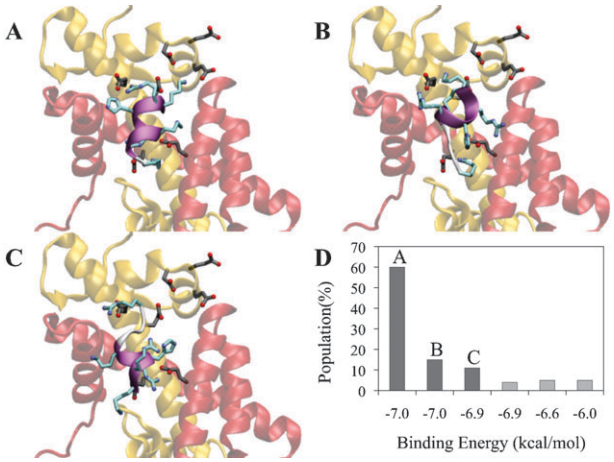

Fig. 6 (A–C) The lowest-energy configurations of the top three most-populated clusters obtained from docking of the wild type H4 tail (H4-WT) fragment at the acidic patch. The H4 tail is shown as purple ribbon and the H2A and H2B histones as yellow and red ribbons, respectively. The basic residues are shown in cyan and the acidic residues in grey. (D) Percentage population of the top six clusters obtained from analyzing results of docking. The lowest-energy member of the cluster (A) represents the most likely binding mode.

clusters is indicative of exhaustive sampling. The top cluster with a population of 60% is the most representative bound configuration. It has a binding free energy (ΔG) of -7.0 kcal mol^{-1}. The next two clusters have populations of 15% and 11% with ΔGs of -7.0 and -6.9 kcal mol^{-1}, respectively. The H4 tail fragment corresponding to the most populated cluster fits "snugly" within the acidic patch groove (Fig. S6, ESI†). The α-helical axis assumes an almost parallel orientation relative to the nucleosome surface and an angle of roughly 10° with respect to the nucleosome dyad axis, with its N-termini pointing towards the northern direction. Interestingly, the tail fragment becomes α-helical over its entire length during the simulation even though only five residues are constrained to be α-helical (Fig. 6A). We also carry out explicit-solvent MD simulations of this most favorable docked configuration and find it to be highly stable. Our results thus suggest that the H4 tail segment can indeed dock very favorably at the acidic patch in an α-helical configuration.

Analysis of docking results for the unconstrained H4 tail segment (without the α-helical constraint) suggests that the low-energy configurations are highly scattered and do not converge to any particular cluster(s), likely due to the greater number of flexible torsional angles. Hence, a direct comparison of the most favorable docked configuration of constrained *versus* unconstrained H4 tails is not possible. However, a rough comparison can be made by analyzing the ΔG distribution for the entire population of α-helical and unconstrained H4 tail fragment configurations obtained from the docking (Fig. S7, ESI†). Docking of the α-helical fragments yields a Gaussian energy distribution that shifts in the positive direction by roughly $+1$ kcal mol^{-1} when the backbone is left unconstrained, suggesting that the α-helical H4 tail segment likely binds more strongly than the unconstrained H4 tail fragment.

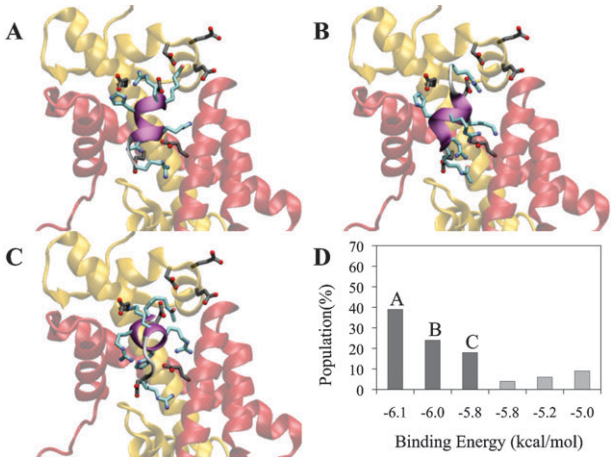

Fig. 7 Same as Fig. 6, except that the docked H4 tail fragment contains acetylated K16 (H4-AcK).

K16 acetylation destabilizes H4 tail/acidic patch binding

Docking calculations for the α-helically constrained H4-AcK tail fragment yield one highly populated cluster and several less-populated clusters while that for the unconstrained H4-AcK tail fragment do not converge to any particular cluster. Fig. 7A–C shows the most favorable configuration within each of the top three clusters. The most populated cluster (Fig. 7A) interestingly exhibits a similar binding conformation and orientation as the most populated cluster of the H4-WT tail fragment (Fig. 6A). However, K16 acetylation is observed to destabilize the binding as noted from the less favorable binding energy (-6.1 kcal mol^{-1} for H4-Ack *versus* -7.0 kcal mol^{-1} for H4-WT) and smaller population of the most populated cluster ($\sim$40% for H4-AcK *versus* $\sim$60% for H4-WT). We have also carried out exhaustive explicit-solvent MD simulations of the most favorable bound configuration of the H4 tail fragment at the acidic patch of the nucleosome to confirm its stability.

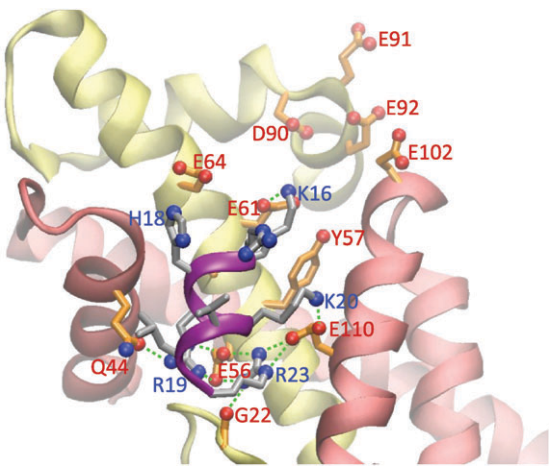

Fig. 8 The most favorable configuration of the H4-WT fragment (helix backbone shown in purple) docked at the H2A (yellow)–H2B (red) acidic patch illustrating the detailed residues and interactions involved in the binding. The green dashed lines indicate hydrogen bonds observed during the MD simulation. The image is produced by the VMD package.

Detailed interactions of the H4-WT and H4-AcK tails with the acidic patch

Fig. 8 shows a representative snapshot of the H4-WT tail bound at the acidic patch groove captured from an exhaustive 5 ns long explicit-solvent MD simulation started from the most favorable docked configuration shown in Fig. 6A. The figure details the most important electrostatic and hydrogen bonding interactions involved in the binding. The hydrogen-bonding interactions are identified by measuring the distances of these interacting residues. Most notably, the four basic residues K16, R19, K20, and R23 facing a common direction in the isolated H4 tail of Fig. 5 interact strongly with acidic patch residues E56, E61, E64, D90, E92, and E102 of histone H2A and Q44 and E110 of histone H2B. In particular, K16 interacts very strongly with E61 through salt bridges and also mediates fairly strong electrostatic interactions with residues E64, D90, E92, and E102. R19 forms hydrogen bonds with the OX group of Q44 and forms salt bridges with E56. K20 interacts very strongly with E110 through salt bridges. R23 interacts strongly with E110 and E56 through salt bridges.

Other H4 tail residues, R17 and H18, also contribute to the binding to some extent: R17 interacts intermittently with E64 through electrostatic interactions and H18 also interacts with E64 through van der Waals and hydrogen bonding interactions. V21 and L22 do not contribute to H4 docking as they face away from the nucleosomal surface. Interestingly, there appears to be a chain of hydrogen bonding interactions: K20 with E110, E110 with R23, R23 with E56, and E56 with R19. We do not currently know the significance of this interaction network but we expect that they strengthen the binding, as these interactions are likely cooperative. That is, the formation of one interaction promotes the formation of others, as the subsequent interactions no longer suffer from the significant translational entropy loss accompanying the formation of the first few interactions.

As noted earlier, K16 acetylation does not change the global configuration of the bound H4 tail but it has three effects on the aforementioned interactions. First, the acetylated K16 no longer interacts strongly with E61, D90, E92, and E102, causing it to occasionally come off the H2A/H2B surface. Second, the strong salt-bridges observed between R23 and E56 become significantly weaker. Third, K20 interactions between E110 also reduce, likely due to its interactions with the acetylated K16. The energetic impact of these effects is discussed below.

Detailed energetics of H4-WT and H4-AcK binding

Though the docking calculations help identify the most favorable binding modes for H4-WT and H4-AcK, they cannot provide detailed energetics of the bound configurations. To this end, we carry out MM-PBSA and MM-GBSA calculations on the most favorable α-helical H4-WT and H4-AcK configurations obtained from docking. These methods allow us to more accurately compute the binding free energy ΔG_{bind} and its polar ($\Delta G_{\mathrm{polar}}$) and non-polar contributions (ΔG_{np}). However, as discussed earlier, we report only contributions to ΔG_{bind} from molecular mechanics (ΔE_{MM}) and solvation (ΔG_{sol}); the contribution from conformational entropy ($-T\Delta S$) is not reported due to the extremely computationally demanding nature of normal mode calculations on ligands and receptors as large as the H4 tail and the nucleosome. Nonetheless, we find that ΔE_{MM} and ΔG_{sol} can provide valuable insights into the binding driving forces. More importantly, the method allows us to compute contributions to the binding from individual tail and acidic patch residues, which does not require normal mode entropy calculations.

Table 1 Binding free energies and its various components computed from MM-PBSA and MM-GBSA. WT and AcK denote wild type and K16-acetylated H4 tails, respectively. σ denotes the uncertainties in the computed energies for the two types of tails. All values are given in kcal mol^{-1}

Contribution	WT	σ-WT	AcK	σ-AcK
ΔE_{ele}	−6496.4	50.2	−5064.7	94.2
ΔE_{vdw}	−44.3	5.2	−43.3	4.6
ΔE_{int}	0	0	0	0
ΔE_{MM}	−6540.7	49.1	−5108.1	95.4
$\Delta G_{\mathrm{sol\text{-}np}}$	−8.6	0.5	−7.5	0.5
$\Delta G_{\mathrm{sol\text{-}pol,PB}}$	6439.8	45.8	5031.5	92.0
$\Delta G_{\mathrm{sol,PB}}$	6431.2	45.9	5024.0	91.7
$\Delta G_{\mathrm{ele,PB}}$	−56.6	13.2	−33.3	10.2
$\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{sol,PB}}$	−109.5	10.0	−84.1	9.2
$\Delta G_{\mathrm{sol\text{-}pol,GB}}$	6432.3	47.7	5027.6	89.9
$\Delta G_{\mathrm{sol,GB}}$	6423.7	47.9	5020.0	89.6
$\Delta G_{\mathrm{ele,GB}}$	−64.0	7.2	−37.2	7.5
$\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{sol,GB}}$	−117.0	4.5	−88.0	7.9

The computed energies and their breakdown are presented in Table 1. Both MM-PBSA and MM-GBSA methods yield similar trends in the energies though the absolute energy values differ slightly between the two methods. The most important finding is that the H4-WT fragment binds much more strongly at the acidic patch compared to H4-AcK. Specifically, the MM-PBSA and MM-GBSA methods, respectively, predict binding energies $\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{sol}}$ of approximately −109.5 kcal mol^{-1} and −117.0 kcal mol^{-1} for H4-WT compared to −84.1 kcal mol^{-1} and −88.0 kcal mol^{-1} for H4-AcK. These values yield $\Delta\Delta E_{\mathrm{MM}} + \Delta\Delta G_{\mathrm{sol}}$ of +25.6 kcal mol^{-1} and +29.0 kcal mol^{-1}, which quantify the destabilization of H4 tail binding from K16 acetylation. Note that the energy values are very large because they do not include the unfavorable contribution from configurational entropy. We expect that the entropy contribution will significantly reduce the energies through the so-called binding enthalpy–entropy compensation effect,[54] but we do not expect it to overturn the energy values (*i.e.*, make the energies positive). Additionally, the main source of the difference between the binding energies of H4-WT and H4-AcK is the polar contribution ($\Delta G_{\mathrm{polar}}$ = −56.6 kcal mol^{-1} for H4-WT *versus* −33.3 kcal mol^{-1} for H4-AcK, *i.e.*, $\Delta\Delta G_{\mathrm{polar}}$ = +23.3 kcal mol^{-1}).

We have further decomposed the energies ($\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{sol}}$) on a residue-level basis $\Delta G_{\mathrm{bind,i}} = \Delta E_{\mathrm{MM,i}} + \Delta G_{\mathrm{sol,i}}$, which is further broken down into polar ($\Delta G_{\mathrm{pol,i}} \equiv \Delta E_{\mathrm{ele,i}} + \Delta G_{\mathrm{sol\text{-}pol,i}}$) and non-polar contributions ($\Delta G_{\mathrm{np,i}} \equiv \Delta E_{\mathrm{vdW,i}} + \Delta G_{\mathrm{sol\text{-}np,i}}$). $\Delta G_{\mathrm{bind,i}}$ and its two contributions $\Delta G_{\mathrm{pol,i}}$, and $\Delta G_{\mathrm{np,i}}$ for each H4-WT and H4-AcK residue are plotted in Fig. 9A–C. We note four interesting findings. First, the main tail residues contributing to its binding are K16, K19, and R23, each of which contributes at least −13 kcal mol^{-1}; K20 and R17 contribute to some extent (roughly −7 kcal mol^{-1}) (Fig. 9A). Second,

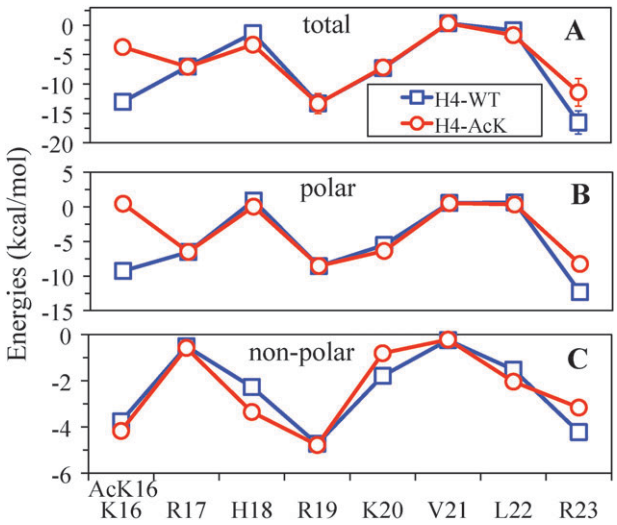

Fig. 9 (A) Contributions of individual H4 tail residues to the net binding free energy ($\Delta E_{MM,i} + \Delta G_{sol,i}$), as obtained from MM-GBSA, and their decomposition into (B) polar ($\Delta E_{ele,i} + \Delta G_{sol\text{-}pol,i}$) and (C) non-polar energies ($\Delta E_{vdw,i} + \Delta G_{sol\text{-}np,i}$). The blue squares and red circles represent energies computed from H4-WT and H4-AcK simulations.

~60–70% of these favorable energies arise from polar contributions (Fig. 9B and C). Third, the acetylation of K16 dramatically reduces the binding capacity of K16 from −13.0 kcal mol^{-1} to −3.7 kcal mol^{-1}, *i.e.*, an increase of ~9 kcal mol^{-1}. Fourth, K16 acetylation does not affect significantly the binding energy of the remaining residues.

We have also plotted the individual residue contributions to the binding energy $\Delta G_{bind,i}$ on the acidic patch side in Fig. S8 (ESI†). E56 with an energy contribution of −11.8 kcal mol^{-1} is the biggest contributor to the binding. The next biggest contribution is from E22 (−3.1 kcal mol^{-1}). Upon K16 acetylation, we find that these two energy contributions decrease to −5.5 kcal mol^{-1} and −1.8 kcal mol^{-1}, respectively.

Discussion

Our study uncovers four major findings related to the molecular mechanisms by the H4 histone tail and its acetylation at K16 compacts and modulates chromatin structure.

The *first* finding is that a stretch of the H4 tail (residues 16–20) has a strong propensity to adopt an α-helical structure, which is demonstrated using three separate approaches: exhaustive implicit-solvent MD simulations using a solvent model-force field combination that minimises secondary structure biases; exhaustive MD simulations with explicit solvent; and five secondary structure prediction algorithms. The observed α-helical propensity also has good support from the literature.[55,56] Specifically, circular dichroism measurements by the Ausio group[56] indicate the presence of a 4–5 residues long stretch within the H4 tail that exhibits high propensity for the α-helical structure, though its exact location cannot be determined. Recent MD simulations by another group[55] also indicate that the H4 tail could form an α-helix between residues 14 and 22. Interestingly, the H3 tails exhibit an even stronger α-helical tendency than that observed for the H4 tails in this study.[38]

The *second* finding is that the H4 tail in its α-helical conformation can bind very strongly to the acidic-patch. Indeed, such a conformation allows most of the basic residues of the tail to orient in one common direction (Fig. 5), facilitating cooperative interactions with the residues of the acidic patch. Thus, the almost periodic positioning of the H4 tail lysines (with a periodicity of ~3–4 residues) may not be a coincidence but a deliberate pattern to facilitate their unidirectional exposure.[57] The shift in the binding energy distributions towards smaller values for unconstrained H4 tails (Fig. S7, ESI†) indicates that the α-helical H4 tail binds stronger than a random coil conformation, though computational limitations in docking fully random tail fragments prevent a conclusive result. Our docking calculations provide a rough estimate of the strength of H4 tail/acidic patch binding (−7 kcal mol^{-1}), which is in good agreement with the strength of internucleosomal interactions (−6–8.5 kcal mol^{-1}) deduced from single-molecule force measurements of chromatin.[58]

More detailed MD simulations and free energy calculations have identified the residues involved in these interactions and also dissected their energetic contributions. We find that salt-bridges between H4 tail residues (K16, K19, and R23) and acidic patch residues (E56, E61, E64, D90, E92, E102, Q44 and E110) are the most responsible for this binding. We also note a network of hydrogen bonds between some tail and acidic patch residues that could additionally stabilize the binding. Thus, rather than posing steric hindrance for the accommodation of the H4 α-helix, the acidic patch groove instead provides a glove-like fit for the α-helix to promote strong interactions with the α-helix (Fig. S6, ESI†).

It is instructive to compare this pattern of residue interactions (Fig. 9) with that proposed by Luger and Richmond, who considered H4 tail binding in an extended configuration (Fig. 3 of ref. 34). Despite this difference, the interaction patterns in both configurations match surprisingly well. Specifically, the strong salt bridges between K16 and E61, K20 and E110, and R23 and E156 are very similar in the two configurations. However, two differences arise between the two configurations stemming from the more compact configuration of our H4 tail: the K16 side chain does not need to stretch backwards to mediate salt bridging with E61 and the R19 interacts with Q44 and E56 instead of interacting with N68.

It is also interesting to note that the observed binding mode requires the K16–R23 region for strong interactions with the acidic patch. That this region resides so close to its "parent" histone octamer implies that the two interacting nucleosomes come very close to each other. Such close packing of nucleosomes is consistent with large nucleosome packing ratios observed in fully compact *in vivo* chromatin (~11–15 nuc/11 nm of chromatin fiber length).[59]

The *third* finding is that K16 acetylation destabilizes to some extent the α-helical conformation of the H4 tail. Specifically, K16 acetylation reduces the propensity of the H4 tail to form the α-helical configuration by >20%. Together with our earlier observation that the α-helical conformation of the H4 tail binds stronger to the acidic patch compared to its extended counterpart, the observed destabilization of the H4 α-helix

via K16 acetylation effectively translates into weaker binding of the H4 tail at the acidic patch and in turn reduced compaction of chromatin.

Interestingly, Liu and Duan[38] observed a similar decrease in the stability of the α-helix of the H3 histone tail upon acetylation of Lys4 and Lys9. Wang *et al.*[56] observed no visible change in the α-helicity of the H4 tails upon acetylation of a *single* lysine but acetylation of *multiple* lysines caused a notable increase in the stability of the α-helix. That K16 acetylation causes some destabilization of the α-helix while acetylation of multiple lysines causes stabilization is *not* contradictory. This can be explained by recognizing that K16 acetylation leads to the formation of "destructive" H-bonds between its C═O group and the N–H groups on the sidechains and backbone of nearby residues. On the other hand, acetylation of multiple lysines could lead to the formation of "constructive" H-bonds between the C═O groups of one lysine with the N–H group of another acetylated lysine that is approximately 3–4 residues away and therefore pointing along the same direction. These H-bonds would thus complement the existing H-bond interactions between the peptide groups of standard α-helices and further stabilize the α-helix.

The *fourth* finding is that K16 acetylation reduces the binding affinity of the H4 tail with the acidic patch, assuming that the tails remain α-helical after acetylation. Our docking calculations yield a rough $\Delta\Delta G$ of $+1$ kcal mol^{-1} upon acetylation, which is not small considering that there are two such tails per nucleosome that could get acetylated, resulting in a total free energy change of 2 kcal mol^{-1}. Even if one of the tail interacts with the acidic patch, a 1 kcal mol^{-1} difference translates to a large change in the equilibrium constant of the tail binding to the acidic patch ($\sim$5.3-fold, as given by $K_{AcK}/K_{WT} = \exp[-\Delta\Delta G/RT]$), which could lead to chromatin unfolding. Moreover, the docking estimate of $\Delta\Delta G$ is rough and should not be taken literally.

Further decomposition of the energies through more accurate MM-PBSA calculations reveals that the main contributor to the reduced affinity is the K16 residue itself. Upon acetylation, K16 no longer mediates strong salt-bridging interactions with E61 and longer-range electrostatic interactions with E64, D90, E91, E92, and E102. The loss of these highly favorable interactions upon acetylation more than compensates the loss in solvation free energy accompanying the acetylation, yielding a free energy loss of $\sim$9 kcal mol^{-1}. Hence, the effect of K16 acetylation on binding could be due to changes in *specific* electrostatic interactions. We note the interesting observation that the E64, D90, E91, E92, and E102 residues do not seem to exhibit any change in their interaction free energies upon K16 acetylation. This suggests that these residues interact as favorably with the K16 residues as they do with the solvent. In contrast, the K16 residue prefers to interact with these residues compared to the solvent, as it can mediate cooperative interactions with all five oppositely charged residues of the acidic patch.

Taken together, our results point to an intriguing mechanism by which K16 acetylation could trigger partial unfolding of chromatin. K16 acetylation decreases the free energy of binding of the H4 tail α-helix with the acidic patch through specific electrostatic interactions, decreasing the strength of internucleosomal interactions that stabilize the compact form of chromatin. Concomitantly, K16 acetylation also destabilizes the H4 α-helical region itself that mediates the strongest affinity for the acidic patch, contributing further to the weakening of internucleosomal interactions.

Our findings are also consistent with the prevailing notion that the histone tails may be intrinsically disordered proteins,[60] *i.e.*, proteins that are unstructured in isolation but form well-defined secondary structures upon interaction with external molecules.[61] Indeed, our simulations show that the K16–K20 region of the H4 tail has a 50–70% propensity of forming an α-helix, suggesting that this region likely fluctuates between disordered and α-helical conformations. However, when this tail region comes in contact with the acidic patch, its intermittent α-helical conformation becomes stabilized due to favorable interactions with the patch. Specifically, the acidic patch provides a glove-like fit to the α-helical conformation and mediates favorable hydrogen bonds and salt-bridges with the H4 tail lysines co-oriented due to the α-helical conformation. Our observations could also explain why Baneres *et al.*[62] and Wang *et al.*[56] observe signatures of α-helices in H4 tails in nucleosomes, which are not observable in isolated tails.

As is the case with all biomolecular simulations, one needs to be aware of inherent approximations arising from force field inaccuracies, solvent modeling, and sampling limitations. Also, the ΔG values computed from docking and MM-PB(GB)SA calculations do not account for the configurational entropy of the H4 tail and the acidic patch residues and also do not account for the steric effects and electrostatic field of the nucleosome connecting the H4 tail. Nevertheless, we believe that our calculations capture well leading-order effects of the H4 tail and the acidic patch and provide new insights into chromatin regulation. Ultimately, further experiments would be required to best validate our findings.

In summary, we have carried out the first detailed computational investigation of the conformation of the H4 tail and the binding of the tail to the nucleosomal acidic patch. We have also investigated how the conformation and binding of the H4 tail gets affected by acetylation of K16, a key posttranslational modification involved in gene regulation. Our analyses indicate that the H4 tail is overall disordered but has a strong propensity to form an α-helix in the residue range A15–K20 and that this conformation surprisingly binds very strongly with the nucleosomal acidic patch. Further analyses suggest that K16 acetylation could partially destabilize the binding of the H4 tail and also reduce its tendency to form the α-helical structure, leading to partial unfolding of the chromatin fiber. We have identified the primary binding modes, residues, and energetics involved in the binding of the H4 tail to the acidic patch and identified the molecular mechanism by which K16 acetylation could disrupt these interactions. These results could be relevant to gene regulation, epigenetics, and drug design.

Acknowledgements

We acknowledge computer time on the Center for Theoretical Biological Physics supercomputing cluster. D.Y. acknowledges

research funding from the McNair Scholars Program and the Amgen Scholars Program. G.A. acknowledges the Hellman Fellowship for partial support of this work. We are also grateful to Dr Robert Konecny and Dr Nan Li for computational advice.

References

1 G. Felsenfeld and M. Groudine, *Nature*, 2003, **421**, 448–453.
2 P. J. Horn and C. L. Peterson, *Science*, 2002, **297**, 1824–1827.
3 K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.
4 C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *J. Mol. Biol.*, 2002, **319**, 1097–1113.
5 P. J. J. Robinson and D. Rhodes, *Curr. Opin. Struct. Biol.*, 2006, **16**, 336–343.
6 T. Schalch, S. Duda, D. F. Sargent and T. J. Richmond, *Nature*, 2005, **436**, 138–141.
7 S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock and T. Schlick, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13317–13322.
8 T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074–1080.
9 T. Kouzarides, *Cell (Cambridge, Mass.)*, 2007, **128**, 693–705.
10 S. L. Berger, *Curr. Opin. Genet. Dev.*, 2002, **12**, 142–148.
11 K. A. Gelato and W. Fischle, *Biol. Chem.*, 2008, **389**, 353–363.
12 W. Fischle, Y. Wang and C. D. Allis, *Nature*, 2003, **425**, 475–479.
13 A. Akhtar and P. B. Becker, *Mol. Cell*, 2000, **5**, 367–375.
14 N. Suka, Y. Suka, A. A. Carmen, J. Wu and M. Grunstein, *Mol. Cell*, 2001, **8**, 473–479.
15 W. Dang, K. K. Steffen, R. Perry, J. A. Dorsey, F. B. Johnson, A. Shilatifard, M. Kaeberlein, B. K. Kennedy and S. L. Berger, *Nature*, 2009, **459**, 802–807.
16 M. F. Fraga, E. Ballestar, A. Villar-Garea, M. Boix-Chornet, J. Espada, G. Schotta, T. Bonaldi, C. Haydon, S. Ropero, K. Petrie, N. G. Iyer, A. Perez-Rosado, E. Calvo, J. A. Lopez, A. Cano, M. J. Calasanz, D. Colomer, M. A. Piris, N. Ahn, A. Imhof, C. Caldas, T. Jenuwein and M. Esteller, *Nat. Genet.*, 2005, **37**, 391–400.
17 P. Allegra, R. Sterner, D. F. Clayton and V. G. Allfrey, *J. Mol. Biol.*, 1987, **196**, 379–388.
18 V. G. Allfrey, in *Chromatin and Chromosome Structure*, ed. H. J. Li and R. Eckhardt, Academic Press, New York, 1977, pp. 167–191.
19 T. R. Hebbes, A. W. Thorne and C. Crane-Robinson, *EMBO J.*, 1988, **7**, 1395–1402.
20 D. Y. Lee, J. J. Hayes, D. Pruss and A. P. Wolffe, *Cell (Cambridge, Mass.)*, 1993, **72**, 73–84.
21 J. D. McGhee and G. Felsenfeld, *Annu. Rev. Biochem.*, 1980, **49**, 1115–1156.
22 M. Garcia-Ramirez, C. Rocchini and J. Ausio, *J. Biol. Chem.*, 1995, **270**, 17923–17928.
23 C. Tse, T. Sera, A. P. Wolffe and J. C. Hansen, *Mol. Cell. Biol.*, 1998, **18**, 4629–4638.
24 M. Shogren-Knaak, H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie and C. L. Peterson, *Science*, 2006, **311**, 844–847.
25 P. J. J. Robinson, W. An, A. Routh, F. Martino, L. Chapman, R. G. Roeder and D. Rhodes, *J. Mol. Biol.*, 2008, **381**, 816–825.
26 H. Walia, H. Y. Chen, J. M. Sun, L. T. Holth and J. R. Davie, *J. Biol. Chem.*, 1998, **273**, 14516–14522.
27 J. D. Anderson, P. T. Lowary and J. Widom, *J. Mol. Biol.*, 2001, **307**, 977–985.
28 G. Arya and T. Schlick, *J. Phys. Chem. A*, 2009, **113**, 4045–4059.
29 S. J. McBryant, J. Klonoski, T. C. Sorensen, S. S. Norskog, S. Williams, M. G. Resch, J. A. Toombs, 3rd, S. E. Hobdey and J. C. Hansen, *J. Biol. Chem.*, 2009, **284**, 16716–16722.
30 J. Mozziconacci, H. Wong and J. M. Victor, *FEBS J.*, 2007, **274**, 78.
31 J. Zhou, J. Y. Fan, D. Rangasamy and D. J. Tremethick, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1070–1076.
32 B. P. Chadwick and H. F. Willard, *J. Cell Biol.*, 2001, **152**, 375–384.
33 J. V. Chodaparambil, A. J. Barbera, X. Lu, K. M. Kaye, J. C. Hansen and K. Luger, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1105–1107.
34 K. Luger and T. J. Richmond, *Curr. Opin. Genet. Dev.*, 1998, **8**, 140–146.
35 A. L. Garske and J. M. Denu, *New Frontiers in Epigenetic Modifications*, Wiley-VCH, Weinheim, 2009, vol. 42.
36 D. A. Case, T. A. Darden, T. E. Cheatham, I, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, University of California, San Francisco, 2008.
37 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.
38 H. Liu and Y. Duan, *Biophys. J.*, 2008, **94**, 4579–4585.
39 A. Onufriev, D. Bashford and D. A. Case, *Proteins: Struct., Funct., Genet.*, 2004, **55**, 383–394.
40 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
41 E. Lin and M. S. Shell, *J. Chem. Theor. Comput.*, 2009, **5**, 2062–2073.
42 H. Lei, C. Wu, H. Liu and Y. Duan, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 4925–4930.
43 H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
44 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
45 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
46 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
47 A. Varshney and F. J. Brooks, UNC at Chapel Hill, Chapel Hill, NC, 1993.
48 E. Lindahl, B. Hess and D. Van Der Spoel, *J. Mol. Model.*, 2001, **7**, 306–317.
49 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
50 V. Munoz and L. Serrano, *Nat. Struct. Biol.*, 1994, **1**, 399–409.
51 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
52 B. Dorigo, T. Schalch, K. Bystricky and T. J. Richmond, *J. Mol. Biol.*, 2003, **327**, 85–96.
53 M. F. Sanner, *J. Mol. Graphics Modell.*, 1999, **17**, 57–61.
54 A. Cornish-Bowden, *J. Biosci.*, 2002, **27**, 121–126.
55 R. D. Lins and U. Rothlisberger, *J. Chem. Theor. Comput.*, 2006, **2**, 246–250.
56 X. Wang, S. C. Moore, M. Laszckzak and J. Ausio, *J. Biol. Chem.*, 2000, **275**, 35013–35020.
57 B. D. Strahl and C. D. Allis, *Nature*, 2000, **403**, 41–45.
58 M. Kruithof, F. T. Chien, A. Routh, C. Logie, D. Rhodes and J. van Noort, *Nat. Struct. Mol. Biol.*, 2009, **16**, 534–540.
59 P. J. Robinson, L. Fairall, V. A. Huynh and D. Rhodes, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6506–6511.
60 J. C. Hansen, X. Lu, E. D. Ross and R. W. Woody, *J. Biol. Chem.*, 2006, **281**, 1853–1856.
61 A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradovic, *Biochemistry*, 2002, **41**, 6573–6582.
62 J. L. Baneres, A. Martin and J. Parello, *J. Mol. Biol.*, 1997, **273**, 503–508.