

Protein disorder-to-order transition enhances the nucleosome-binding affinity of H1

Akshay Sridhar¹, Modesto Orozco^{2,3} and Rosana Collepardo-Guevara^{1,4,5,*}

¹Maxwell Centre, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK,

²Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Baldiri i Reixac,

19, 08028 Barcelona, Spain, ³Department of Biochemistry and Biomedicine, University of Barcelona, Av. Diagonal

647, 08028 Barcelona, Spain, ⁴Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2

1EW, UK and ⁵Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Received February 04, 2020; Revised April 02, 2020; Editorial Decision April 09, 2020; Accepted April 14, 2020

ABSTRACT

Intrinsically disordered proteins are crucial elements of chromatin heterogenous organization. While disorder in the histone tails enables a large variation of inter-nucleosome arrangements, disorder within the chromatin-binding proteins facilitates promiscuous binding to a wide range of different molecular targets, consistent with structural heterogeneity. Among the partially disordered chromatin-binding proteins, the H1 linker histone influences a myriad of chromatin characteristics including compaction, nucleosome spacing, transcription regulation, and the recruitment of other chromatin regulating proteins. Although it is now established that the long C-terminal domain (CTD) of H1 remains disordered upon nucleosome binding and that such disorder favours chromatin fluidity, the structural behaviour and thereby the role/function of the N-terminal domain (NTD) within chromatin is yet unresolved. On the basis of microsecond-long parallel-tempering metadynamics and temperature-replica exchange atomistic molecular dynamics simulations of different H1 NTD subtypes, we demonstrate that the NTD is completely unstructured in solution but undergoes an important disorder-to-order transition upon nucleosome binding: it forms a helix that enhances its DNA binding ability. Further, we show that the helical propensity of the H1 NTD is subtype-dependent and correlates with the experimentally observed binding affinity of H1 subtypes, suggesting an important functional implication of this disorder-to-order transition.

INTRODUCTION

In Eukaryotic cells, genomic DNA is packaged into chromatin, a nucleoprotein complex whose fundamental repeat-

ing unit is the nucleosome (1,2). Nucleosomes are formed by ~147 bp of DNA wrapped around a histone protein octamer core (two copies each of H3, H4, H2A and H2B) and are joined together by free DNA linker segments. Most organisms contain an additional type of histone, the H1 linker histone protein. H1 is a chromatin architectural protein that binds to nucleosomes near the entry/exit site of the linker DNA (3,4), interacts with ~20 bp of the DNA linkers (5), and critically influences chromatin organization (6–8).

The family of H1 proteins is highly divergent (9). In mammals, up to 11 different H1 subtypes (H1.1–H1.10 and H1.0) have been identified (10–12) and characterized as having differing expression timings (13), extent of post-translational modifications (11,14), chromatin compacting capabilities (15), preference for eu-/heterochromatin regions (16), and interactions with other chromatin architectural proteins (17,18).

From the protein structure point of view, the H1 proteins contain a ~75 residue-long winged-helix globular domain, an unstructured N-terminal domain (NTD) of up to 45 residues, and a long mostly unstructured C-terminal domain (CTD) of ~100 residues (19,20) (Figure 1A). The globular domain is highly conserved within the H1 family (21,22) but the NTD and CTD vary significantly in sequence and length. This has led to the hypothesis that the differential behaviour among H1 subtypes stems from the heterogeneity in the diverging unstructured terminal domains (14). Because the CTD is believed to account for most of the subtype heterogeneity (23–25), significant efforts have been made to understand the conformational landscape of the CTD in isolation (26), when bound to DNA (27) or to a negatively charged IDP with high affinity (28), and when bound to a full chromatosome (4,29).

H1's shortest domain, the NTD, plays a negligible role in H1's chromatin compaction capability (30). However, its deletion reduces the binding affinity of H1 to the nucleosome significantly (30,31). Across H1 subtypes, an analysis of the NTD role has shown contrasting results with swaps

*To whom correspondence should be addressed. Tel: +44 1223 747371; Email: rc597@cam.ac.uk

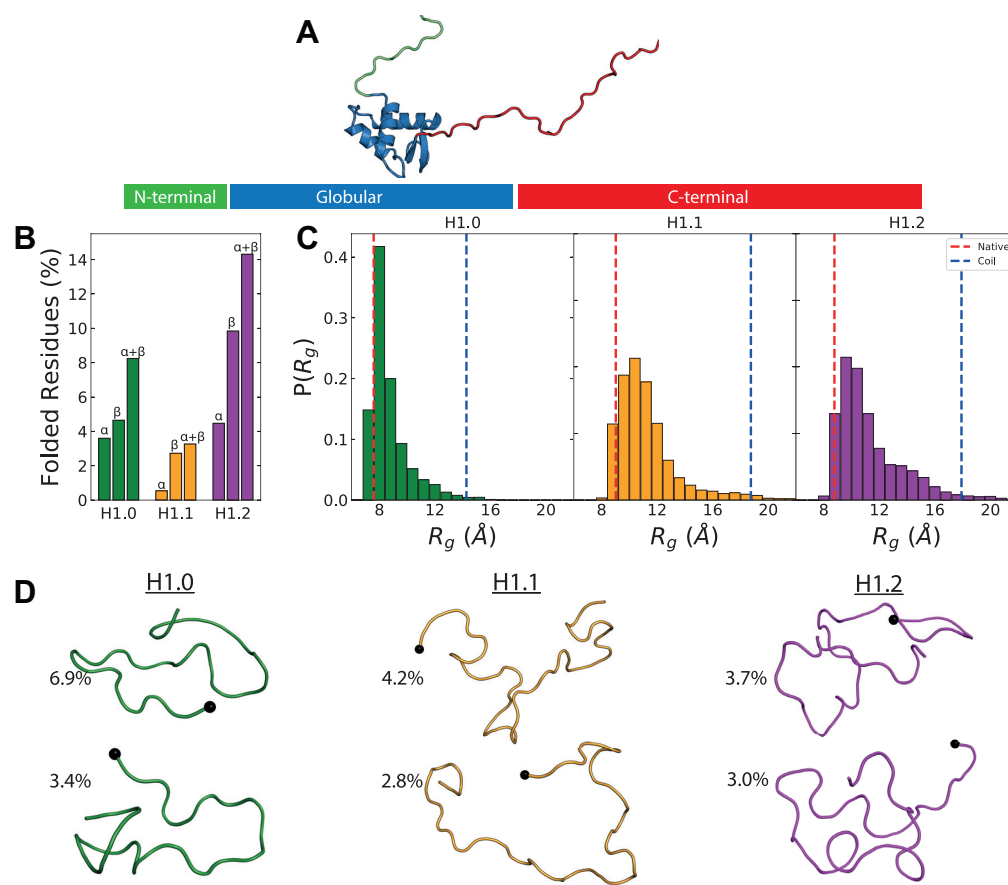


Figure 1. Structure of H1 (A) The constituent domains of H1 – winged-helix globular domain (blue), unstructured NTD (green) and unstructured CTD (red). (B) Ensemble averages of the percentage of secondary structures in each of the three NTDs from the final 200 ns of the 300 K T-REMD trajectory. (C) Distributions of the radius of gyration (R_g) of the NTD C α atoms. The experimentally predicted regressions for random-coil and globular proteins are shown in blue and red respectively. (D) Illustrations of the highest populated clusters of the three NTDs. The final residue of the NTDs (point of attachment to the winged-helix globular domain) is shown as a black sphere.

of the H1.0/H1.2 and H1.0/H1x NTDs resulting in an exchange of their binding characteristics (32,33), but swaps of H1.1/H1.5 NTDs displaying no effects (23). The NTDs are rich in basic residues crucial for DNA interactions and otherwise predominantly composed of G/E/P/S/T/A amino acids – an archetypical composition of IDPs (34). Indeed, the NTD has been shown to be unstructured in aqueous solution and to have a propensity to fold in the presence of short DNA segments (35). However, whether the NTD remains disordered or undergoes a disorder-to-order transition when H1 binds to the nucleosome is not known, which limits our mechanistic understanding of its functional role (36,37).

Given the unstructured and flexible nature of the NTD, characterizing its conformational landscape experimentally at atomistic resolution is challenging. Complementary to the experimental efforts, advanced simulation methods, such as enhanced sampling molecular dynamics (MD) simulations, have been successfully used in recent years to investigate the configurational ensembles of core histone tails (38,39) and the CTD of H1 (29), and to understand the role of post-translational modifications within them (40,41). Here, we take advantage of recent advances in enhanced-sampling methods and the development of state-of-the-art

biomolecular force fields (42,43) for the description of IDPs. In particular, we use the Parallel-Tempered Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) method (44)—a combination of the Temperature Replica-Exchange (T-REMD) and metadynamics techniques developed to facilitate sampling of phase space. While T-REMD assists in the crossing of high energy barriers via high temperature replicas, MetaD-WTE accelerates sampling through the use of a history-dependent biasing potential acting along specific collective variables (CV) (see Materials and Methods). The combination of both helps overcome two well-known limitations of using T-REMD or metadynamics independently: (1) thorough sampling along slow diffusing CVs not considered explicitly (44), and (2) increase of the energetic overlap among temperature replicas and, hence, significant reduction of the total number of replicas needed and the associated computational costs (45) (see Materials and Methods).

By means of a large set of microsecond-long PTMetaD-WTE, T-REMD, and Biased-Exchange Metadynamics (BEMD) simulations (Supplementary Table S1), we fully characterize the difference in structure across various NTD subtypes. We compare how the structural behaviour of the NTD changes when it is free in solution, versus when it is

bound to DNA or to a full 211-bp nucleosome (a 147-bp nucleosome with two symmetric 32-bp linker DNA arms). Our results show that the NTD of H1 is completely unstructured in solution, but upon DNA or nucleosome binding transitions into an amphiphilic helix with charged and uncharged residues on opposing faces. Further, we demonstrate that formation of this amphiphilic helix depends on the charge neutralization of the Lys sidechains by DNA. Crucially, our work demonstrates that the amphiphilic helical propensity of the NTD of H1 varies proportionally to the experimentally observed binding affinity of H1 subtypes, suggesting a potential implication of the NTD disorder-to-order transition in the function of linker histones.

MATERIALS AND METHODS

System and simulation setup

The sequences of the human H1 subtypes were obtained from the Uniprot Database (46) and their N-terminal domains were built in an extended conformation using Avogadro (47) without capping of the N- or C-terminus. For computational expedience associated with reducing the solvent molecules required, we bent the initial linear configurations through a short 5 ns simulation in GBSA implicit solvent (39,48). The resulting structures were then used as initial configurations for all further simulations. The basic sub-region of each NTD was defined as commencing three residues (one-turn) before the first Lys of the NTD.

The simulations were performed using Gromacs 2016 (49) with Plumed 2.3.0 (50) in explicit solvent and 0.15 M Na⁺/Cl⁻ ions. Two state-of-the-art force fields were used for the T-REMD simulations: Charmm36M (42), and Amber99SB-ILDN (51) with Helix-Coil transition specific backbone (52) and charge (53) corrections (Amberff99sb*-ildn-q). With the Amber force field, water was modelled using the TIP3P parameters (54), and ions with the parameters of Aqvist (55) and Dang (56). With Charmm36M, water was modelled using the modified TIP3P parameters with increased Hydrogen atom Lennard-Jones well-depth (42), and ions with the parameters of Beglov and Roux (57), including the nbfx changes of Luo and Roux (58) and Venable *et al.* (59).

For the 20-bp B-DNA strands, we used two different force fields: Charmm DNA and parmbsc0 (60,61). The Settle algorithm (62) was used to constrain bond lengths and angles of water molecules and P-LINCS was used for all other bond lengths. Temperatures were maintained using the v-rescale thermostat (63) and the pressure at 1 bar using the Parrinello-Rahman barostat (64). Long range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) (65) algorithm with a cut-off of 1.0 nm.

Temperature replica-exchange simulations

Temperature Replica-Exchange (T-REMD) simulations were performed starting from representative NTD configurations obtained through implicit-solvent equilibration. A temperature range of 300–450 K was used and the number/temperature of replicas were calculated using the predictor of Patriksson and van der Spoel (66) and an acceptance probability of 20%. This resulted in 56, 96 and 72

replicas for the H1.0, H1.1 and H1.2 subtypes respectively. The replicas were simulated for 250 ns each and exchanges between adjacent temperatures were attempted every 10 ps. The trajectory frames from the lowest temperature (300 K) were then used for analysis. Convergence of the T-REMD simulations were assessed by monitoring changes in the helical and beta content of the residues in 40 ns intervals after discarding the initial 50 ns for equilibration.

Parallel-tempered metadynamics in the well-tempered ensemble

We begin by applying a time-dependent biasing potential, $V(S,t)$, defined as a sum of Gaussians deposited along the collective variables (CV) space (S)

$$V(S, t) = W \sum_{t' = \tau_G, 2\tau_G, \dots} \exp\left(-\frac{(S - S_{t'})^2}{2\sigma^2}\right)$$

where τ_G is the time interval at which the Gaussians are added with height W , width σ and mean $S_{t'}$. The biasing potential discourages the trajectory from visiting states already sampled. Choosing an accurate set of CVs to adequately promote sampling is often non-trivial (45) with energy barriers along unaccounted CVs affecting sampling efficiency (67). The Parallel-Tempered Metadynamics (PT-MetaD) (68) method combines the above described T-REMD with metadynamics to deal with unaccounted slow degrees of freedom. The metadynamics biases of individual replicas at different temperatures are evolved independently and the adjacent replicas can exchange with a probability of

$$P = \min \left\{ 1, \exp \left[(\beta_i - \beta_j)(U_i - U_j) + \beta_i (V_i(S_i) - V_i(S_j)) + \beta_j (V_j(S_j) - V_j(S_i)) \right] \right\}$$

where U is the potential energy, V is the bias potential and β is the inverse of the thermodynamic temperature ($1/k_B T$).

However, energy distributions of the replicas 'i' and 'j' need to sufficiently overlap for exchanges to occur (Supplementary Figure S1). This in-turn requires the use of a large number of replicas as predicted for the T-REMD simulations. Here, we use the Well-Tempered Ensemble (WTE) (44) to improve efficiency. The method enhances energetic fluctuations of each replica and thereby its overlap with adjacent replicas through the use of the potential energy as an independent CV. For the current work, we used the procedure of Bernetti *et al.* (69) with eight temperatures geometrically distributed between 298 and 400 K as

$$T_i = T_{\min} \left(\frac{T_{\max}}{T_{\min}} \right)^{(i-1)/(N-1)}$$

where N is the number of replicas. The configurations obtained from implicit solvent equilibration were equilibrated in the eight temperatures before a preliminary 10 ns PTMetaD-WTE run was carried out. This initial run used only the potential energy (U) as the CV and deposited Gaussians of height 2.5 kJ/mol, width 500 kJ/mol and bias factor of 50. The Gaussians were deposited every 250 steps and sufficient overlap of the replica potential energies was ensured through a histogram analysis at the end of this run (see Supplementary Figure S1).

The bias from this preliminary run was then kept fixed in the production PTMetaD-WTE simulation and ensured exchanges between adjacent replicas. This production run used two CVs: (1) the alpha helical content (S_α) and (2) the radius of gyration (S_{rg}) of C α atoms. The alpha helical content CV (S_α) has been used extensively as a CV within metadynamics simulations to sample the random coil-helix transformation of proteins, with back-calculated NMR shifts of the resulting conformational ensembles being in good agreement with experimental values (69–71). It is defined following the work of Pietrucci and Laio (72) as

$$S_\alpha = \sum \frac{1 - \left(\frac{\Delta\text{RMSD}}{R_0}\right)^n}{1 - \left(\frac{\Delta\text{RMSD}}{R_0}\right)^m}$$

where R_0 , n and m are 0.08 nm, 8 and 12 respectively. ΔRMSD is the root mean square difference of six residue segments between the configuration and ideal alpha conformations. The Backbone (C α , C, N, O) and C β atoms were included in the RMSD calculation.

The radius of gyration CV (S_{rg}) of the C α atoms is defined as

$$S_{rg} = \left(\frac{\sum_i^n m_i |r_i - r_{com}|^2}{\sum_i^n m_i} \right)^{1/2}$$

where r_i and m_i are the positions of and mass of C α atom i . r_{com} is the centre of mass of the NTD C α atoms.

DNA–protein metadynamics simulations

The H1.0 subtype was used to examine the effects of DNA binding on the conformational preferences of the NTD. The nucleosomal on-dyad structure of the H1.0 globular domain (PDB ID: 5NL0 (4)) was obtained and Modeller (73) was used to add the NTD basic subregion in an extended random coil configuration. The relative orientations of this built basic subregion and one of the nucleosomal linker arms (20 bp) was used as the initial configuration (Figure 3A) for metadynamics simulation.

The metadynamics simulations utilized two CVs: (1) the alpha helical content (S_α) and (2) the number of NTD–DNA electrostatic contacts (S_{cont}). The CVs thus simultaneously allow exploring NTD–DNA binding/unbinding and changes in helical propensity. The S_{cont} CV was defined as

$$S_{cont} = \sum_{i \in H^+} \sum_{j \in O^-} \frac{1 - \left(\frac{\vec{r}_i - \vec{r}_j}{R_0}\right)^n}{1 - \left(\frac{\vec{r}_i - \vec{r}_j}{R_0}\right)^m}$$

where R_0 , n and m are 0.2 nm, 8 and 10 respectively. H^+ is the set of all positively charged sidechain hydrogen atoms within Lys/Arg residues and O^- is the set of backbone oxygen atoms within the central 11-bp of DNA. Metadynamics gaussians were deposited every 10 ps and the simulation was run for 600 ns and the initial 50 ns was disregarded for analysis. To assist convergence, the lowest terminal base-pair (BP 20) and the backbone of the NTD's carboxy-terminal residue were both restrained with a force constant of 750 kJ/mol/nm².

Trajectory analysis

The simulation trajectories were analysed using a combination of Plumed/Gromacs tools together with Python scripts using the MDAnalysis and MDTraj libraries (74,75). The secondary structures were assigned using DSSP (76) with residues being part of either alpha- or 3_{10} -helix motifs counted towards the total helical propensity. The theoretical compaction (radius-of-gyration) of proteins with globular (77) and random-coil (78) structures were calculated as

$$R_g^{glob}(N) = 2.2N^{0.38}$$

$$R_g^{coil}(N) = 2.02N^{0.60}$$

Clustering was performed using the single-linkage clustering method where the structure ' j ' is assigned to cluster ' i ' if the RMSD of C α atoms between them is $< 2.5 \text{ \AA}$.

In the metadynamics simulations, the addition of history dependent bias potentials precludes a direct ensemble averaging of the system's characteristics as simulation time is without physical meaning. The methodology of Tiwary and Parrinello (79) was thus used to reweight trajectory frames and subsequently calculate the unbiased equilibrium averages. Briefly, the Probability Distribution of the biased system $P(R, t)$ can be expressed as

$$P(R, t) = \frac{\exp(-\beta[U(R) + V(s(R), t)])}{\int \exp(-\beta[U(R) + V(s(R), t)]) dR}$$

where $U(R)$ is the internal potential and $V(s(R), t)$ is the bias potential as given by Equation (1). The introduction of the delta function $\delta(s - S(R))$ and the unbiased probability density function $P_0(R, t)$ allows the expression of equation (7) as

$$P(R, t) = P_0(R, t) \cdot \exp(-\beta[V(s(R), t) - c(t)])$$

where $c(t)$ is the time-dependent bias offset (80) defined as

$$c(t) = \frac{1}{\beta} \ln \left[\frac{\int \exp(-\beta F(s)) ds}{\int \exp(-\beta [F(s) + V(s, t)]) ds} \right]$$

The $c(t)$ offset is calculated using Plumed following the protocol of Tiwary and Parrinello (79) and used to assign weights $w(t) \propto \exp(\beta[V(s(R), t) - c(t)])$ from the biased probability density. See Bonomi *et al.* (80) for a complete derivation. The equilibrium average of any system characteristic O is then calculated as

$$\langle O \rangle = \frac{\sum w_i(t) O_i}{\sum w_i}$$

For the calculation of per-residue helical content, per-frame O_i of the residue was set as 1 if DSSP (76) predicted an alpha/ 3_{10} motif and 0 otherwise. Similarly, the clusters were reweighted using $w(t)$ of the trajectory frames assigned to them.

Docking

Docking of the clustered conformations from the PTMetaD-WTE simulations were performed using the HADDOCK webserver (81). The B-DNA for docking was built as a 20-bp long strand using the Nucleic Acids

Builder within Amber16 (82). To allow for all possible orientations of the NTD basic sub-regions when interacting with this strand, the central 12-bp (one-turn) were considered the ‘active’ residues for docking. For the NTD basic subregions, the Lys residues within the basic ‘face’ were considered the ‘active’ residues. The His residues were considered neutrally charged and no segment within the DNA or NTD was considered flexible.

Potential of mean force calculations

The Potential of Mean Force (PMF) calculations were performed using a protocol similar to that of Wieczor and Czub (83). Firstly, the helical axis of the 20-bp strand was aligned along the z-axis and the radial-distance between the Central BP and the Helix C α atoms along the XY-plane was used as the CV. Initial frames for the umbrella windows were generated from the docked configuration by increasing this radial distance CV using 20 windows at intervals of 0.1 nm. Each window was then simulated for 100 ns using an umbrella biasing potential of 750 kJ/mol/nm² along the CV. To reduce diffusion of the DNA strand and thereby assist convergence (84), the two terminal base-pairs at each end (BP 1,2,19,20) were restrained with a force constant of 500 kJ/mol/nm². The initial 25 ns were discarded, and the rest of the trajectory was used for analysis. PMF profiles were generated using the GROMACS implementation (g-wham) (85) of the weighted histogram analysis method (WHAM). Bayesian bootstrapping with 100 bootstraps were used to estimate the errors for each profile.

RESULTS AND DISCUSSION

The isolated H1 NTD peptide is disordered regardless of subtype

The amino acid sequences of the various H1 NTD subtypes display several distinguishable features of IDPs: low hydrophobicity, high net charge, and low sequence complexity (86). Thus, we first determine if the varying amino acid sequences of the different subtypes of the H1 NTD are sufficient to induce different structural behaviours, such as, random coil conformations, disordered structures with flickering secondary structural elements, or stable secondary structural folds. To answer this, we assess the conformational landscape of the isolated NTD peptides of three different H1 human subtypes—H1.0, H1.1 and H1.2—in solution through T-REMD simulations. We consider temperature replicas ranging from 300 to 450 K and run each replica for 250 ns for an accumulated sampling of up to 24 μ s per system (see Materials and Methods).

We find that for all subtypes, the H1 NTD emerges as a highly unstructured domain displaying only transient secondary structural elements (Figure 1), irrespective of the force field used (Amberff99sb*-ildn-q (51–53) and Charmm36M (42); Supplementary Figure S2). Indeed, for both force fields, the conformational ensembles include only a small fraction (<15%) of residues with α -helical or β -strand structural elements (Figure 1B). These observations are consistent with circular dichroism experiments (35,87) showing that the spectra for the NTDs of H1.0 and mH1.4

(homologous to hH1.2, Supplementary Figure S3) in aqueous solution are dominated by contributions of the random coil.

To further explore the structural heterogeneity of the NTDs, their size distributions were calculated and compared to the mean radius of gyration (R_g) of a globular (77) and thermally denatured random-coil proteins (78). The structural heterogeneity of the NTDs is evident from the wide distributions of radius of gyration (Figure 1C). However, despite the large structural variation, most of the conformations adopted by the NTDs are compact, as shown by radii of gyration that are closer to the value predicted for a globular protein of similar length (Figure 1C and Materials and Methods). In contrast, Replica-Exchange with Solute Tempering (REST2) (88) simulations of the CTD of H1.0 (29) reveal extended conformations (Supplementary Figures S4 and S5). The differences in compaction between the two unstructured domains might plausibly be attributed to the uneven charge density within them. The 97-residue H1.0 CTD is predominantly composed of basic residues (42.2%, 41 Lys) while the 26-residue H1.0 NTD only has six basic residues (23.1%). To test this hypothesis, we split the H1.0 NTD into hydrophobic and basic subregions (Figure 2D) and recalculated the radius of gyration of each region from the T-REMD simulations (Supplementary Figure S6). Consistently, we observe that the basic subregion, with higher charge density, adopts extended conformations that yield the typical distribution of the radius of gyration of IDPs (Supplementary Figure S6). In contrast, the hydrophobic subregion adopts predominantly collapsed conformations that contribute to the comparatively compact NTD radius of gyration observed in Figure 1B.

This distinction between the two unstructured N- and C-terminal domains might provide a molecular explanation for their experimentally observed contrasting roles in mediating H1-nucleosome interactions (23,89). Extended states of the CTD might allow it to initiate the process of nucleosome recognition through long-range electrostatic interactions (89) while the collapsed conformations of the NTD might preclude this (23).

H1 NTD undergoes a subtype-specific disorder-to-order transition upon charge neutralization

The NTD can be divided into two distinct subregions (90): the N-terminal subregion that lacks charged residues and is enriched in Ala, Pro and other non-aromatic hydrophobic residues and the basic subregion (close to the globular domain), which contains a high proportion of positive Lys/Arg residues crucial for electrostatic interactions with the DNA (Figure 2D) (90). Binding of H1 NTD to DNA is, hence, expected to significantly transform the structural behaviour of the second half of the NTD due to screening of the electrostatic repulsion among its numerous positively charged residues enabled by the interactions with the phosphate backbone (91).

We perform PTMetaD-WTE (68) simulations of the basic subregions of the three NTDs explored above in two conditions: (1) with standard charged Lys (WT), and (2) with neutralized Lys sidechains (neutralized; as an approximation to NTD charge neutralization by DNA). We use

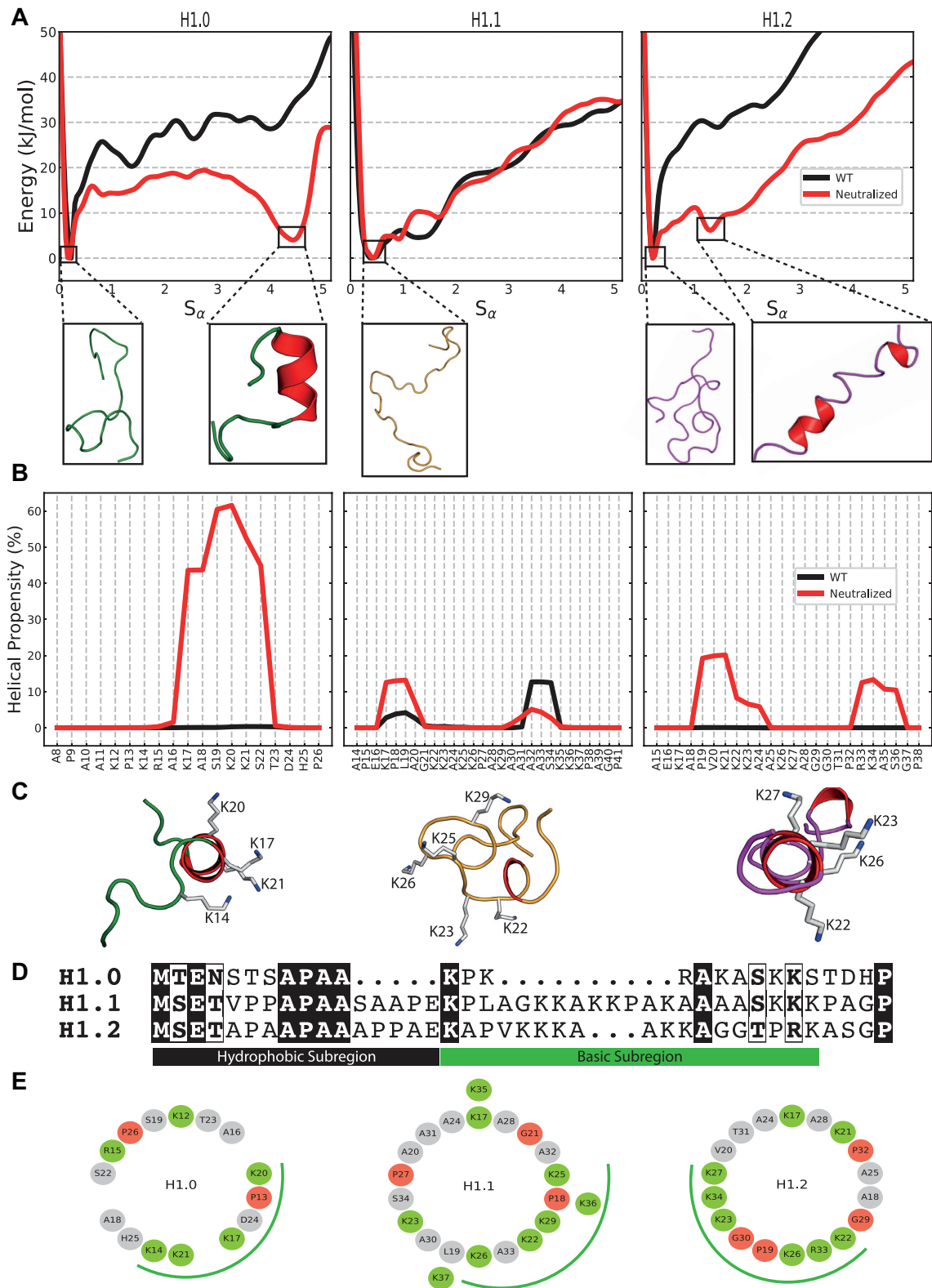


Figure 2. Inducible amphipathic helicities of the H1 NTDs. **(A)** The one-dimensional free energies along the S_{α} CV for the WT NTDs (black) and with neutralized Lys sidechains (red). The conformations at each of the free energy minima are illustrated with the α -helical motifs coloured in red. **(B)** The reweighted per-residue helical content from PTMetaD-WTE simulations of the WT NTDs (black) and with neutralized Lys sidechains (red). **(C)** Top-down illustrations of the induced NTD conformations from PTMetaD-WTE simulations with neutralized Lys sidechains. The α -helical motifs are coloured red and the residues within the positive ‘face’ are illustrated. **(D)** Sequence comparison of the of the hH1.0, hH1.1 and hH1.2 linker histone NTDs. **(E)** Helical wheel projections of the basic subregions of the three H1 NTDs. Basic Arg/Lys residues are in green while the helix-breaking Pro/Gly residues are in red. The basic ‘face’ of each subtype is illustrated with a green curve.

two collective variables (CV) that show complete diffusivity along the simulation (Supplementary Figure S7)—the radius of gyration (S_{Rg}) and a function of the alpha-helical content (S_α) (72)—and eight geometrically spaced temperature replicas from 300 to 450 K (see Materials and Methods) together with enhanced potential energy fluctuations (WTE - Supplementary Figure S1) (44). To measure alpha-helical content, S_α is defined as a function of the RMSD between the NTD and a six-residue ideal α -helix (72) (see Methods). We chose these parallel-tempering metadynamics method as it allows us to characterize the effects of charge neutralization in the statistical ensembles of the H1 NTDs and on their free energy surfaces as a function of their secondary structural content. In all cases, we run the simulations for 3.6 μ s as we achieve convergence after 2.8 μ s of accumulated sampling (350 ns per replica; see Supplementary Figure S8).

For the WT NTD systems, independently of the subtype, the free-energy profile along the S_α collective variable shows a single minimum at values of low helical content $S_\alpha \sim 0.1$, which evidences that the most energetically favourable conformations for all subtypes are those lacking secondary structure (Figure 2A). This is in agreement with the predominantly unstructured ensembles observed in our T-REMD simulations for the same systems (Figure 1). Upon charge neutralization, all three variants preserve a global minimum at S_α values of ~ 0.1 , signalling favourable unstructured conformations in all cases. However, interestingly, the neutralized NTDs of the H1.0 and H1.2 subtypes exhibit a second low energy minima at a higher value of S_α (Figure 2A), corresponding to conformations with helical secondary structural elements. Transient short-lived β -motifs were also observed, but with propensities that remain unchanged upon charge neutralization (Supplementary Figure S10). To discard the existence of other energy minima, the two-dimensional free energy profiles along the two CVs were plotted and compared between the two Lys charge states (see free energy surface (FES) plots in Supplementary Figure S9).

The minima of the H1.0 NTD at large S_α values ($S_\alpha=4.4$) is nearly as stable as that of the unstructured conformations ($\Delta E = 3.6$ kJ/mol). The helical propensity of the neutralized H1.0 NTD (Figure 2B), estimated using the reweighted probability distributions stemming from our PTMetaD-WTE simulations (see Materials and Methods) (79), shows that this is a single seven residue-long helix spanning the region $_{16}AKASKK_{22}$. To validate this structural transition, we performed Biased-Exchange Metadynamics (BEMD) (Supplementary Discussion, Supplementary Table S2) of the whole length H1.0 NTD with neutralized Lys sidechains using varying force fields/water/ion parameters (Supplementary Table S3). While the extent of helicity varied, the span of the motif remained invariant throughout (Supplementary Figure S11).

Further inspection of this motif (Figure 2C), shows that forming a helix would concentrate the positively charged residues K14, K17, K20 and K21 on the same face of the helix, creating a charged face on the NTD that should favour interactions with DNA. Indeed, the projection of sidechains down the axis of an ideal helix (3.6 residues per turn, 100°

Table 1. Characteristic differences in the sequences of the NTD basic subregion of the H1.0, H1.1 and H1.2 subtypes. The basic subregion was considered as the segment between the first and last Arg/Lys residues within the NTD

Subtype	Length	No. of Arg/Lys	No. of Pro/Gly
H1.0	10	6 (60.0%)	1
H1.1	21	9 (42.8%)	3
H1.2	18	8 (44.4%)	4

between residues), known as helical wheel analysis, reveals that the sequence of amino acids in the H1.0 NTD is ideal for clustering charge along one face (Figure 2E), as occurs in many proteins that undergo a disordered-to-ordered transition upon DNA binding (92,93). Hence, once electrostatic repulsion is screened, the sequence of the H1.0 NTD leads to focussing of positive charge on one face of an α -helix favouring then the stability of H1.0 NTD-DNA contacts.

Although the H1.2 NTD with uncharged Lys sidechains also exhibits a second minimum corresponding to a partial helical conformation, this occurs at a much lower S_α value of 1.3 and with a free energy slightly higher ($\Delta E = 5.8$ kJ/mol) (in-comparison to the unstructured conformation). In line with the lower S_α value, the per-residue helicity for this subtype shows two short and more transient helical motifs involving residues $_{19}PVKKKA_{24}$ and $_{33}RKAS_{36}$ (Figure 2B). Inspection of the H1.2 NTD sequence explains why this subtype is unable to form a single long helix: each of the two short helices starts with a ‘helix-initiating’ Pro and are separated by a helix-breaking double Gly (G29/G30) motif (94). This is consistent with a helix-Gly-Gly-helix motif observed in the mouse H1.4 NTD (>95% sequence similarity to hH1.2 NTD, Supplementary Figure S3) in the helix-stabilizing solvent TFE (87). Despite helicity in the H1.2 NTD case being shorter-lived than in the H1.0 subtype, transient partial folding and rigidification of the H1.2 NTD upon DNA binding would imply enabling a more compact configuration that concentrates positive charge (in this case three and two basic residues per helical motif) better than a random coil and favours stronger interactions with DNA.

In contrast to the two other H1 NTDs, the H1.1 subtype exhibits a one-dimensional free-energy surface along S_α that is invariant with the Lys charge state. The per-residue helicity plot in Figure 2B confirms only two small segments with low propensity (<10%).

The differences in folding propensities that we observe correlate with the percentage of positively charged residues and the lack of helix-breaking residues that each system has (see Table 1). That is, the H1.0 NTD subtype that possesses the highest folding propensity, has the highest concentration of positive residues and only one ‘helix-destabilizing’ residues. Consistently, the H1.1 NTD subtype that exhibits the weakest folding propensity, has the lowest positive charge concentration in the set and is interspersed with four ‘helix-destabilizing’ residues (95) (see Table 1 and Figure 2D). The helical-wheel analysis in Figure 2E further illustrates this trend with the H1.0 and H1.2 subtypes that display helical propensity also exhibiting significant built-

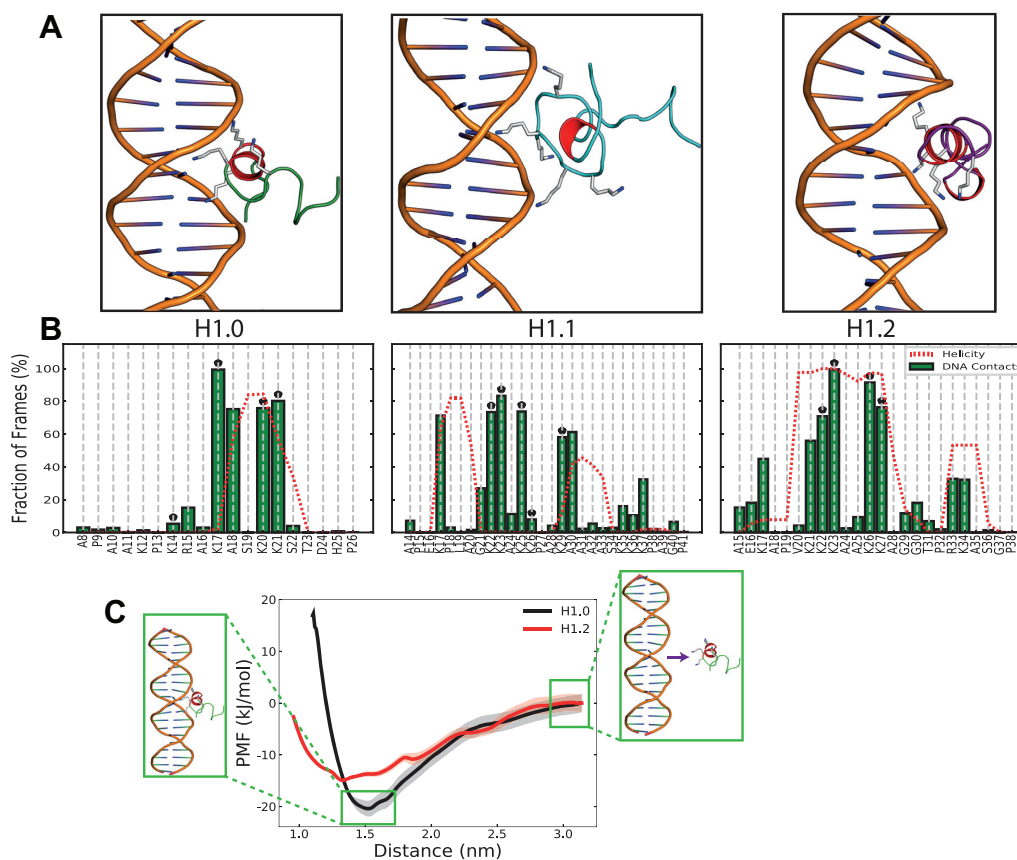


Figure 3. Interactions of the induced helical conformations with DNA. (A) Lowest energy docked conformations of the induced helical conformations with a linear 20-bp β -DNA strand. The α -helical motifs are in red and the Lys within the basic face are illustrated. (B) Interactions of specific residues within the NTD with DNA from the final 250 ns of simulations of the docked conformations. A contact was assumed if a non-hydrogen atom of the residue was within 3.2 Å of a non-hydrogen DNA atom. The average per-residue helicity during this final 250 ns is shown in red. The Lys residues within the basic face of each H1 NTD are marked with a black '*'. (C) PMF profiles for the binding of the H1.0 (black) and H1.2 (red) with the DNA major groove. The standard deviations of the calculated PMFs estimated from the bootstrapping are shown as shaded regions. Snapshots illustrate the initial and final window configurations for the H1.0 variant and the radial distance used as the CV (purple arrow). The energy profiles are shifted so that the energy at large distances is set to zero.

in amphiphilicity and charge concentration within the basic 'face' (green curve) in contrast to the unstructured H1.1.

H1 NTD also undergoes a disorder-to-order transition upon DNA binding

We now focus on determining if the charge-neutralization-induced helical conformations observed above are indeed reproducible when charge repulsion among Lys in the H1 NTD is instead screened via DNA binding. For this, we start by using HADDOCK (81) to rigid-body dock a NTD configuration representative of either the second minima (for the H1.0 and H1.2 subtypes) or the first minimum (for the H1.1 subtype) from our PTMetaD-WTE simulations (on the neutralized peptides) to a 20-bp ds-DNA segment (see Materials and Methods). Before the docking, we restore the full charges on the NTD Lys. In the case of H1.0 and H1.2, we consider the Lys residues within the amphiphathic face (Figure 2C, E) as active residues for the docking interaction, in the case of H1.1 use K22, K23, K25, K26 and K27, as they are within a region with the highest charge density. Figure 3A shows the lowest energy HAD-

DOCK cluster for the three subtypes. For the H1.0 and H1.2 subtypes, docking predicts the helix regions to fit snugly within the DNA major groove (the four lowest energy clusters were similar for all three cases), facilitating interactions between the four Lys within the basic faces and DNA. Binding of helices to the major groove is amongst the most common DNA-protein interaction motifs with such interfaces present within helix-turn-helix proteins, zinc-binding proteins and leucine zipper proteins (96). In the case of H1.1 NTD docked configuration instead, there are only three basic residues (all within the unstructured region) that interact with DNA around the minor groove.

To assess the viability of the docked conformations and to investigate the implications of NTD differential folding across subtypes on DNA interactions, we run unbiased MD simulations of the DNA-docked NTD basic subregion systems with fully charged Lys sidechains. Our results confirm that charge-screening by DNA binding favours disorder-to-order transitions in all the H1 NTD subtypes studied. When interacting with the DNA strand, the induced helical configuration in the H1.0 NTD is stable across the 400 ns MD simulation (Supplementary Figure S12), with residues in the

basic face (K17, K20, K21) of the helix accounting for the majority of interactions with the DNA (Figure 3B).

Simulations of the DNA-bound H1.2 NTD also show a prominent disorder-to-order transition upon DNA binding, with region ₁₉PVKKKAAKKAGG₃₀ retaining the stable Lys-rich helix throughout the course of the simulation (see Supplementary Figure S12). Here, the four Lys K22, K23, K26 and K27 residues fall on the same side of the helix, making two contiguous helical turns that interact strongly with the DNA (Figure 3B), mainly at the major groove. A secondary small set of residues (₃₃RKA₃₅) with helical properties can also be observed. However, this secondary motif is too short to form a full helix and correlates with less prominent NTD-DNA interactions.

Contrary to the neutral-Lys PTMetaD-WTE simulations, the H1.1 NTD also forms two short helices (₁₇KPLA₂₀ and ₂₉KAAAA₃₃) that are joined together by a Lys-rich unstructured loop (₂₁GKKAKKPA₂₈). Because the helices formed here are mainly devoid of charged residues, DNA interactions are instead mediated by the unstructured loop and the single charged residues at the edge of each helix. Enhanced fluidity of the binding region is consistent with the H1.1 NTD showing a reduced number of DNA-bound states (~83% of the simulation frames) than the H1.0 and H1.2 cases (100%), where rather than focusing on the DNA major groove, the NTD promiscuously binds/unbinds the entire DNA strand (Supplementary Figure S13). Although the specific DNA interaction patterns of the disordered H1.1 segment depend on the initial configurations generated by HADDOCK's rigid body docking algorithm, our results (Figure 3B, Supplementary Figure S13) demonstrate a distinction between the promiscuous interactions of disordered segments and the stable interactions of structured segments.

To quantify the impact of this observed subtype-specific helical propensity on the protein's DNA binding ability, we took the two partially structured H1 NTD subtypes (i.e., H1.0 and H1.2) and, for each of them, independently simulated the binding process to a 20-bp ds-DNA strand by means of umbrella sampling simulations (see Methods). From the umbrella sampling simulations, we calculate the potential of mean force (PMF) as a function of the radial protein-DNA distance normal to the DNA's helical axis (Figure 3C). The PMFs reveal that while the subtype with highest helical propensity, the H1.0 NTD, is stabilized by ~20 kJ/mol (~8 $k_B T$) upon DNA binding, the longer H1.2 NTD—with a decreased helical propensity—is stabilized by only ~15 kJ/mol (~6 $k_B T$) upon DNA binding. However, it should be noted that within the force fields, interactions between the oppositely charged amine and phosphate groups have not been explicitly calibrated (97). Thus, we repeated the PMF calculations with the Amberff14SB (98) and parmbsc1 (99) force fields together with the complete set of non-bonded fix (cufix) corrections (100). While the inclusion of the non-bonded amine-phosphate interaction corrections modified the free energy profile at short distances, the depth of the resulting well is equivalent (~20 kJ/mol) (Supplementary Figure S14). These results suggest that a higher tendency of the H1 NTD subtypes to undergo a disorder-to-order transition (favoured by a shorter length

and higher charge density) enhances their DNA binding propensity.

Finally, to further investigate the stability of the disorder-to-order NTD transition, we performed metadynamics simulations of the basic subregion of the H1.0 NTD (with normally charged Lys sidechains) in contact with a 20-bp ds-DNA strand, starting from an unstructured NTD configuration. We used two CVs: the alpha helicity (S_α) and the number of electrostatic contacts (S_{cont}) between the phosphate backbone and Lys/Arg sidechains (see Methods). Convergence was assessed using the time-evolution of the free-energy profiles (Supplementary Figure S15).

The simulations predict that upon DNA binding the NTD transitions from the unstructured configuration ($S_\alpha = 0$, Figure 4) to a helical state; that is, the global minimum appears at $S_\alpha \sim 5.8$ and $S_{\text{cont}} \sim 12.5$, which corresponds to the helical peptide bound to DNA (Figure 4B, Supplementary Figure S16). The resulting helix spans a larger region (₁₃PKRAKASKKST₂₃ in Figure 4C) than that predicted by our simulations with neutralized Lys sidechains and no DNA (₁₇KASKKS₂₂ in Figure 2B), suggesting that DNA binding effects are stronger than simple charge neutralization in triggering a disorder-to-order transition of the NTD. In addition, the interaction between the NTD and DNA emerges as non-specific with homogeneous contacts across the DNA strand (Supplementary Figure S17).

Implications of NTD disorder-to-order transition on H1 nucleosome binding

After having determined that the NTDs of H1 undergo a subtype-specific disorder-to-order transition upon DNA binding, we explore the implications of this structural transition for H1-nucleosome binding. We first verify if a similar transition is observed when the H1 NTD forms part of a full H1 protein (with both globular and CTD domain) that is bound to a nucleosome. For this, we compare the behaviour of an isolated H1.0 NTD peptide to that of the NTD of H1 in our recent 5- μs long Biased-Exchange Metadynamics simulation of the full H1.0 bound to a 211-bp chromatosome (29) (SI Discussion, Supplementary Table S4). For a direct comparison to the nucleosomal results, we additionally repeated the 'neutralized' PTMetaD-WTE simulations with the full length H1.0 NTD (in-contrast to only the basic subregion discussed in the preceding sections).

Figure 5B shows the per-residue structural propensity of the NTD within the nucleosome using multiple force-fields (Amber99SBildn, Amber ff03ws, Charmm36M (Supplementary Table S5)) and compares the results to those obtained from neutralized H1.0 segments. For all systems, the NTD displays a significant tendency to transition into a helical state. In comparison to the modelling of the shorter basic subregion (cyan, Figure 5B), modelling the entire NTD (dotted line, Figure 5B) allows the formation of a larger helical motif.

Within the nucleosome-bound systems, regardless of the force field, the helical motif is formed by the residues ₁₂KPKRAK₁₇ and mediates interactions with DNA (Supplementary Figure S18). Upon nucleosome binding, the exact region where the helix forms is shifted slightly away from

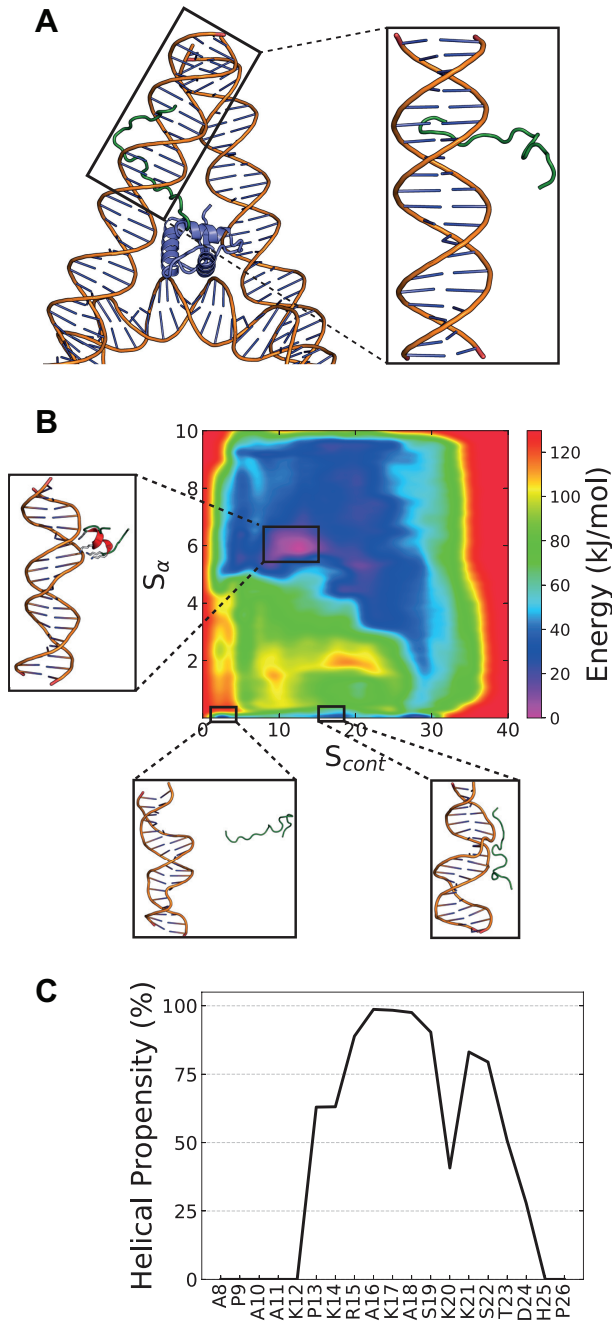


Figure 4. DNA induced secondary structure within the basic subregion of the H1.0 NTD. (A) Left: Constructing the initial configuration for metadynamics simulations – the NTD basic subregion (green) was built in an extended random coil configuration onto the on-dyad H1.0 globular domain (blue, PDB ID: 5NL0). Right: A 20-bp linker arm and the constructed NTD were then together considered the initial configuration. (B) Two-dimension free energy profiles of the NTD when interacting DNA along the two CVs – S_{α} and S_{cont} . Three different configurations are illustrated: (1) the primary energy minima with a helical motif when bound to DNA, (2) unstructured and unbound and (3) DNA bound but unstructured. (C) The reweighted per-residue helical content of the H1.0 NTD basic subregion when interacting with DNA.

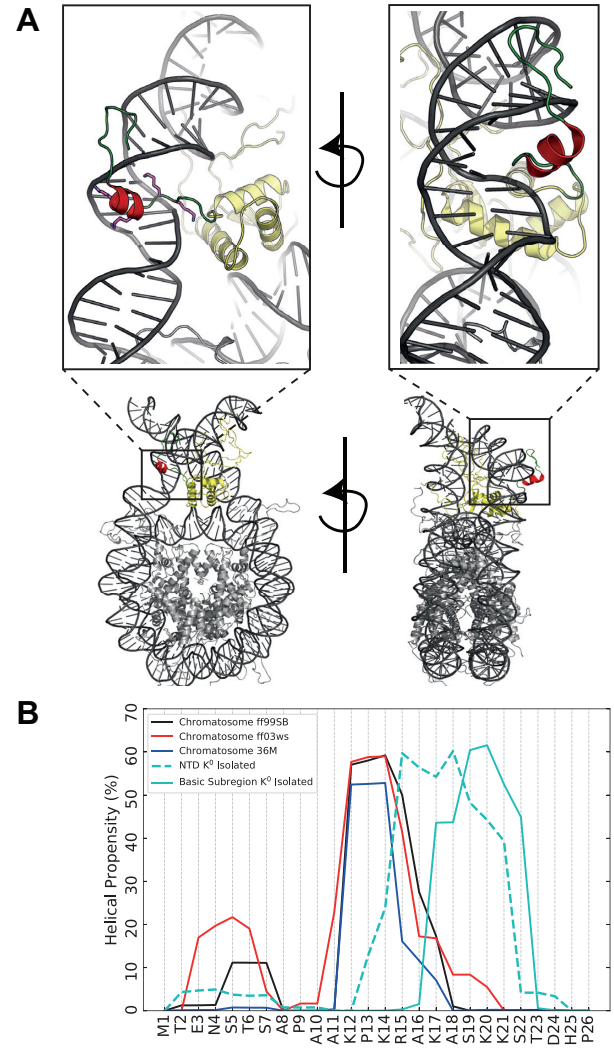


Figure 5. The N-terminal domain within the nucleosome. (A) Illustrations of the most-populated cluster of the NTD from 5- μ s BEMD simulations of the full-length hH1.0 within a 211-bp nucleosome. The globular-domain is shown in yellow while the helical-motif within the NTD is in red. The four residues that constitute the basic ‘face’ from the helical-wheel analysis are illustrated. (B) Comparison of the per-residue helicity of the NTD within the nucleosome across different force fields (ff99SB-ildn, ff03ws, Charmm36M (SI Discussion)) and when isolated with neutralized Lys sidechains.

the point of attachment of the NTD to the globular domain, when compared to the helix in isolated NTD peptide. This difference may be plausibly attributed to the NTD within the nucleosome-bound system being subjected to additional configurational restraints imposed by its attachment to the globular domain of H1.

What is the meaning of these results for the binding of H1 to the nucleosome? Several different classifications for the relative nucleosomal binding affinities of the various H1 subtypes have been proposed (see Table 2). Despite variance across experimental methods, most studies have classified H1.0 as possessing the highest nucleosome binding affinity, followed by H1.2 and finally H1.1 (see Table 2). This trend is also consistent with H1.0 subtype overexpression being associated with quiescent chromatin/repressed gene expres-

Table 2. Summary of the experimental nucleosomal binding affinities of the linker histone subtypes. Experiment 5 only ranked the subtypes into two affinity categories. Experiments 6 and 7 only compared two specific subtypes and thus have no ‘intermediate’ ranking

ID	Low	Intermediate	High	Reference
1	H1.1, H1.10	H1.0, H1.2, H1.3	H1.4, H1.5	Clausell <i>et al.</i> (15)
2	H1.1	H1.2, H1.5	H1.0, H1.3, H1.4	Orrego <i>et al.</i> (104)
3	H1.1, H1.2	H1.3, H1.4, H1.5	H1.0	Flanagan <i>et al.</i> (23)
4	H1.1, H1.2	H1.0, H1.3	H1.4, H1.5	Th’ng <i>et al.</i> (16)
5	H1.2, H1.3	-	H1.4, H1.0	Oberg <i>et al.</i> (105)
6	H1.2	-	H1.0	George <i>et al.</i> (21)
7	H1.2	-	H1.0	Vyas <i>et al.</i> (32)

sion (101), and the H1.2 and H1.1 subtypes localizing in transcriptionally active euchromatic regions (16). Remarkably, the observed binding affinities vary congruently with the ability of the different H1 NTD subtypes to form helical conformations upon charge-neutralization. Hence, it is plausible to propose that the NTD contributes to the differential nucleosomal binding of the subtypes through differences in its conformational preferences when within the nucleosome. That is, the higher nucleosome binding affinity of the H1.0 subtype might be in part due to the H1.0 NTD being able to form a stable amphipathic helix that increases the stability of the DNA-NTD complex, helping ‘glue’ the H1 to the nucleosome. Conversely, the longer and less-positive H1.1 NTD is disordered, which in-turn might result in weaker, fluctuating, and promiscuous interactions with the nucleosomal DNA (Figure 3B). Our simulations thus suggest a molecular mechanism contributing to explain the differential nucleosome binding affinities of the H1 subtypes.

CONCLUSIONS

We have used a combination of microsecond-long T-REMD, PTMetaD-WTE and BEMD simulations to sample the conformational landscape of the NTD of H1 subtypes H1.0, H1.1 and H1.2. When free in solution, our simulations predict that all three NTDs would predominantly adopt unstructured but partially collapsed conformations. Interestingly, when the Lys sidechains are neutralized either artificially by increasing the pH or physiologically by binding to DNA, the simulations show that reduced inter-sidechain repulsion is expected to be sufficient to induce a subtype-specific disorder-to-order transition. However, DNA-binding has a stronger effect than simple charge-reduction at triggering such transition. The ordered NTD structures in subtypes H1.0 and H1.2 are amphipathic helical conformations that orient the basic residues of the NTD along one ‘face’ of the helix, facilitating electrostatic interactions with the DNA major groove along that face which drives the peptide to a folded conformation acting as an ‘anchor’ between the DNA and H1 as previously hypothesized (102). By combining docking, unbiased MD simulations, and metadynamics simulations we show that the NTD conformational preferences translate into different patterns of DNA/NTD interactions and, hence, a differential ability of the different subtypes to mediate interactions of the H1 and the nucleosome.

Our simulations further reveal that the different charge density per unit length, the total length of the NTD, and interspersing of ‘helix-breaking’ Pro/Gly residues across

subtypes, give rise to different degrees of helicity of the NTD; thereby, shorter charge-dense positive peptides devoid of Pro/Gly residues (H1.0 and H1.2) emerge as optimal for helical folding. Importantly, the varying degree of helicity upon DNA binding correlates with the experimentally observed binding affinities of the specific H1 subtypes, and the differences in the free energy stabilization of the various DNA-NTD complexes, with higher affinity subtypes exhibiting more prominent helical conformations. Specifically, upon charge neutralisation, the strongest binding H1.0 subtype forms a single contiguous helix spanning four turns, while the weakest binding H1.1 displays no helicity. This computationally predicted structural mechanism correlates with the experimentally observed effects of the H1.0/H1.2 NTD domain-swap that resulted in an exchange of their binding characteristics (32).

Our simulations postulate how subtle sequence differences in chromatin binding proteins can have a significant impact in their structural behaviour, and subsequently, their nucleosome binding affinity. Our results suggest a mechanism by which the structural behaviour of short charged-rich disordered domains of chromatin binding proteins could be crucially transformed after nucleosome binding. Prospectively, our observations can be exploited in the design of novel experiments to assess the structural preferences of the NTD of H1 *in vitro* and enhance our molecular understanding of the link between H1 NTD structure and the different roles of the various H1 subtypes *in vivo* (9). Our work also suggests that charge-reducing post-translational modifications like Lys acetylation (38,41) and Arg citrullination (103) can strongly impact the structural behaviour of the histone’s terminal domains, especially those that are short and highly enriched in positively charged amino acids.

DATA AVAILABILITY

The initial configurations, temperature distributions, docking results, PLUMED input files and GRO-MACS .mdp are available at the Cambridge Research repository (<https://doi.org/10.17863/CAM.46225>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Brian Jimenez-Garcia for help with the HADDOCK docking.

FUNDING

European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [803326]; R.C.-G. is an advanced fellow from the Winton Programme for the Physics of Sustainability; Red Española de Supercomputación [QCM-2018-3-0041, QCM-2018-2-0037]; Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (<http://www.hpc.cam.ac.uk>) funded by EPSRC Tier-2 capital grant [EP/P020259/1]. Funding for open access charge: ERC starting grant.

Conflict of interest statement. None declared.

REFERENCES

- Olins,A.L. and Olins,D.E. (1974) Spheroid chromatin units (v bodies). *Science*, **183**, 330–332.
- Kornberg,R.D. and Lorch,Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, **98**, 285–294.
- Zhou,B.R., Jiang,J., Feng,H., Ghirlando,R., Xiao,T.S. and Bai,Y. (2015) Structural mechanisms of nucleosome recognition by linker histones. *Mol. Cell*, **59**, 628–638.
- Bednar,J., Garcia-Saez,I., Boopathi,R., Cutter,A.R., Papai,G., Reymer,A., Syed,S.H., Lone,I.N., Tonchev,O., Crucifix,C. et al. (2017) Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. *Mol. Cell*, **66**, 384–397.
- Simpson,R.T. (1978) Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, **17**, 5524–5531.
- Thoma,F., Koller,T. and Klug,A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.*, **83**, 403–427.
- Shen,X., Yu,L., Weir,J.W. and Gorovsky,M.A. (1995) Linker histones are not essential and affect chromatin condensation in vivo. *Cell*, **82**, 47–56.
- Öztürk,M.A., Cojocaru,V. and Wade,R.C. (2018) Dependence of chromatosome structure on linker histone sequence and posttranslational modification. *Biophys. J.*, **114**, 2363–2375.
- Hagop Parseghian,M. (2015) What is the role of histone H1 heterogeneity? A functional model emerges from a 50 year mystery. *AIMS Biophys.*, **2**, 724–772.
- Happel,N. and Doenecke,D. (2009) Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene*, **431**, 1–12.
- Fyodorov,D.V., Zhou,B.R., Skoultchi,A.I. and Bai,Y. (2017) Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.*, **19**, 192–206.
- Roque,A., Ponte,I. and Suau,P. (2016) Interplay between histone H1 structure and function. *Biochim Biophys Acta.*, **1859**, 444–454.
- Talbert,P.B., Ahmad,K., Almouzni,G., Ausiá,J., Berger,F., Bhalla,P.L., Bonner,W.M., Cande,W.Z., Chadwick,B.P., Chan,S.W.L. et al. (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenet. Chromatin*, **5**, 7.
- Roque,A., Ponte,I. and Suau,P. (2017) Post-translational modifications of the intrinsically disordered terminal domains of histone H1: effects on secondary structure and chromatin dynamics. *Chromosoma*, **126**, 83–91.
- Clausell,J., Happel,N., Hale,T.K., Doenecke,D. and Beato,M. (2009) Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS One*, **4**, e0007243.
- Th'ng,J.P.H., Sung,R., Ye,M. and Hendzel,M.J. (2005) H1 family histones in the nucleus: control of binding and localization by the C-terminal domain. *J. Biol. Chem.*, **280**, 27809–27814.
- Daujat,S., Zeissler,U., Waldmann,T., Happel,N. and Schneider,R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J. Biol. Chem.*, **280**, 38090–38095.
- Hale,T.K., Contreras,A., Morrison,A.J. and Herrera,R.E. (2006) Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1 α . *Mol. Cell*, **22**, 693–699.
- Allan,J., Hartman,P.G., Crane-Robinson,C. and Aviles,F.X. (1980) The structure of histone H1 and its location in chromatin. *Nature*, **288**, 675–679.
- Hartman,P.G., Chapman,G.E., Moss,T. and Bradbury,E.M. (1977) Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur. J. Biochem.*, **77**, 45–51.
- George,E.M., Izard,T., Anderson,S.D. and Brown,D.T. (2010) Nucleosome interaction surface of linker histone H1c is distinct from that of H10. *J. Biol. Chem.*, **285**, 20891–20896.
- Öztürk,M.A., Pachov,G.V., Wade,R.C. and Cojocaru,V. (2016) Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Res.*, **44**, 6599–6613.
- Flanagan,T.W., Files,J.K., Casano,K.R., George,E.M. and Brown,D.T. (2016) Photobleaching studies reveal that a single amino acid polymorphism is responsible for the differential binding affinities of linker histone subtypes H1.1 and H1.5. *Biology Open*, **5**, 372–380.
- Izzo,A. and Schneider,R. (2016) The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochim Biophys Acta.*, **1859**, 486–495.
- Brown,D.T., Izard,T. and Misteli,T. (2006) Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat. Struct. Mol. Biol.*, **13**, 250–255.
- Clark,D.J., Hill,C.S., Martin,S.R. and Thomas,J.O. (1988) Alpha-helix in the carboxy-terminal domains of histones H1 and H5. *EMBO J.*, **7**, 69–75.
- Turner,A.L., Watson,M., Wilkins,O.G., Cato,L., Travers,A., Thomas,J.O. and Stott,K. (2018) Highly disordered histone H1–DNA model complexes and their condensates. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 11964–11969.
- Borgia,A., Borgia,M.B., Bugge,K., Kissling,V.M., Heidarrson,P.O., Fernandes,C.B., Sottini,A., Soranno,A., Buholzer,K.J., Nettels,D. et al. (2018) Extreme disorder in an ultrahigh-affinity protein complex. *Nature*, **555**, 61–66.
- Sridhar,A., Farr,S.E., Portella,G., Schlick,T., Orozco,M. and Collepardo-Guevara,R. (2020) Emergence of chromatin hierarchical loops from protein disorder and nucleosome asymmetry. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 7216–7224.
- Oberg,C. and Belikov,S. (2012) The N-terminal domain determines the affinity and specificity of H1 binding to chromatin. *Biochem. Biophys. Res. Commun.*, **420**, 321–324.
- Henzel,M.J., Lever,M.A., Crawford,E. and Th'ng,J.P.H. (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J. Biol. Chem.*, **279**, 20028–20034.
- Vyas,P. and Brown,D.T. (2012) N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isoforms H10 and H1c. *J. Biol. Chem.*, **287**, 11778–11787.
- Okuwaki,M., Abe,M., Hisaoka,M. and Nagata,K. (2016) Regulation of cellular dynamics and chromosomal binding site preference of linker histones H1.0 and H1.X. *Mol. Cell Biol.*, **36**, 2681–2696.
- Baker,C.M. and Best,R.B. (2014) Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation. *Wiley Interdiscipl. Rev.: Comput. Mol. Sci.*, **4**, 182–198.
- Vila,R., Ponte,I., Collado,M., Arrondo,J.L.R., Jiménez,M.A., Rico,M. and Suau,P. (2001) DNA-induced α -helical structure in the NH2-terminal domain of histone H1. *J. Biol. Chem.*, **276**, 46429–46435.
- Liu,Z. and Huang,Y. (2014) Advantages of proteins being disordered. *Protein Sci.*, **23**, 539–550.
- Berlow,R.B., Dyson,H.J. and Wright,P.E. (2015) Functional advantages of dynamic protein disorder. *FEBS Lett.*, **589**, 2433–2440.
- Collepardo-Guevara,R., Portella,G., Vendruscolo,M., Frenkel,D., Schlick,T. and Orozco,M. (2015) Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Am. Chem. Soc.*, **137**, 10205–10215.

39. Potoyan, D.A. and Papoian, G.A. (2011) Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.*, **133**, 7405–7415.
40. Potoyan, D.A. and Papoian, G.A. (2012) Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *PNAS*, **109**, 17857–17862.
41. Winogradoff, D., Echeverria, I., Potoyan, D.A. and Papoian, G.A. (2015) The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects. *J. Am. Chem. Soc.*, **137**, 6245–6253.
42. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B.L., Grubmüller, H. and MacKerell, A.D. (2016) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, **14**, 71–73.
43. Best, R.B., Zheng, W. and Mittal, J. (2014) Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.*, **10**, 5113–5124.
44. Bonomi, M. and Parrinello, M. (2010) Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.*, **104**, 190601.
45. Barducci, A., Bonomi, M. and Parrinello, M. (2011) Metadynamics. *Wiley Interdiscipl. Rev.: Comput. Mol. Sci.*, **1**, 826–843.
46. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2004) UniProt: the universal protein knowledge base. *Nucleic Acids Res.*, **32**, 115–119.
47. Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchison, G.R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.*, **4**, 1758–2946.
48. Onufriev, A., Bashford, D. and Case, D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Genet.*, **55**, 383–394.
49. Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19–25.
50. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. and Bussi, G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.*, **185**, 604–613.
51. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct. Funct. Bioinform.*, **78**, 1950–1958.
52. Best, R.B. and Hummer, G. (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, **113**, 9004–9015.
53. Best, R.B., De Sancho, D. and Mittal, J. (2012) Residue-specific α -helix propensities from molecular simulation. *Biophys. J.*, **102**, 1462–1467.
54. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
55. Åqvist, J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.*, **94**, 8021–8024.
56. Dang, L.X. (1995) Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. *J. Am. Chem. Soc.*, **117**, 6954–6960.
57. Beglov, D. and Roux, B. (1994) Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.*, **100**, 9050–9063.
58. Luo, Y. and Roux, B. (2010) Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett.*, **1**, 183–189.
59. Venable, R.M., Luo, Y., Gawrisch, K., Roux, B. and Pastor, R.W. (2013) Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B*, **117**, 10183–10192.
60. Hart, K., Foloppe, N., Baker, C.M., Denning, E.J., Nilsson, L. and MacKerell, A.D. (2012) Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.*, **8**, 348–362.
61. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
62. Miyamoto, S. and Kollman, P.A. (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, **13**, 952–962.
63. Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.
64. Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
65. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
66. Patriksson, A. and van der Spoel, D. (2008) A temperature predictor for parallel tempering simulations. *PCCP*, **10**, 2073–2077.
67. Branduardi, D., Gervasio, F.L., Cavalli, A., Recanatini, M. and Parrinello, M. (2005) The role of the peripheral anionic site and cation- π interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.*, **127**, 9147–9155.
68. Bussi, G., Gervasio, F.L., Laio, A. and Parrinello, M. (2006) Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.*, **128**, 13435–13441.
69. Bernetti, M., Masetti, M., Pietrucci, F., Blackledge, M., Jensen, M.R., Recanatini, M., Mollica, L. and Cavalli, A. (2017) Structural and kinetic characterization of the intrinsically disordered protein SeV N-TAIL through enhanced sampling simulations. *J. Phys. Chem. B*, **121**, 9572–9582.
70. Camilloni, C., Schaal, D., Schweimer, K., Schwarzinger, S. and De Simone, A. (2012) Energy landscape of the prion protein helix 1 probed by metadynamics and NMR. *Biophys. J.*, **102**, 158–167.
71. Granata, D., Camilloni, C., Vendruscolo, M. and Laio, A. (2013) Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *PNAS*, **110**, 6817–6822.
72. Pietrucci, F. and Laio, A. (2009) A collective variable for the efficient exploration of protein beta-sheet structures: application to SH3 and GB1. *J. Chem. Theory Comput.*, **5**, 2197–2201.
73. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
74. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. and Beckstein, O. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**, 2319–2327.
75. McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J. and Pande, V.S. (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.*, **109**, 1528–1532.
76. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
77. Kolinski, A. and Skolnick, J. (1997) Determinants of secondary structure of polypeptide chains: Interplay between short range and burial interactions. *J. Chem. Phys.*, **107**, 953–964.
78. Ding, F., Jha, R.K. and Dokholyan, N.V. (2005) Scaling behavior and structure of denatured proteins. *Structure*, **13**, 1047–1054.
79. Tiwary, P. and Parrinello, M. (2015) A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B*, **119**, 736–742.
80. Bonomi, M., Barducci, A. and Parrinello, M. (2009) Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, **30**, 1615–1621.
81. Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastriitis, P.L., Karaca, E., Melquiond, A.S.J., Van Dijk, M., De Vries, S.J. and Bonvin, A.M.J.J. (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.
82. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
83. Wieczór, M. and Czub, J. (2017) How proteins bind to DNA: target discrimination and dynamic sequence search by the telomeric protein TRF1. *Nucleic Acids Res.*, **45**, 7643–7654.
84. Wieczór, M., Tobiszewski, A., Wityk, P., Tomiczek, B. and Czub, J. (2014) Molecular recognition in complexes of TRF proteins with telomeric DNA. *PLoS One*, **9**, e89460.

85. Hub, J., de Groot, B. and Van Der Spoel, D. (2010) g-wham - a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.*, **6**, 3713–3720.
86. Uversky, V.N., Oldfield, C.J. and Dunker, A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, **37**, 215–246.
87. Vila, R., Ponte, I., Jiménez, M.A., Rico, M. and Suau, P. (2002) An inducible helix-Gly-Gly-helix motif in the N-terminal domain of histone H1e: a CD and NMR study. *Protein Sci.*, **11**, 214–220.
88. Wang, L., Friesner, R.A. and Berne, B.J. (2011) Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B*, **115**, 9431–9438.
89. Stasevich, T.J., Mueller, F., Brown, D.T. and McNally, J.G. (2010) Dissecting the binding mechanism of the linker histone in live cells: An integrated FRAP analysis. *EMBO J.*, **29**, 1225–1234.
90. Böhm, L. and Mitchell, T.C. (1985) Sequence conservation in the N-terminal domain of histone H1. *FEBS Lett.*, **193**, 1–4.
91. Roque, A., Ponte, I. and Suau, P. (2009) Role of charge neutralization in the folding of the carboxy-terminal domain of histone H1. *J. Phys. Chem. B*, **113**, 12061–12066.
92. Agou, F., Waller, J.P., Yang, Y., Guittet, E. and Gesquière, J.C. (1995) Polyanion-induced α -Helical structure of a synthetic 23-residue peptide representing the lysine-rich segment of the N-terminal extension of yeast cytoplasmic aspartyl-tRNA synthetase. *Biochemistry*, **34**, 569–576.
93. Johnson, N.P., Lindstrom, J., Baase, W.A. and von Hippel, P.H. (1994) Double-stranded DNA templates can induce alpha-helical conformation in peptides containing lysine and alanine: functional implications for leucine zipper and helix-loop-helix transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 4840–4844.
94. Kim, M.K. and Kang, Y.K. (1999) Positional preference of proline in α -helices. *Protein Sci.*, **8**, 1492–1499.
95. Fisher, B.F., Hong, S.H. and Gellman, S.H. (2017) Helix propensities of amino acid residues via thioester exchange. *J. Am. Chem. Soc.*, **139**, 13292–13295.
96. Pabo, C.O. and Nekludova, L. (2000) Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.*, **301**, 597–624.
97. Yoo, J. and Aksimentiev, A. (2016) Improved parameterization of amine-carboxylate and amine-phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *J. Chem. Theory Comput.*, **12**, 430–443.
98. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.
99. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A. *et al.* (2015) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
100. Yoo, J. and Aksimentiev, A. (2018) New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *PCCP*, **20**, 8432–8449.
101. Brown, D.T., Alexander, B.T. and Sittman, D.B. (1996) Differential effect of H1 variant overexpression on cell cycle progression and gene expression. *Nucleic Acids Res.*, **24**, 486–493.
102. Allan, J., Mitchell, T., Harborne, N., Bohm, L. and Crane-Robinson, C. (1986) Roles of H1 domains in determining higher order chromatin structure and H1 location. *J. Mol. Biol.*, **187**, 591–601.
103. Christophorou, M.A., Castelo-Branco, G., Halley-Stott, R.P., Oliveira, C.S., Loos, R., Radziszewska, A., Mowen, K.A., Bertone, P., Silva, J.C.R., Zernicka-Goetz, M. *et al.* (2014) Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature*, **507**, 104–108.
104. Orrego, M., Ponte, I., Roque, A., Buschati, N., Mora, X. and Suau, P. (2007) Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol.*, **5**, 22.
105. Öberg, C., Izzo, A., Schneider, R., Wrangé, Ö. and Belikov, S. (2012) Linker histone subtypes differ in their effect on nucleosomal spacing in Vivo. *J. Mol. Biol.*, **419**, 183–197.