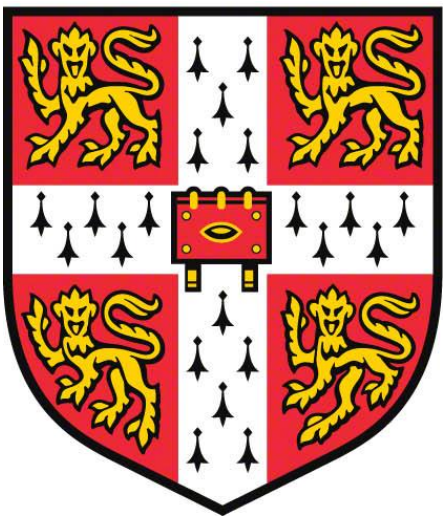
Intrinsically Disordered Proteins within the Genome



Akshay Sridhar

Department of Physics
University of Cambridge

A thesis submitted for the degree of
Doctor of Philosophy

Hughes Hall December 2020

Declaration

The work described in this thesis was carried out by the author in the Department of Physics at the University of Cambridge between September 2017 and July 2020. The contents are the original work of the author and contain nothing that is the outcome of collaboration. The contents have not previously or concurrently been submitted for any other degree or qualification at the University of Cambridge or another institution. The number of words does not exceed 60,000.

Akshay Sridhar
December 2020

Abstract

The hundreds of millions of DNA base-pairs within eukaryotic cells are not found free but packed inside the micrometre-sized nuclei through the formation of a macromolecular structure known as chromatin. Chromatin consists of a chain of nucleosomes – nucleoprotein complexes where the DNA makes $\sim$1.75 turns around a protein octamer core composed of two copies each of H2A, H2B, H3 and H4 histones. A fifth histone H1 binds on the nucleosomal surface close to the entry/exit site of DNA, interacts with linker DNA and aids in chromatin compaction. Enabling the condensation of DNA to fit into the nucleus is however only one-half of chromatin's role. The three-dimensional spatial organization of chromatin serves a second important role in allowing the capability to exert control over gene expression. The chromatin structure thus serves as an additional layer of complexity above the genome code and permits the transcription of different proteins varying with cell lineages/cycles.

The proteins that makeup, modify and read the chromatin structure are particularly enriched in 'Intrinsic Disorder' – a class of proteins lacking a well-defined structure but existing as a dynamic ensemble of rapidly interchanging states. While folded proteins with well-defined structures are amenable to be characterized through standard methods of protein structure determination, the 'plasticity' of disordered proteins challenges the use of such ensemble averaged techniques. In this thesis, Molecular Dynamics simulations are used to characterize the disordered regions of three proteins that form the core of chromatin structure: histones, linker histones (H1) and heterochromatin protein (HP1). The carboxy-terminal domain of H1 when within the nucleosome, adopts a compact but unstructured conformation that allows its positioning between the two linker DNA strands. In contrast, the amino-terminal domain of H1 undergoes a disorder-to-order transition to an amphiphilic helical conformation. The transition to the amphiphilic helix is however subtype-dependant with the degree of condensation varying with the subtypes' nucleosomal affinity. Finally, the simulations demonstrate that the affinity of HP1 subtypes for the H3 histone is caused by the synergetic effects of both the proteins' unstructured amino-terminal domain and the structured chromodomain.

Acknowledgements

Right at the start, I would have to thank my supervisor Dr. Rosana Collepardo-Guevara for help, guidance and patience over the course of the past four years. It has been a great pleasure working with a brilliant biochemist who simultaneously displays such unmatched dedication to her students. Meetings and discussions with her have always been very stimulating and have opened many avenues for the project. I am also thankful to her for the help in putting my work together both in this thesis and within the publication manuscripts. Finally, I must also appreciate her efforts in helping me through my career choices and the multiple postdoctoral fellowship applications.

Science is however never performed in a vacuum and the Cavendish Laboratory has been the ideal setting for me to meet, network and collaborate in an effort to widen my research horizons. I must particulary thank Jeffrey Gorman, Florian Auras and Sarah Orsborne from the Richard Friend group for allowing me to be involved in their DNA-coupling research. It has been a wonderful side-project that was simultaneously fun, novel and de-stressing to work on. Towards the final months of my doctoral work, I also had the pleasure of meeting and being involved with the fantastic microfluidics work being done by Georg Krainer, Timothy Welsh and Tuomas Knowles on IDPs. Working with MD quite often leads to tunnel-vision and I am grateful for being given this opportunity to appreciate the plethora of other research methodologies.

I must also acknowledge plenty of people who quite unselfishly offered their time to help me develop the research skills to carry out the work within this thesis – Ignacio Faustino for initially helping me when I was considering the use of Martini FF, Brian Jimenez-Garcia for teaching me the ropes of HADDOCK docking and Stephen Farr for opening my eyes on making wonderful visualizations of MD structures. Here, I must also thank Modesto Orozco for his ideas with the H1 modelling work and access to the computational resources at the Barcelona Supercomputing Centre.

Finally, I must thank my parents and Punicha, whose enduring encouragement and support motivates me to aspire towards greater heights of professional and personal achievement. Thank you all very much.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Contents

1.1 DNA, Histones and Nucleosomes

Deoxyribonucleic Acid (DNA) is the building block of life and carries information for the growth, development, functioning and reproduction of the organism [1]. It is primarily composed of four complementary subunits called nucleotides which are covalently linked together by a backbone of phosphodiester bonds. Under physiological conditions, DNA adopts a double helical structure [2] held together through complementary base pair hydrogen bonding and sequential nucleotide stacking. The sizeable nucleotide sequence required to completely encode genetic information (upto hundreds of millions in eukaryotic cells) is compressed within cells through complexation with an equal mass of proteins called histones.

The fundamental unit of this nucleo-protein macromolecular structure is called a nucleosome and is composed of a protein core around which ~1.7 turns of DNA is wrapped [3, 4]. The protein core is an octamer made up of four pairs of dimers of the H2A, H2B, H3 and H4 histone proteins [5]. These four 'core' histones are structurally similar [6] and possess a common 'histone-fold' motif consisting of three alpha helices linked by two loops (Figure 1.1) [7, 8].

Figure 1.1 The conserved 'histone-fold' motif of the four core histones. Adapted from the PDB structure **4QLC** [9].

In addition to the motif illustrated in Figure 1.1, the core histones contain highly basic unstructured regions rich in Lysine and Arginine residues known as histone tails [10]. While limited structural information about the histone tails are available, their positive charges are essential for the compaction of DNA [11–13] and have been identified to promote interactions with the DNA backbone, both within and across nucleosomes [14, 15]. Figure 1.2 illustrates the structure of wrapped DNA around the octamer core and the hitherto best resolved histone tail configuration [16].

A fifth histone (H1) binds on the nucleosomal surface close to the entry/exit site of DNA [17]. In addition to the core DNA, this protein interacts with

about 20 nucleotide base pairs of linker DNA [18] and is thus aptly titled linker histone.

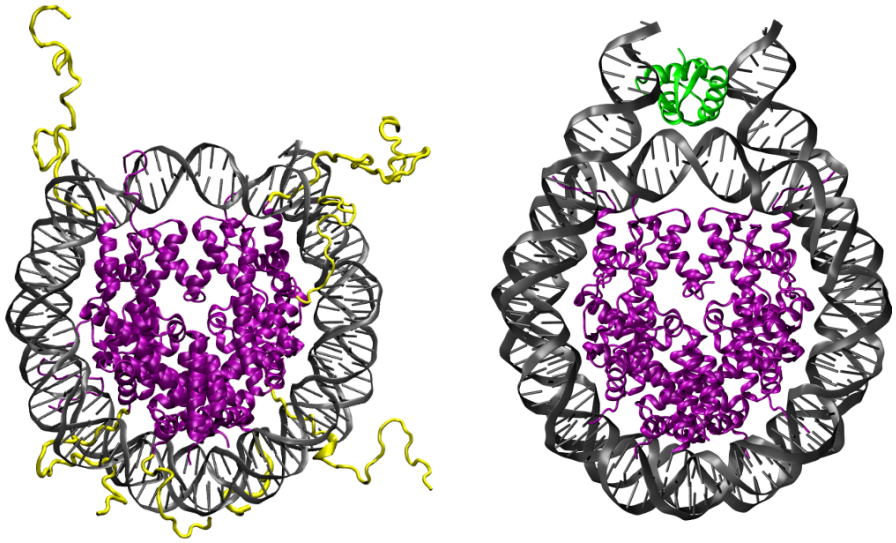


Figure 1.2 Left: The structure of the nucleosome with the DNA in grey and core histones in purple. The unstructured tails extending beyond the nucleosome are in yellow. PDB ID: 1KX5 [16]. **Right:** The structure of the linker histone (green) bound nucleosome. PDB ID: 4QLC [9]. Images visualized using VMD [19].

1.2 Chromatin Fiber Structure

Nucleosomal arrays of wrapped DNA, octamer core and H1 histone are joined by linker DNA to form 'extended' chromatin – a model often compared to 'beads-on-a-string'. Under physiological conditions, the extended 'beads-on-a-string' chromatin condenses into a secondary fibrous structure roughly 30 nm in diameter [20]. Tertiary condensations of this '30-nm fiber' form the highly recognizable metaphase structure of a chromosome. Figure 1.3 illustrates these hierarchical levels of DNA folding in association with histones.

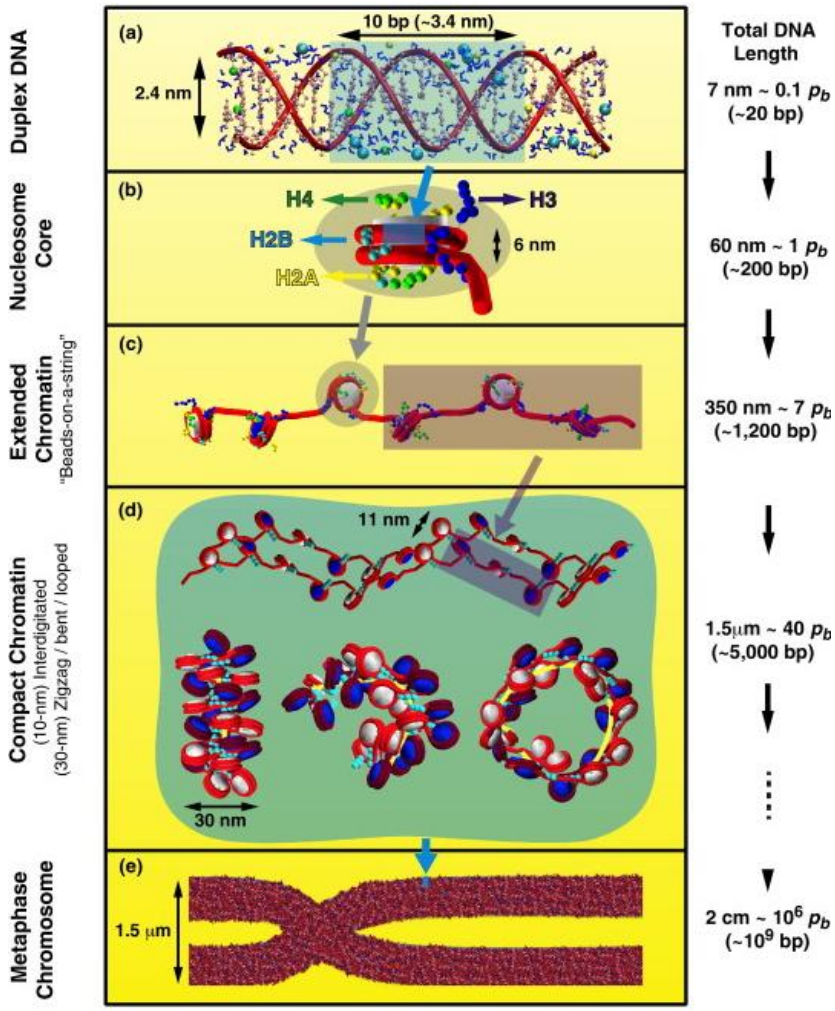


Figure 1.3 The hierarchical levels of DNA folding. **(a)** Simple double stranded DNA. **(b)** DNA wrapped around the octamer core of H2A, H2B, H3 and H4 histones. **(c)** Sequence of nucleosomes with linker DNA. **(d)** Condensed secondary structure of chromatin commonly referred as the '30 nm fiber'. **(e)** Tertiary condensation of chromatin fibers into chromosome. The length scales of each structure are mentioned on the right. Image from Ozer et al. [21].

The secondary '30 nm fiber' structural model was postulated to explain the formation of the easily visualizable tertiary chromosome structure (Figure 1.3e) from the primary nucleosomal array. However, despite considerable scientific effort, conclusive evidence regarding this secondary structure has remained elusive [22, 23]. Challenges in ascertaining this structure stem from its highly compact nature hindering the DNA folding paths from being visually inspected through electron microscopy [24]. Thus, *in situ* experiments with minimum perturbations to physiological conditions have been limited [25].

In vitro attempts at deciphering this structure have proposed two major hypotheses for the arrangement of nucleosomes in chromatin – the zigzag and solenoidal models. In the zigzag model, nucleosome core 'i' primarily interacts

with nucleosomes $i \pm 2$ [26]. In contrast, within the solenoidal model, the nucleosome 'i' interacts sequentially with cores $i \pm 1$ [27]. Figure 1.4 illustrates these differences in structure between the two models.

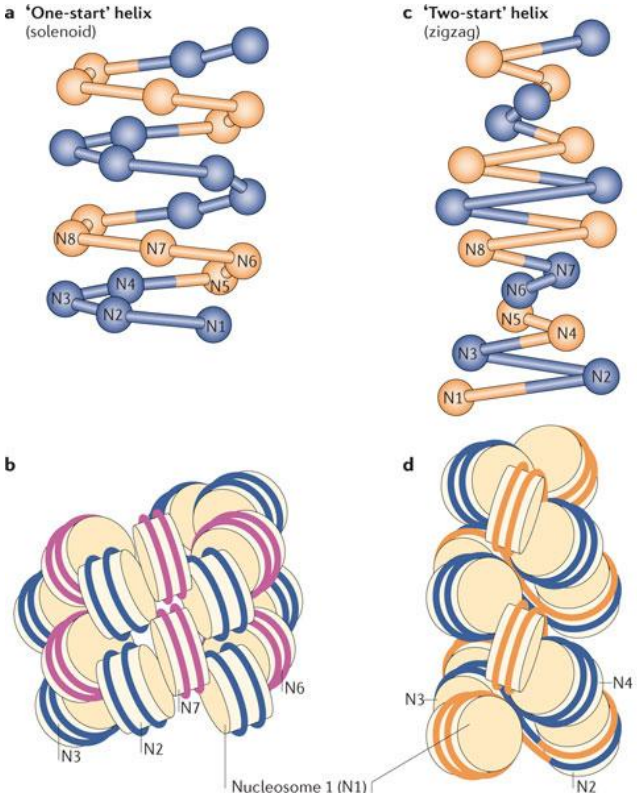


Figure 1.4 Illustrations of the solenoidal and zigzag models of chromatin structure. The nucleosomes are labelled sequentially to illustrate the inter-core interactions. **(a,b)** The solenoidal model. **(c,d)** The zigzag model. Image from Li et al. [28].

A range of experimental and computational studies have provided evidence in support of both models. Electron micrography of Williams et al. [29] supported the zigzag model with a direct correlation between the DNA length and chromatin width. Dorigo et al. [30] offered additional evidence in support of the zigzag model through their analysis of cysteine crosslinking in nucleosome cores. In 2005, Schalch et al. [31] further supported the zigzag model by crystallizing a compact high-resolution tetranucleosome structure. Despite these results, the zigzag model was deemed to insufficiently explain chromatin structure with the physiological nucleosome packing ratio nearly double of the maximal packing ratio possible under this model [32].

In contrast, the solenoidal model with the linker DNA oriented towards the centre of toroidal fiber can accommodate larger packing ratios closer to experimental results [33]. The solenoidal model was also supported by the electron microscopy results of Robinson et al. [24] and X-ray diffraction results of Widom et al. [34]. However, computational modelling studies by Perisic et

al. [35] have suggested the bending of the linker DNA between nucleosome cores to be energetically unfeasible and also hindered by linker histone binding. This lack of conclusive evidence in support of either the zigzag or solenoidal systematic arrangement models has led to the suggestion of chromatin adopting a fluid organization that permits interconversion between conformations [36, 37] – a hypothesis supported by the recent obesrvations of histone proteins to undergo liquid-liquid phase separation [38–40].

1.3 Chromatin and Genomic regulation

The above discussed condensation of DNA from a length of $\sim$2 m in stretched form to fit within a nucleus of 6 μm [21] is remarkable in itself. However, this considerable compression is only one half of chromatin's role. Highly compact chromatin regions are transcriptionally inactive [41, 42] and for genomic expression, this packed structure must unravel to allow transcription factors and RNA polymerases to access DNA [43]. The three-dimensional spatial organization of DNA and DNA-associated proteins thus serve a second and possibly more important role in allowing the capability to exert complex levels of control over gene expression [44–46]. The remodelling of the oligonucleosome structure thus allows additional complexity above the genetic code to permit the same DNA molecules to transcribe different proteins depending on cell lineages/cycles [47–55]. Such dynamic modifications of chromatin structure are carried out through two different mechanisms: (**1**) Restructuring of nucleosomes by ATP-dependent remodelling complexes and (**2**) Covalent amino and nucleic acid modifications by alteration specific enzymes [43, 56–58].

1.3.1 ATP-Dependent Remodelling

The nucleosome is the basic repeating unit of chromatin and the DNA within it is constrained through at least 14 salt-bridges with Arginine residues of the histone octamer [5, 16]. Adenosine-Tri-Phosphate (ATP)-dependent remodelling complexes use ATP hydrolysis to drive the modification of this constrained nucleosomal DNA and increase its accessibility [59, 60]. Within eukaryotes, at least five distinct families of remodelling complexes have been identified – switch/sucrose non-fermentable (SWI/SNF), imitation-switch (ISWI), chromodomain helicase DNA-binding (CHD), inositol-requiring 80 (INO80) and sick with RSC1 (SWR1) [61–64]. Although all remodelling complexes share a com-

mon ATPase domain, they are structurally diverse and distinct methodologies have been proposed for their action [65].

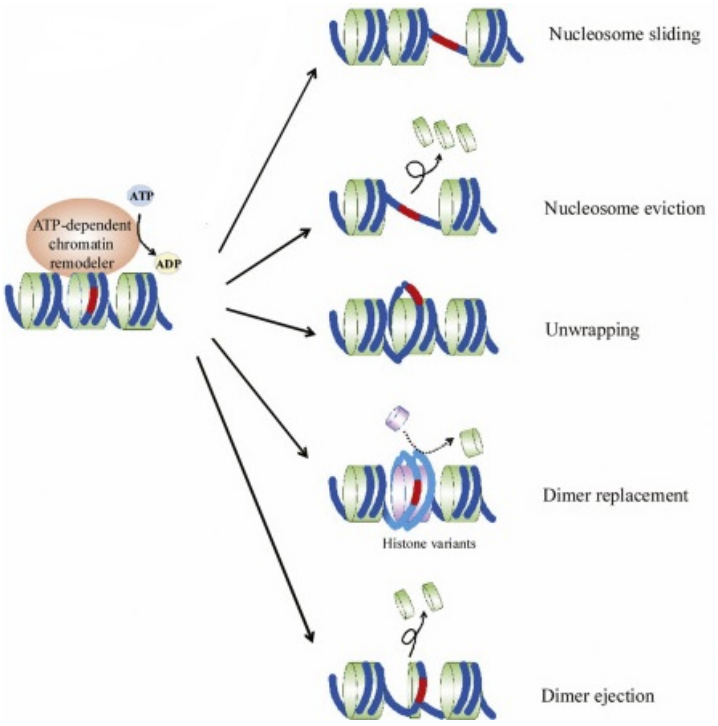


Figure 1.5 Illustrations of the proposed methodologies of ATP-dependent remodeller action. DNA segments made accessible through remodeller action are shown in red. Image from Chen et al. [66].

Figure 1.5 illustrates these proposed mechanisms which range from simple modulation of histone-DNA interactions like sliding/bulging to those of the INO80/SWR1 remodellers that completely disrupt the nucleosome through replacement/ejection of core histones [64–67].

1.3.2 Post-Translational Modifications

Post-translational modifications (PTM) are chemical changes to amino and nucleic acids performed by highly specific enzymes [55]. While an extremely wide range of PTMs have been identified in vivo including phosphorylation, acetylation, methylation, citrullination and ubiquitination [68–70], mass-spectrometric analyses have been identifying increasingly more novel modifications [71–73] and modification sites [74, 75]. Figure 1.6 illustrates one such PTM (citrullination) of Arginine where its sidechain is neutralized through the replacement of its guanidium group.

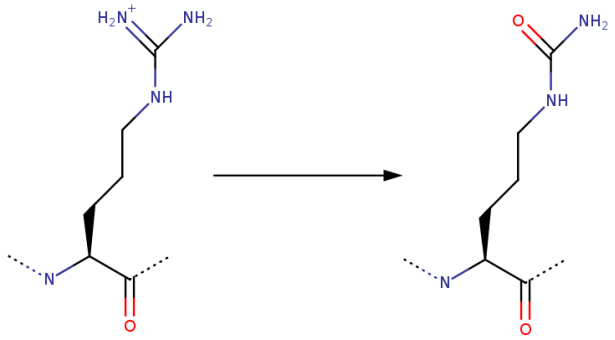


Figure 1.6 Schematic of Arginine citrullination performed in-vivo by PAD enzymes [76, 77]. The positively charge sidechain is neutralized through the replacement of the guanidium group. Image drawn using Marvinsketch [78]

The modifications can impact the structure of chromatin directly by shifting the interaction free-energy of the histone-DNA interface [79, 80]. For example, phosphorylation introduces negative charge and adds steric bulk which together can impede the binding of histones to the DNA backbone and thereby impede the free energy of nucleosome formation by up to 5 kcal/mol [81, 82]. Alternatively, the PTMs can impact chromatin structure indirectly by regulating the actions of the ATP-Dependent remodellers discussed in Section 1.3.1 [79, 80, 83]. For example, ISWI type remodellers preferentially bind to nucleosomes methylated at H3K4 and acetylated at H4K16 [84]. This importance and wide-reaching effects of PTMs have linked their malfunctioning to a range of diseases including cancer, Alzheimer's and multiple-sclerosis [68]. Figure 1.7 provides a summary of the PTM sites hitherto identified within core histones.

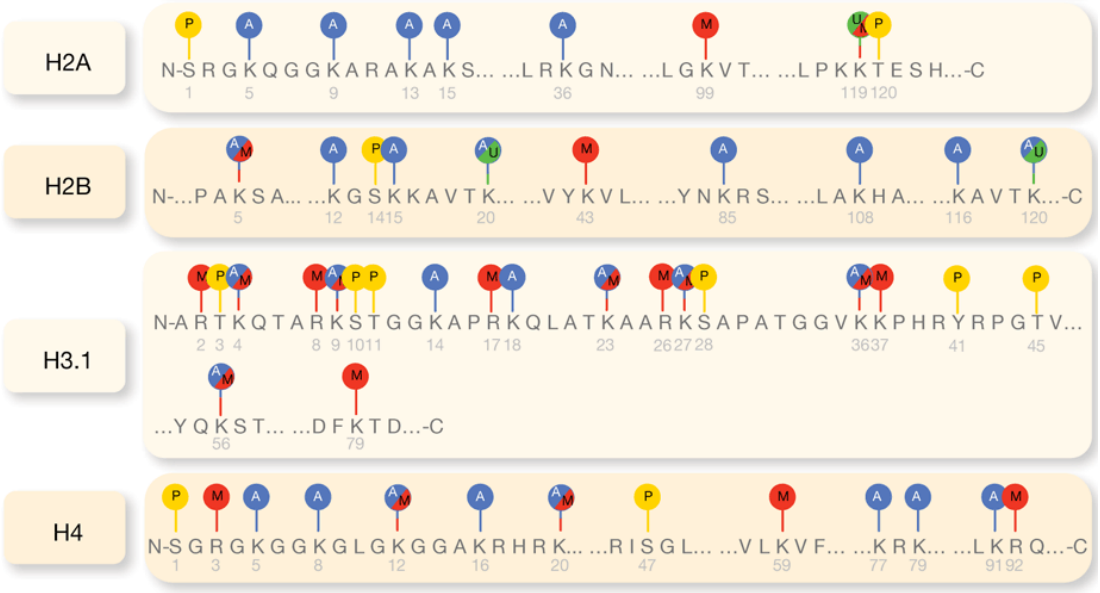


Figure 1.7 Summary of the post-translational modification sites observed in the H2A, H2B, H3 and H4 core histones. The phosphorylation sites are labelled 'P', ubiquitination 'U', methylation 'M' and acetylation 'A'. Image from Portela et al. [68].

1.4 Intrinsic Disorder

The plethora of proteins that makeup, modify and read chromatin structure are particularly enriched in 'intrinsic disorder' – a class of proteins lacking a well-defined secondary structure but existing as a dynamic ensemble of rapidly interchanging states (Figure 1.11) [85–87]. For example, histones that makeup the octamer core contain disordered tails that extend beyond the nucleosomal surface (Figure 1.2). Methyl- and acetyl-transferases that implement post-translational modifications contain disordered regions of up to 300 residues in length [88, 89]. In fact, over half of eukaryotic transcription factors have been identified over the past decade to possess significant disordered nature [90, 91].

As the name 'intrinsic' disorder suggests, the disordered nature is encoded within the peculiarities of the amino acid sequences of these proteins [92]. IDPs are typically depleted in bulky hydrophobic residues like Trp, Tyr, Cys, Ile, Phe, Val and Leu that normally form the core of folded globular proteins. Instead, they are enriched in small and charged residues like Pro, Ala, Gly, Lys, Glu and Asp that allow favourable interactions with the solvent environment [93–95]. In fact, some amino acids like Pro and Glu are nearly twice as likely to be present within IDPs than within resolved globular regions (Figure 1.8). This low sequence space associated with IDPs results in them commonly containing 'low-complexity' domains [96–98] characterized by repeating motifs [99].

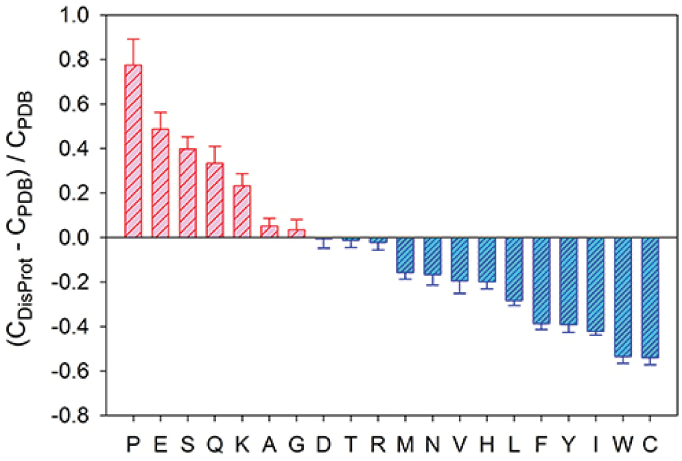


Figure 1.8 A comparison of the compositions of amino acids within disordered proteins ($C_{DisProt}$) and within the PDB database (C_{PDB}). Image from Uversky et al. [100].

Despite this low amino acid space and sequence simplicity associated with protein disorder, IDPs are not homogeneous, but are a heterogeneous mix of proteins that display varying levels of transient secondary structure and tertiary contacts [101, 102]. For example, the c-Myb oncogene transcription

factor displays transient helicity of up to 70% within its transactivation domain [103]. In contrast, the nucleosome-binding domain of the transcription activator CBP is completely disordered and can be described as a random coil [104]. Figure 1.9 illustrates this heterogeneous range of disorder displayed by IDPs ranging from nearly structured to completely random coil.

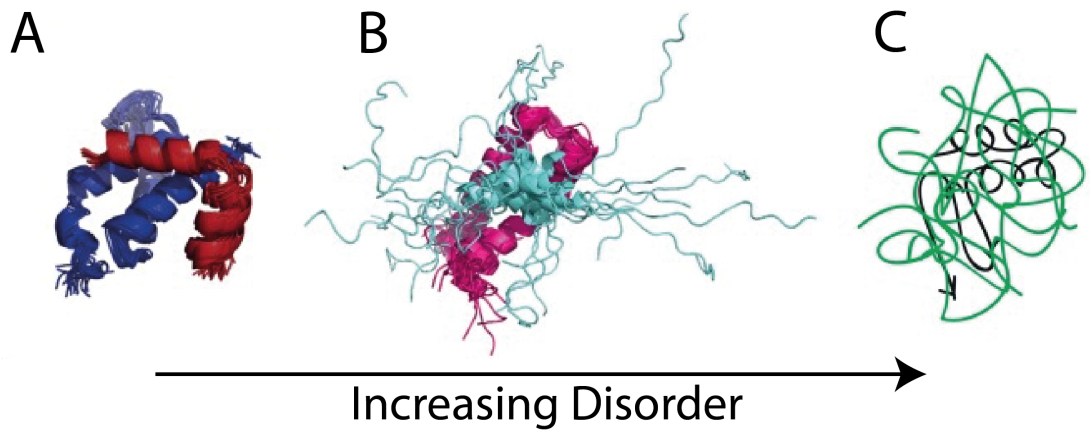


Figure 1.9 Illustration of the structural heterogeneity displayed by Intrinsically Disordered Proteins. **(A)** Disorder is limited to small segments flanking the structured regions and does not majorly affect the protein structure (PDB ID: 1KDX [105]). **(B)** The IDP displays transient α-helicity but not a well-defined tertiary structure (PDB ID: 2LPB [106]). **(C)** The IDP remains highly disordered throughout. Image from Shammas et al. [107].

1.5 Functional Roles of Disorder

IDPs with their lack of a well-defined three dimensional structure are in stark contrast to the structure-function paradigm of traditional structural biology [108, 109]. The functional role of this disorder has thus remained an open question with multiple mechanisms of action proposed for the heterogeneous IDPs [95].

1.5.1 Fly-Casting Mechanism

The 'fly-casting' mechanism hypothesizes the extended nature of IDP structures to afford them a large capture-radius that enables initial contact with binding partners at larger separations [110] (Figure 1.10A). From this large-separation contact, IDPs and their associated domains are 'reeled-in' to subsequently bind in their preferred orientation [95]. Functionally, this mechanism predicts kinetic advantage in binding for IDPs over their structured counterparts with a considerably smaller capture radius [111, 112] (Figure 1.10A).

1.5.2 Monkey-Bar Mechanism

The 'monkey-bar' mechanism is an extension of the fly-casting model and was initially proposed to account for the inter-DNA strand transfer of multi-domain proteins occurring at much faster timescales than those of single-domain proteins [113, 114]. In this proposed mechanism, the protein's disordered subdomain extends to recognize the second DNA strand while the protein's binding motif remains bound to the primary DNA strand. Subsequently, the disordered region 'swings' the protein's binding motif to the new DNA strand [107] (Figure 1.10B).

1.5.3 DNA Sliding Mechanism

The 'DNA-sliding' mechanism hypothesizes the role of disorder in acting as an 'entropic'-lock precluding the binding of proteins to non-specific DNA sequences [115]. Large proteins with multiple DNA-binding motifs can often be kinetically trapped to bind non-specific DNA sites through electrostatic interactions with the phosphate backbone. The entropic penalty of restraining disordered regions of the protein at such non-specific sites is hypothesized to allow the disassociation of the protein from such sites [116] and in this manner 'slide' along the DNA until reaching the enthalpically favoured cognate site [117].

1.5.4 Promiscuous Binding

In contrast to the mostly singular role of a majority of cellular proteins, chromatin-architectural proteins partake in a multitude of different processes [118]. The multifaceted capabilities of the proteins have been hypothesized to stem from the conformational plasticity of their IDRs [119] that allows them to achieve complementarity with a range of binding partners [120, 121]. For example, the cationic disordered regions of the linker histone forms complexes [122] with nucleic acids [123], the nucleosome [124], chaperone proteins [125, 126], and heterochromatic proteins [127] (Figure 1.10D). Promiscuous binding however requires the protein's disassociation from one complex before binding to the next – a characteristic hypothesized to be enabled by the IDP's low free energy of binding due to the entropic penalty of restraining disordered regions [95, 116]. In fact, analyses of datasets of protein-protein complexes have identified IDPs on-average to possess higher disassociation constants (K_d) than their structured counterparts [128, 129]. These differences are however only average and sev-

eral IDPs, especially in the genomic context, have been identified to possess extremely high binding affinities of up to the picomolar range [130–132] – in fact much higher than that of most structured proteins.

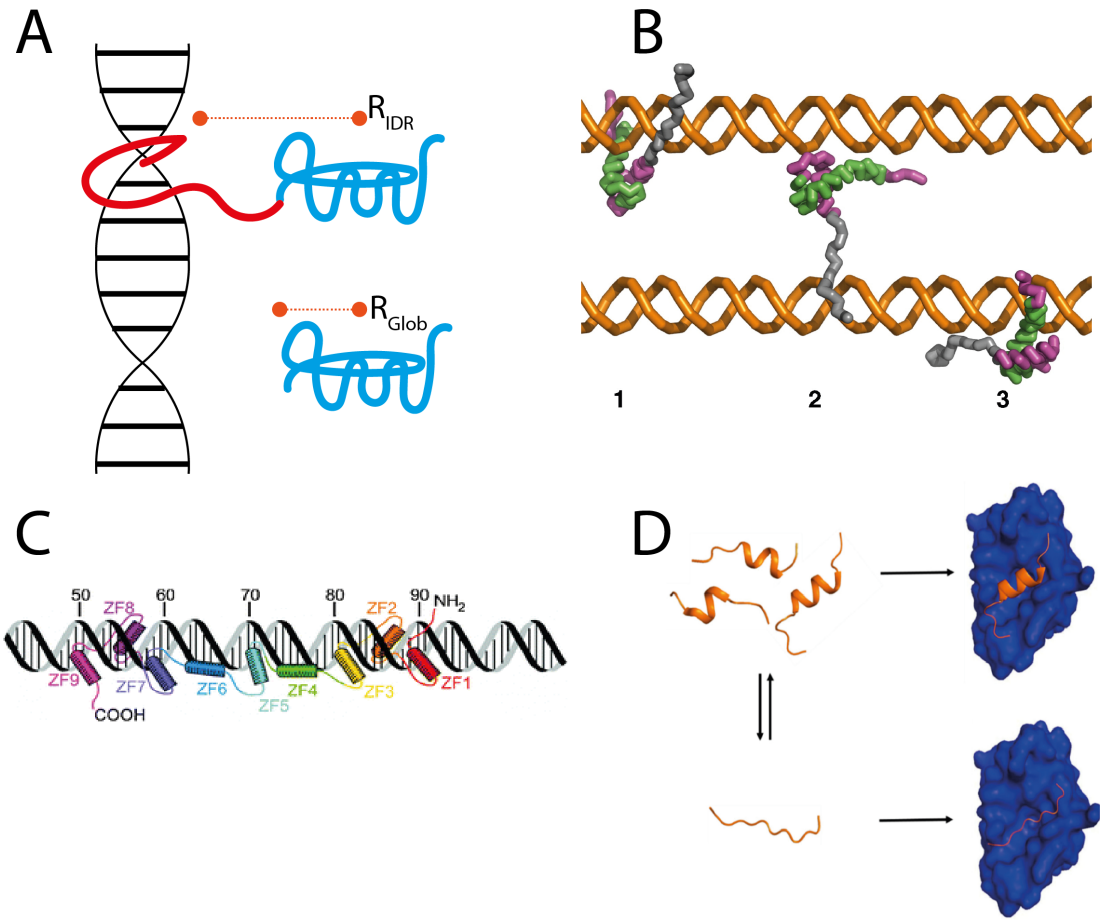


Figure 1.10 Illustration of the proposed functional advantages of Intrinsically Disorder Proteins. **(A)** Fly-Casting Mechanism. The disordered region of the protein (red) enable a larger capture radius (R_{IDP}) for the protein compared to globular proteins (R_{glob}). Image adapted from Shoemaker et al. [110]. **(B)** The monkey-bar mechanism for the transfer of proteins from one DNA strand to the adjacent. Image from Vuzman et al. [114]. **(C)** Schematic representation of the multi-zinc finger protein TFIIIA that binds specific DNA sequences through a sliding mechanism. The structured zinc fingers are shown as cylinders. Image from Dyson et al. [117]. **(D)** Illustration of the IDP's structural heterogeneity enabling its binding to multiple binding partners. Image from Kim and Han [133].

1.6 Modifying the IDP Landscape

The behaviour of Intrinsically Disordered Proteins as a highly frustrated system with no single folded state is reflected in its energy landscape – a 'hilly-plateau' with multiple similarly preferred interchanging states that are separated by barriers [100]. The conformations of these proteins are thus expressed as an

'ensemble' that together makeup its overall behaviour. As discussed in Section 1.4, this dynamic ensemble is a result of a finely tuned balance of the protein's interactions with itself and the surrounding solvent. Hence, the underlying folding free energy and thereby the conformational ensemble of disordered proteins are susceptible to be modified in two physiologically significant ways – **(1)** binding to partner proteins and **(2)** post-translational modifications [134, 135].

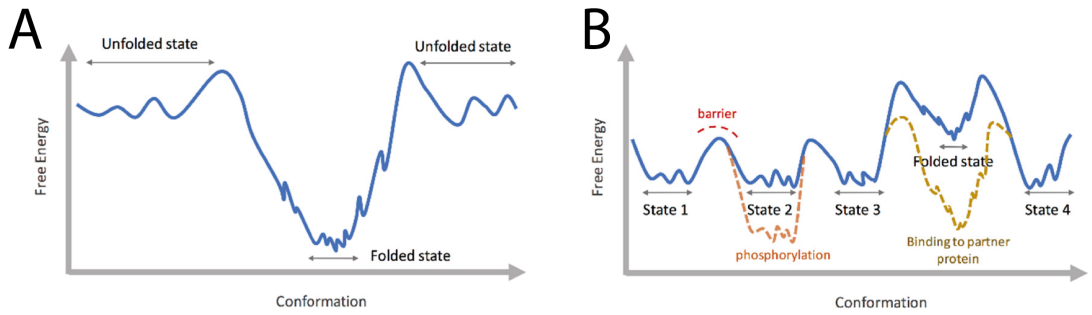


Figure 1.11 Illustration of the folding free energy landscape of **(A)** globular proteins and **(B)** Intrinsically-Disordered Proteins. The folding free energies of the disordered proteins can be modified by post-translational modifications like phosphorylation (orange) and the binding to partner proteins (yellow). Image from Bhattacharya and Lin [134].

1.6.1 Binding of IDPs

IDPs are a fascinating class of proteins that rarely function alone. Physiologically, disordered segments are often found as a part of larger multi-domain proteins that contain other structured/disordered regions [117] or found bound to partner biomolecules (proteins/DNA/RNA) [135]. In fact, within the human proteome, over 100,000 linear binding motifs have been identified to be located within intrinsically disordered regions [136]. This association of disordered interaction motifs with their partners has been predicted to occur through two broadly different structural mechanisms – Folding-upon Binding and Fuzzy Binding [137].

1.6.1.1 Folding-upon Binding

Folding-upon binding refers to the interaction process where disordered proteins that are unstructured in solution often fold into a well-defined three-dimensional structure upon interaction with a binding partner [138, 139]. Remarkably, this disorder-to-order transitioning has been observed to occur regardless of the nature of the biomolecular counterpart including another IDP

[140], globular protein [141], DNA [142] or RNA [143]. For example, the disordered H3 histone tail [15] condenses to form a segment of a three-stranded β-sheet when interacting with the Heterochromatin Protein 1 (HP1) that can then be crystallographically resolved [144].

The structural transition however raises a hitherto unanswered question over the mechanism of the process – does folding occur before binding or vice-versa [145]. The 'induced folding' hypothesis suggests that the IDP associates with its partner molecule in a fully disordered state and that the newly-formed interactions subsequently 'induce' structural transition. Contrarily, the 'conformational selection' hypothesis suggests that the conformational ensemble of the IDP in the isolated state includes the folded conformation that is then 'selected' by the target biomolecule [139, 146]. Figure 1.12A illustrates the two competing hypotheses of IDP folding-upon binding.

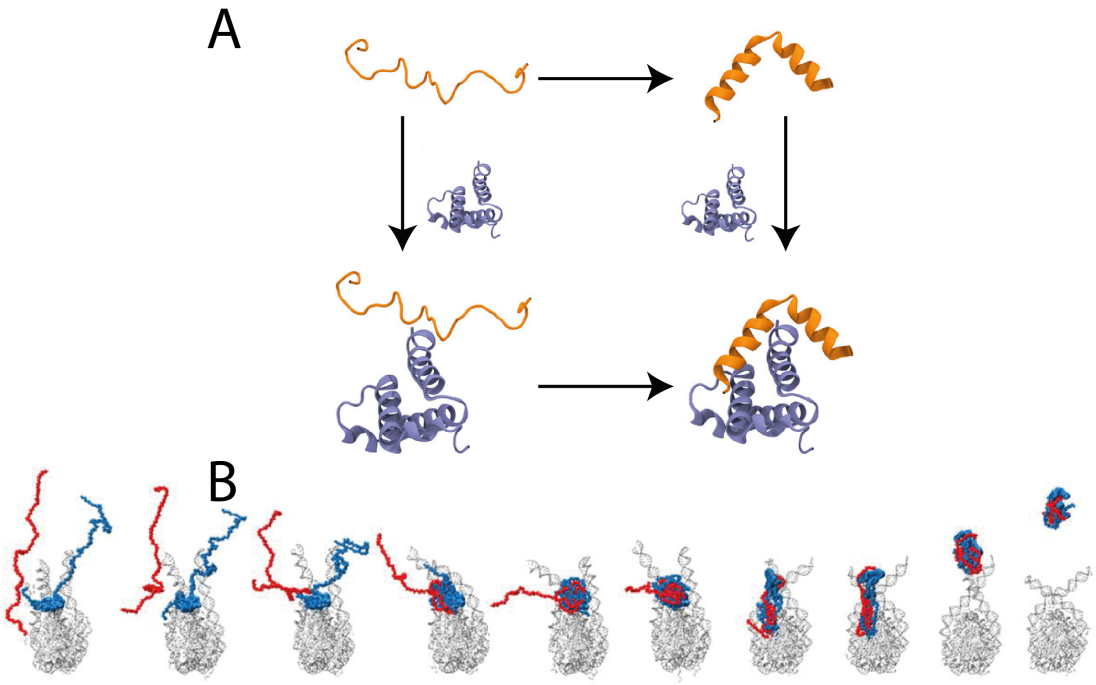


Figure 1.12 Illustration of the binding mechanisms of IDPs **(A)** Illustrations of the induced-folding (Bottom) and conformation selection (Top) mechanisms of IDP's (orange) folding-upon binding to its partner protein (purple). Image from Baker et al. [95]. **(B)** Illustration of the physiological role of fuzzy interactions where the fluctuating interactions of the disordered linker histone (blue) with both ProTα (red) and the nucleosome (grey) enables its chaperoning. Image from Heidarsson et al. [126].

1.6.1.2 Fuzzy Binding

The folding-upon binding mechanism allows the functioning of IDPs to be viewed within the traditional structure-function paradigm of biology. However,

not all disordered proteins undergo folding transitions upon interactions with their partner proteins [147–149]. Some IDPs form 'fuzzy' complexes where significant disorder remains within the binding interface [100]. Figure 1.12B illustrates one such interaction where the chaperone protein ProTα forms fuzzy interactions with the disordered domains of the Linker histone (LH) and the fluctuating nature of the binding enables chaperoning by permitting the LH's simultaneous interactions with both ProTα and DNA [126].

While fuzzy interactions were identified and studied over the past decade, they were overwhelmingly considered weak binding where the entropic penalty of restraining disordered regions is inadequately compensated by the enthalpy of fluctuating interactions [150–153]. However, within the past year breakthrough experiments and simulations have identified such disordered complexes to attain picomolar affinity levels [125, 126] and thus suggesting the presence of hitherto unidentified physiological roles for fuzzy interactions [100].

1.6.2 Post-Translational Modifications

As discussed in Section 1.3.2, post-translational modifications (PTM) are chemical changes to amino acids performed by highly specific enzymes. However, the unstructured extended nature of IDPs make them particularly accessible to modifying enzymes [135, 138] and thus susceptible to post-translational modification [154]. Figure 1.7 illustrates this where an overwhelming majority of the identified PTMs are situated at the disordered terminal tails of histones that extend out of the nucleosomal surface.

The impact of PTMs is two-fold – firstly, the modified proteins can be considered as possessing a 'new' amino acid composition [155] and thus a different folding free-energy profile in-comparison to the unmodified protein (Figure 1.11B). For example, the unstructured H4 histone tail [15] displays a significant propensity to form 3_{10}-helices upon K16 acetylation due to the reduced inter-sidechain repulsion within the protein [156]. Further, PTMs can have a second-hand effect on IDP structure where the modification enables the protein's binding to a counterpart [157, 155] which in-turn enables a structural transition [139, 158]. For example, the unstructured N-terminal domain of linker histone H1.4, upon K26 methylation can bind HP1 to form a β-strand [159].

1.7 Thesis Summary

The contents of this chapter provides an Introduction to two topics that together form the overarching focus of this thesis – chromatin structure and the role of disordered proteins within it. Disordered proteins are however the rule rather than the exception within the context of genomic proteins and in this thesis I hope to shed light on the functioning of a small subsegment of these proteins that form the core of chromatin structure – histones, linker histones and the heterochromatin protein. A more in-depth introduction to the specific proteins studied, their structures, roles and functional mechanisms are provided within each Chapter.

The remaining parts of this thesis are organized as follows

- Chapter 2 introduces the computational methods employed within this thesis.
- Chapter 3 explores the binding and functioning of the Linker Histone H1 and its disordered domains within the nucleosome.
- Chapter 4 explores the role of the disordered H1 N-terminal domain in differentiating the functioning of the LH subtypes.
- Chapter 5 explores the synergetic functioning of disordered and ordered domains in modulating the differential histone affinity of HP1 subtypes.
- Chapter 6 highlights and summarizes the key findings of the dissertation and proposes future research directions.

Chapter 2

Theory and Methods

Contents

2.1 Molecular Modelling

The behaviour of any molecule can be described by the interactions of its atoms, including both nuclei and electrons. The time-independent non-relativistic Schrödinger equation provides this description of nuclear r and electronic r_e coordinates as

$$\hat{H}\psi\left(r, r_e\right) = E\psi\left(r, r_e\right) \tag{2.1}$$

where $\hat{H}$ is the molecular Hamiltonian, E is the total energy and $\psi\left(r, r_e\right)$ is the molecular wavefunction in terms the nuclear and electronic coordinates. The Hamiltonian operator $\hat{H}$ can be expressed in terms of its constituent kinetic and potential energy operators as

$$\begin{aligned} \hat{H} &= \hat{T} + \hat{V} \\ &= \hat{T}_N + \hat{T}_n + \hat{V}_{NN} + \hat{V}_{nn} + \hat{V}_{Nn} \end{aligned} \tag{2.2}$$

where $\hat{T}_N$, $\hat{T}_n$ are the nuclear and electronic kinetic energy operators respectively, and $\hat{V}_{NN}$, $\hat{V}_{nn}$ and $\hat{V}_{Nn}$ are the three pair-wise interaction potential energy operators. Thus, the $\hat{H}$ Hamiltonian operator has the explicit form

$$\begin{aligned} \hat{H} = &-\sum_{I}^{N} \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_{i}^{n} \frac{\hbar^2}{2m_e}\nabla_i^2 + \frac{e^2}{4\pi\varepsilon_0}\sum_{i}^{n}\sum_{j>i}^{n} \frac{1}{|r_{ei} - e_{ej}|} \\ &+ \frac{e^2}{4\pi\varepsilon_0}\sum_{I}^{N}\sum_{J>I}^{N} \frac{Z_I Z_J}{|r_I - r_J|} - \frac{e^2}{4\pi\varepsilon_0}\sum_{i}^{n}\sum_{J}^{N} \frac{Z_I}{|r_I - r_{ej}|} \end{aligned} \tag{2.3}$$

where M_I and Z_I are the masses and charge of the nucleus, and m_e is the electron mass. An analytical solution of this Equation 2.3 is however intractable due to the coupling of the nuclear and electronic motions. But, with the mass of a proton being 1836 times greater than that of an electron ($M_I >> m_e$), the Born-Oppenheimer approximation rationalizes that the nuclei are fixed in-relation to the electronic motions. The molecular wave-function can thus be expressed as

$$\psi\left(r, r_e\right) = \phi\left(r_e; r\right)\Theta\left(r\right) \tag{2.4}$$

where $\Theta\left(r\right)$ is the nuclear wavefunction and $\phi\left(r_e; r\right)$ is the electronic wave-function that depends on the nuclear positions r. Combining this simplified

wavefunction with Equation 2.1 and expanding $\hat{T}_N$ provides

$$\begin{aligned}\hat{H}\phi\left(r_e;r\right)\Theta\left(r\right) = &-\sum_{I}^{N}\frac{\hbar^2}{2M_I}\phi\left(r_e;r\right)\nabla^2\Theta\left(r\right)\\ &-\sum_{I}^{N}\frac{\hbar^2}{2M_I}\Theta\left(r\right)\nabla^2\phi\left(r_e;r\right) - 2\sum_{I}^{N}\frac{\hbar^2}{2M_I}\nabla\Theta\left(r\right)\nabla\phi\left(r_e;r\right)\\ &+\left[\hat{T}_n + \hat{V}_{NN} + \hat{V}_{nn} + \hat{V}_{Nn}\right]\phi\left(r_e;r\right)\Theta\left(r\right)\end{aligned} \tag{2.5}$$

Disregarding the partial derivatives of the electronic wavefunction with regards to the nuclear coordinates, the equation simplifies as

$$\hat{H}\phi\left(r_e;r\right)\Theta\left(r\right) = E\phi\left(r_e;r\right)\Theta\left(r\right) \tag{2.6}$$

Combining Equations 2.5 and 2.6 provides

$$\frac{\left[\hat{T}_n + \hat{V}_{nn} + \hat{V}_{Nn}\right]\phi\left(r_e;r\right)}{\phi\left(r_e;r\right)} = E - \frac{\left[\hat{T}_N + \hat{V}_{NN}\right]\Theta\left(r\right)}{\Theta\left(r\right)} \tag{2.7}$$

where the left-hand side reduces to the energy of the electronic interactions E_e, and thus reducing the equation to

$$\left[E_e\left(r\right) + \hat{V}_{NN}\right]\Theta\left(r\right) = E\Theta\left(r\right) \tag{2.8}$$

if the kinetic energy contributions are excluded. Equation 2.8 thus provides that a potential energy function U that together approximates the nuclear and the ground-state electronic interactions $\left(U = \hat{V}_{NN} + E_e\right)$ can be used to describe the state (r) of the nuclei of the system. The Boltzmann distribution provides a relationship between this potential energy of the system and its probability of existing at a state with coordinates r_i as

$$P_i = \frac{\exp\left(-U\left(r_i\right)/k_BT\right)}{\sum_j^N \exp\left(-U\left(r_j\right)/k_BT\right)} \tag{2.9}$$

where k_B is the Boltzmann constant and T is the temperature. The complete potential energy surface of a complex biomolecule however is a function of many internal coordinates and constitutes a multidimensional surface that is incalculable. Atomistic molecular modelling methods use potential energy

functions (U) known as force fields to calculate this sum of states

$$Z = \sum_{j}^{N} \exp\left(-U\left(r_j\right)/k_B T\right) \tag{2.10}$$

called the partition function (Z) *in silico* and thereby allow the determination of the state preferences of the system [160].

Molecular Dynamics (MD) and Monte-Carlo (MC) simulations are the two distinct molecular modelling techniques developed for sampling this multidimensional phase space. In Monte Carlo simulations, a system is evolved in a stochastic manner and an ensemble average provides the configurational preferences. In contrast, Molecular Dynamics simulations allow the system to evolve in a deterministic manner and a time-average provides the configurational ensemble. While work in this thesis primarily utilizes Molecular Dynamics simulations (Section 2.2), the enhanced sampling techniques employed (Section 2.4) use elements of the MC method within them.

2.2 Molecular Dynamics Simulation

Atomistic Molecular Dynamics simulations begin with an initial structure of the atoms (position (r), velocity (v)) and force field parameters. The forces on the atoms in the system are then computed as a function of their positions using the force field potentials. The calculated forces are then used to compute the changes in atomic displacement over a small time interval. These steps are then repeated iteratively to define the motion of atoms and thereby the behaviour of the system (Figure 2.1).

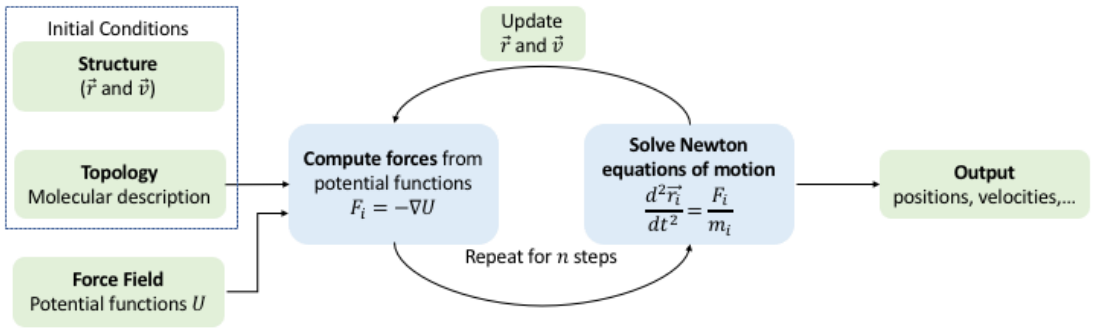


Figure 2.1 Schematic illustration of the MD simulation steps.

2.2.1 Initial Atomic Velocities

The initiation of Molecular Dynamics simulations requires the position (r) and velocity (v) to be defined. The initial velocity of atoms in the system are generated randomly using the Maxwell-Boltzmann distribution. This function calculates the probability of an atom possessing a velocity (v) at temperature T and is described as

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} \exp\left(\frac{m_i {v_i}^2}{2kT}\right) \tag{2.11}$$

where v_i is the velocity of the particle, k is the Boltzmann constant, m_i is the mass of the particle, T is the temperature and p is the probability.

2.2.2 Force Fields

The 'force fields' used in the simulations refer to the functional form of the potential energy functions that describe inter-particle interactions. [161]. This potential energy function of atomic positions $U\,(r)$ can be broken into the nature of inter-particle interactions – bonded and nonbonded.

$$U\,(r) = U_{bonded} + U_{non\text{-}bonded} \tag{2.12}$$

2.2.2.1 Bonded Interactions

The bonded interactions describe forces that arise from the covalent linkages of particles and thus include terms describing bond stretching (U_{bonds}), angle bending (U_{angles}) and the torsional rotation around bonds ($U_{dihedrals}$).

$$U_{bonded} = U_{bonds} + U_{angles} + U_{dihedrals} \tag{2.13}$$

The potential energies for bond stretching and angle bending are generally described using a potential derived from Hooke's law where the energy varies with the square of the displacement from the reference bond length (b_0) or angle (θ_0) as

$$U_{bonds} = \sum_{bonds} \frac{k_b}{2}(b - b_0)^2 \tag{2.14}$$

$$U_{angles} = \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.15}$$

The torsional rotation around bonds are calculated as the angle between two planes defined by the coordinates of four particles *i j k l*, where one plane is defined by atoms *i j k* and the other by *j k l*. The torsional energy functions are thus periodic in nature and are typically described as

$$U_{dihedrals} = \sum_{dihedrals} k_\phi \left(1 + \cos\left(n\phi - \phi_0\right)\right) \tag{2.16}$$

where ϕ_0 is the reference torsional value and n is the periodicity. In addition to this periodic dihedral function, *improper* dihedrals are often introduced to maintain the planarity of molecules like benzene or to keep the stereochemistry of molecule constant during the simulation. These potentials are generally introduced using a harmonic function similar to those describing bond or angle terms.

A number of different force fields like the OPLS-AA [162], GROMOS [163], AMBER [164] and CHARMM [165] have differently parameterized these force constants and equilibrium values by fitting to experimental measurements like electron diffraction and spectroscopy or to theoretical quantum mechanic calculations. Work in this thesis primarily used the Amber99SB-ildn [166] and Amber03ws [167] iterations of the AMBER family of force fields and the Charmm36M [168] iteration of the CHARMM family of force fields for their capabilities to better describe the conformational ensembles of IDPs [169]. To improve the quality of the fit to QM calculations, the CHARMM family of force fields slightly modify the form of the angle bending potential (Equation 2.15) using the Uray-Bradley correction [170]. Here, the angle bending potential of three atoms *i-j-k* is represented by a harmonic potential on the angle θ_{ijk} and an additional harmonic correction on the distance between *i-k* atoms:

$$U\left(\theta_{ijk}\right) = \frac{1}{2}k^{\theta}_{ijk}\left(\theta_{ijk} - \theta^{0}_{ijk}\right)^2 + \frac{1}{2}k^{UB}_{ik}\left(r_{ik} - r^{0}_{ik}\right)^2 \tag{2.17}$$

2.2.2.2 Non-bonded Interactions

Non-bonded interactions describe the interatomic potentials between particles and are represented using a combination of van der Waals (U_{vdW}) and electrostatic (U_{elec}) interactions. The van der Waals interactions are modelled using the Lennard-Jones potential as a function of the inter-atomic separation r_{ij}

$$U_{vdW} = \sum_{i,j}^{n} 4\epsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{2.18}$$

where the r^{-12} and r^{-6} exponential terms describe the repulsion and attractive forces respectively. The σ_{ij} and ϵ_{ij} values describe the collision radius and well-depth respectively, and are computed for the two interacting atoms using the Lorentz-Berthelot combination rules:

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \tag{2.19}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.20}$$

Electrostatic interactions between pairs of particles are modelled using the coulombic potential

$$U_{elec} = \sum_{i,j}^{n} \left(\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \tag{2.21}$$

where ϵ_0 is the permittivity and q_i/q_j are the the atomic partial charges. Similar to the bonded interactions, the multitude of different force fields adopt varying methodologies for the optimization of the non-bonded interaction parameters including matching experimental data like liquid densities or heats of vaporization [171].

2.2.3 Solvation Models

In physiological conditions, most biomolecules are surrounded by water and an accurate description of this environment is thus crucial for physically meaningful simulation results. Atomistic simulations generally describe this aqueous environment in an explicit manner using simple fixed charge multi-site water models. Figure 2.2 schematises the geometry of the three, four and five-site water models that have been developed to reproduce several key factors of bulk water including density, heat of vaporization and the O_w-O_w radial distribution function.

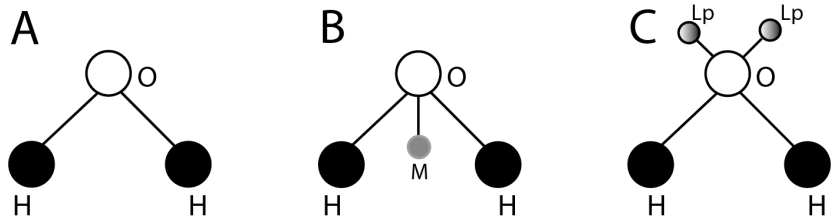


Figure 2.2 Schematic illustration of the geometry of the three, four and five site water models. 'O' is the oxygen atom, 'H' is the hydrogen atom, 'M' is a massless particle used in the four-site models and 'LP' is the Oxygen atom lonepair in five-site models.

The TIP3P [172] and SPC [173] water models follow the three-site geometry and are parameterized to have a single vdW interaction site on the oxygen atom and three electrostatic interaction sites (Table 2.1). The four-site models like TIP4P [174] and OPC [175] are parameterized to allow better dipole interactions and thus place an additional negative charge (Table 2.1) on a massless site ('M') located on the bisector of the HOH bond angle. In five-site models like TIP5P [176], this additional negative charge is instead placed on two dummy particles that represent the oxygen atom lone-pairs ('LP'). However, with the predominance of these solvent interactions in determining biomolecular interactions, the water models must be chosen to be consistent with the parameterization of the force fields. Work in this thesis used the TIP3P/SPC models with the Amber99SB-ildn force field, TIP4P-2005 [177] model with the Amber03ws force field and the modified TIP3P model [168] with the Charmm36M force field.

Table 2.1 A comparison of the parameters of the four water models used in this thesis. 'O' is the oxygen atom, 'H' is the hydrogen atom and 'M' is the massless particle. mTIP3P refers to the Charmm36M [168] modified TIP3P parameters.

	TIP3P [172]	mTIP3P [168]	SPC [173]	TIP4P-2005 [167]
r_{OH} (Å)	0.9572	0.9572	1.0	0.9572
r_{OM} (Å)	-	-	-	0.1594
$\angle_{H\text{-}O\text{-}H}$ (°)	104.52	104.52	109.47	104.52
q_O	-0.834	-0.834	-0.82	0.0
q_H	0.417	0.417	0.41	0.5564
q_M	-	-	-	-1.1128
ϵ_O (kJ/mol)	0.6364	0.6364	0.6506	0.7749
σ_O (Å)	3.1506	3.1506	3.1666	3.1589
ϵ_H (kJ/mol)	0.0	0.1925	0.0	0.0
σ_H (Å)	0.0	0.4000	0.0	0.0

2.2.4 Equations of Motion

The computation of the time evolution of the positions and velocities for a system of particles first requires the calculation of forces from the force field potential energy functions. This force on a particle ($\overrightarrow{F_i}$) is calculated as the negative gradient of the net potential energy of the particle.

$$\overrightarrow{F} = -\nabla U \tag{2.22}$$

Thus, for example, the force on a particle i along the x-direction due to its Lennard-Jones interaction with particle j at a separation of r_{ij} is calculated as

$$\overrightarrow{F_{x\cdot i}^{LJ}} = -\frac{\partial U_{vdW}}{\partial r_{ij}} \cdot \frac{\partial r_{ij}}{\partial x_i} \tag{2.23}$$

$$= \frac{24\epsilon}{r_{ij}} \left[2\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right] \cdot \frac{x_i - x_j}{r_{ij}} \tag{2.24}$$

$$= (x_i - x_j) \cdot \frac{24\epsilon}{r_{ij}^2} \cdot \left[2\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right] \tag{2.25}$$

Subsequently, the motion of particles can be obtained from these forces through the integration of Newton's equation of motion.

$$\overrightarrow{F_i} = m_i \frac{d^2 r}{dt^2} \tag{2.26}$$

However, with the velocity (v) of particles changing between time-steps, the direct update of atomic positions through an implementation of the Taylor expansion such as

$$r_i\left(t + \Delta t\right) = r_i\left(t\right) + \dot{r}_i \Delta t + \frac{1}{2}\ddot{r}_i \Delta t^2 + \frac{1}{6}\dddot{r}_i \Delta t^3 + \cdots \tag{2.27}$$

$$r_i\left(t + \Delta t\right) = r_i\left(t\right) + v_i\left(t\right)\Delta t + \frac{\overrightarrow{F_i}}{2m_i}\Delta t^2 \tag{2.28}$$

leads to a loss of time-reversibility and thereby a loss of the deterministic nature of MD simulations. Several *finite difference methods* [160] have thus been developed to allow the accurate integration of the propagation of atomic positions. The leap-frog algorithm [178] used in this thesis performs this integration by first calculating the velocity at time $t + \dfrac{\Delta t}{2}$

$$v_i\left(t + \frac{\Delta t}{2}\right) = v_i\left(t - \frac{\Delta t}{2}\right) + \frac{\overrightarrow{F_i}(t)}{2m_i}\Delta t \tag{2.29}$$

This half time-step velocity is used to calculate the displacement (r) of the particle.

$$r_i\left(t + \Delta t\right) = r_i(t) + v_i\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{2.30}$$

Finally, the velocity at time-step t is calculated as the average of two half time-step velocities.

$$v_i(t) = \frac{1}{2}\left(v_i\left(t - \frac{\Delta t}{2}\right) + v_i\left(t + \frac{\Delta t}{2}\right)\right) \tag{2.31}$$

2.2.5 Timestep and Constraints

The integration of the propagation of the atomic positions using discretised timesteps of Δt necessitates the force F_i to be nearly constant within this interval. The choice of the integration timestep is thus dependant on a range of factors including the force field, system composition and the integration algorithm. A timestep larger than the atomic motion frequencies would result in unstable simulations caused by a significant deviation from the equilibrium distances. In particular, bond stretching vibrations between hydrogen and heavy atoms occur on extremely fast timescales despite not being of particular interest to biomolecular MD simulations. However, using a small integration Δt timestep to match these vibrations is computationally inefficient and limits the sampling of phase space. Constraining algorithms are thus used to correct deviations of bond lengths from equilibrium values after the update of particle coordinates by the integrator and thereby enable the use of larger integration timesteps. In this thesis, a 2 fs timestep is used with the h-bonds within biomolecules and water restrained using the P-LINCS [179, 180] and SETTLE [181] algorithms respectively.

2.2.6 Periodic Boundary Conditions

Molecular Dynamics simulates a box of finite particles that encompasses the system of interest in a solvent. However, this leads to artefacts in the forces experienced by the particles at the edges of the box. Thus, Periodic Boundary Conditions (PBC) are applied where the system is treated as a lattice with infinite copies of itself extending in all directions. A particle in such a system that leaves a cell through one face into an adjacent box is placed back into the box from its periodic image at the opposite face of the box (Figure 2.3). The interactions of molecules across periodic boundaries enables the effect of modelling bulk solvent with a finite number of particles. Systems in this thesis were modelled using a combination of cubic, *truncated octahedron* and *rhombic dodecahedron* simulation boxes.

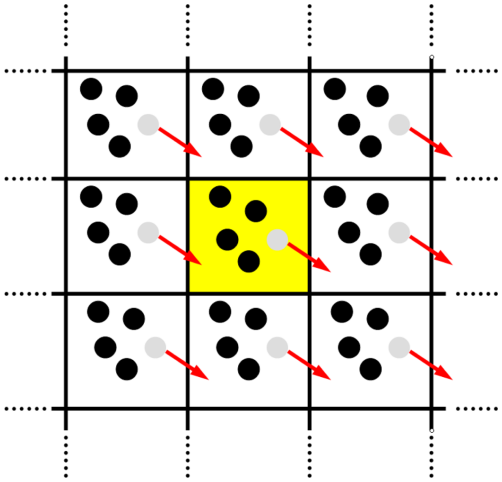


Figure 2.3 Two dimensional illustration of periodic boundary conditions. Particle trajectories are replicated in every direction with particles (grey) that exit from one face replaced from the opposite face.

2.2.7 Long Range Interactions

The calculation of non-boned inter-particle interactions encompasses the most computationally demanding aspect of the Molecular Dynamics routine [182]. In fact, within a system of N particles, the interaction of each particle with all others leads to the number of calculations scaling as N^2. A number of techniques have thus been developed to address this challenge of computational expedience [183, 184].

In the case of van der Waals interactions, the use of the Lennard Jones potential (Equation 2.18) results in the potential energy decaying with a r^{-6} dependence and thus rapidly approaches zero. Thus, van der Waals energies are only calculated for particles within a certain distance and assumed to be zero beyond this cut-off. Work in this thesis used a Lennard-Jones cut-off of 10 Å.

In contrast, the electrostatic interactions modelled using the coulombic potential (Equation 2.21) decay at a significantly slower rate of r^{-1}. Thus, the use of a direct cut-off can introduce significant errors in the calculation of inter-particle interactions [185]. Work in this thesis used the Particle Mesh Ewald (PME) [186, 187] variant of the Ewald technique to calculate long range electrostatic interactions. Briefly, the technique splits the r^{-1} sum into two terms as

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{2.32}$$

where the first term encompasses short-range interactions and the second encompasses long-range interactions. The electrostatic interactions are thus calculated in two parts where the first term is evaluated in real-space and the sum of the second term converges rapidly in Fourier space. The calculation of these long-range interactions in the Fourier space are further expediated by first mapping the charges onto a grid and subsequently using the Fast Fourier Transform (FFTW) algorithms [188] to perform discrete Fourier transforms. In this thesis, the breakup between short- and long-range interactions used a cut-off distance of 10 Å and a grid spacing of 1.4 Å was used for the Fourier transformations.

2.2.8 Coupling of Temperature and Pressure

The direct use of Molecular Dynamics gives rise to simulations in the 'NVE' ensemble – the modelling of a system with a constant number of particles, volume and energy. However, for a more accurate comparison to both in vitro and in vivo experiments that are conducted when exposed to atmospheric conditions, the 'NPT' ensemble with constant pressure and temperature is more appropriate for the molecular simulations. Thus, a number of algorithms termed thermostats and barostats have been developed to allow the control of temperature and pressure respectively within simulations.

2.2.8.1 Thermostat

The kinetic energy of a system with N particles can be expressed as

$$U_{Kinetic} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \tag{2.33}$$

where m_i and v_i are the mass and velocity of particle i. This kinetic energy can also be expressed as

$$U_{Kinetic} = \frac{3}{2} N k_B T \tag{2.34}$$

where k_B is the Boltzmann constant and T is the temperature. Combining the two equations, the temperature can be expressed as a function of particle velocities as

$$T = \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i v_i^2 \tag{2.35}$$

Thus, to control the system temperature at a temperature T_0, the particle velocities at each timestep can be scaled by a factor λ calculated as

$$\lambda = \sqrt{T_0/T} \tag{2.36}$$

where T is the instantaneous temperature. However, such a rescaling of velocities results in the kinetic energy of the system remaining unrealistically constant with zero fluctuations. The different thermostat algorithms thus differently calculate this rescaling λ parameter to both simultaneously maintain temperature while allowing kinetic energy fluctuations. Work in this thesis used a combination of the Berendsen [189] and Bussi [190] thermostats.

The *Berendsen* thermostat maintains temperature through a coupling of the system to an external bath of temperature T_0. The strength of this coupling is determined by a parameter τ as

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.37}$$

Thus, instead of a complete scaling of velocities, the thermostat uses the coupling parameter τ to tune fluctuations around the target temperature. At each timestep, the velocities are thus scaled by a factor λ calculated as

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T} - 1\right)} \tag{2.38}$$

where Δt is the integration timestep.

The *Bussi* thermostat is an extension of the Berendsen thermostat that utilizes an additional stochastic term to promote fluctuations of the kinetic energy without an oscillating behaviour around the target temperature. This coupling of the system temperature to that of an external bath is expressed as

$$dT = (T_0 - T)\frac{dt}{\tau} + 2\sqrt{\frac{TT_0}{N\tau}}dW \tag{2.39}$$

where dW is Wiener random noise whose values vary within a normal distribution of mean 0.

2.2.8.2 Barostat

The pressure in the system can be similarly maintained using the Berendsen barostat by coupling to a reference pressure P_0 using a time-constant τ_P.

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \tag{2.40}$$

Thus, at each timestep, the particle coordinates and box sizes are scaled by a factor μ

$$\mu = \sqrt[3]{1 - \frac{\beta \Delta t}{\tau_P} (P_0 - P)} \tag{2.41}$$

where β is the isothermal compressibility of the system. Simulations in this thesis used a reference pressure of 1 bar, a pressure coupling constant of 2.0 ps and a compressibility of 4.5 x 10^{-5}.

2.3 Energy Minimization

The initial atomic coordinates for Molecular Dynamics simulations are usually obtained through experimental methods like Cryo-EM [191], crystallography [192] or NMR [193], or through computational homology modelling [194]. Unfavourable molecule geometries are thus common and often lead to a 'blow-up' of the simulation system caused by the accumulation of large forces on particles. Energy minimization (EM) protocols are thus commonly employed before the initiation of MD simulations. The EM process attempts to optimize the geometry of the molecules by attempting to locate a local minimum on the Potential Energy surface of the system [160, 171]. Thus, given a potential energy function $U(r)$, the algorithm attempts to find the nearest set of coordinates where

$$\frac{\partial U}{\partial r} = 0 \tag{2.42}$$

$$\frac{\partial^2 U}{\partial r^2} > 0 \tag{2.43}$$

Work in this thesis used the *steepest descent* algorithm for performing the energy minimization. In this method, for each step n, the net forces on each atom (F_n) is first calculated as $-\nabla U$. Subsequently, for each particle, its new

positions for the next EM step $(n+1)$ is calculated as

$$r_{n+1} = r_n + \frac{F_n}{\max|F_n|} h_n \tag{2.44}$$

where $\max|F_n|$ is the scalar of the largest force on any atom within the system and h_n is the displacement distance. While such a first-derivative algorithm can rapidly guide a system towards a local minimum, the method can lead to oscillatory behaviour around the minimum. Thus, the displacement distance is scaled each step as

$$h_{n+1} = \begin{cases} 1.2 \cdot h_n & \text{if } V_{n+1} < V_n \\ 0.2 \cdot h_n & \text{if } V_{n+1} > V_n \end{cases} \tag{2.45}$$

Larger displacement steps are thus attempted when farther from the minima and iteratively scaled down if the minima is 'overshot'. In this work, an initial step-size of 0.1 Å was used and EM was performed until the maximum force on any atom $(\max|F_n|)$ was < 100 kJ mol^{-1} nm^{-1}.

2.4 Enhanced Sampling Molecular Dynamics

In classical Molecular Dynamics simulations as detailed in Section 2.2, the equilibrium value of an observable $\langle O \rangle$ is calculated as its time-averaged value [160].

$$\langle O \rangle = \frac{1}{N} \sum_{t=T_1,T_2\cdots T_N} O_i \tag{2.46}$$

However, this assumes the applicability of the ergodic hypothesis – a simulated trajectory of the system will sample all relevant phase space points in proportions that they makeup the partition function (Equation 2.10) . Despite the timescales attainable by MD simulations growing significantly over the past decade, with even the millisecond mark being breached recently [195], the hypothesis is not always applicable; especially with the presence of significant energy barriers within the system [95]. Figure 2.4 illustrates this issue of non-ergodicity where a system initiated from a state within the red-box samples only a small segment of the phase space.

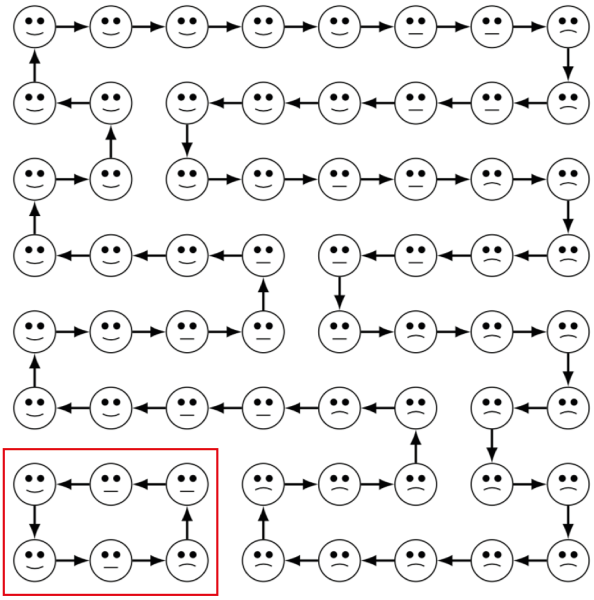


Figure 2.4 Illustration of the breakdown of the ergodic hypothesis where the states within the red-box form a disconnected region. Simulations initiated from states outside the box do not sample these states and the simulations initiated from within the box sample an underwhelming segment of the phase space. Image from Allen and Tildesley [196].

Enhanced sampling techniques have thus been developed to promote the sampling of 'rare' events like the crossing of energy barriers within MD simulations. These techniques primarily function by modifying the Hamiltonian to drive the system towards the configurations of interest while employing a 'reweighting' procedure to preserve the Boltzmann distribution of states. Work in this thesis used a combination of enhanced sampling methods including Temperature Replica-Exchange [197], Solute Tempering [198], Umbrella Sampling [199] and Metadynamics [200].

2.4.1 Temperature Replica Exchange Molecular Dynamics

Temperature Replica Exchange Molecular Dynamics (T-REMD) enhance sampling by performing multiple simulations of the system in parallel at different thermodynamic states [197]. Thus, to determine the ensemble of conformations at a particular temperature of interest T_0, a set of N simulations at temperatures of $T_0 < T_1 < T_2 \ldots T_N$ are performed in parallel (Figure 2.4).

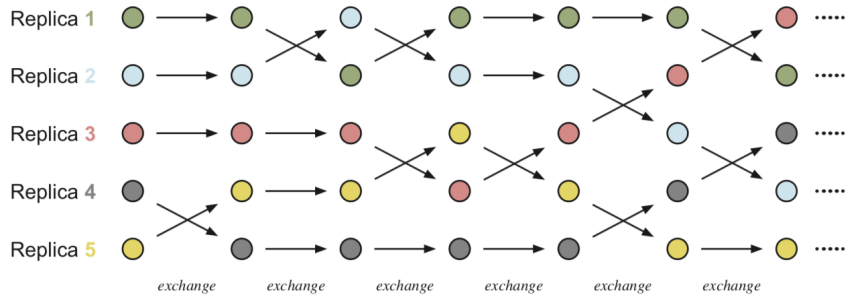


Figure 2.5 Schematic illustration of the temperature replica-exchange MD method.

At regular intervals, the system configurations between replicas are allowed to swap according to a Metropolis criterion that depends on the potential energy difference between the replicas. The exchange probability for this swap between replicas i and j is calculated as

$$P\left(i \leftrightarrow j\right) = \min\left\{1, \exp\left[\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)\left(U\left(r_i\right) - U\left(r_j\right)\right)\right]\right\} \tag{2.47}$$

where k_B is the Boltzmann constant. The algorithm thus provides the kinetic energy to overcome energetic barriers through the inclusion of higher temperature replicas while simultaneously using the Metropolis criteria for swaps to preserve the Boltzmann ensemble of conformations at the temperature T_0 [201]. However, the Metropolis acceptance probability implies that frequent exchanges of conformations occur only if the energetic overlap between replicas is sufficient. Thus, for computational expedience, T-REMD simulations in this work only attempted swaps between adjacent replicas (Figure 2.5) close-together in temperature,

2.4.2 Replica Exchange with Solute Tempering

In T-REMD simulations, a sufficient overlap of the replicas' energies are necessary for efficient Metropolis exchanges – a facet that necessitates the use of a large number of replicas and thereby makes the method intractable for large systems. Hamiltonian replica-exchange (H-REX) algorithms are a more general form of the method and enhance sampling by varying the underlying Hamiltonian between replicas. Exchanges between the replicas i and j are thus attempted at regular intervals with the probability

$$P\left(i \leftrightarrow j\right) = \min\left\{1, \exp\left[\left(\frac{1}{k_B T_i}\right)\left(\left(U_i\left(r_j\right) - U_i\left(r_i\right)\right) + \left(\left(U_j\left(r_i\right) - U_j\left(r_j\right)\right)\right)\right)\right]\right\} \tag{2.48}$$

where U_i and U_j are the modified Hamiltonians of the two replicas. Replica Exchange with Solute Tempering (REST2) [202, 203] is a H-REX method designed to promote fluctuations of protein conformations within aqueous solvents. N replicas are simulated in parallel at the relevant temperature T_0 with a λ_i parameter used to scale the Hamiltonian of the solute's (protein) interactions within each replica i as

$$q_i^{pro} = \sqrt{\lambda_i} q_i^{pro} \tag{2.49}$$

$$\epsilon_i^{pro} = \lambda_i \cdot \epsilon_i^{pro} \tag{2.50}$$

$$U_{dihedrals}^i = \lambda_i \cdot U_{dihedrals}^i \tag{2.51}$$

Thus, if the potential energy of the system U_{tot} is divided into its constituent interactions – solute-solute $\left(U_{pp}\right)$, solute-solvent $\left(U_{pw}\right)$ and solvent-solvent $\left(U_{ww}\right)$, they are scaled by factors of λ_i, $\sqrt{\lambda_i}$ and 1 respectively. The energy differences between the protein conformational changes are thus scaled while solvent-solvent interactions are unaffected. The λ_i parameters of each replica is calculated as

$$\lambda_i = \frac{T_0}{T_i} \tag{2.52}$$

where $\left(T_i\right)$ is a 'pseudo-temperature' $> T_0$ that replicates the effect of higher temperatures in T-REMD. However, the greater energy fluctuations from the modified Hamiltonians results in increased overlaps and thereby necessitates the use of significantly fewer parallel replicas [198].

2.4.3 Umbrella Sampling and WHAM

The Umbrella Sampling algorithm enforces sampling along a Collective Variable (CV, ξ) through the addition of an external energy term [204]. Thus, the effective biased potential energy surface $U_b\left(\xi\right)$ can be expressed as

$$U_b\left(\xi\right) = U\left(r\right) + U_{umb.}\left(\xi\right) \tag{2.53}$$

where $U\left(r\right)$ is the underlying force field potential and the external biased potential is an 'umbrella'-shaped harmonic term of the form

$$U_{umb.}\left(\xi\right) = \frac{K}{2}\left(\xi - \xi_0\right)^2 \tag{2.54}$$

where K is the strength of the harmonic bias. This external potential thus promotes the sampling of the phase space close to the chosen CV value ξ_0.

Typically, the umbrella-sampling algorithm is used to calculate the Free-energy $(F(\xi))$ for the simulation system to sample regions of the phase space along the CV. Thus, N non-interacting simulations are performed with the ξ_0 evenly distributed between the initial and final values of the CV of interest (Figure 2.6).

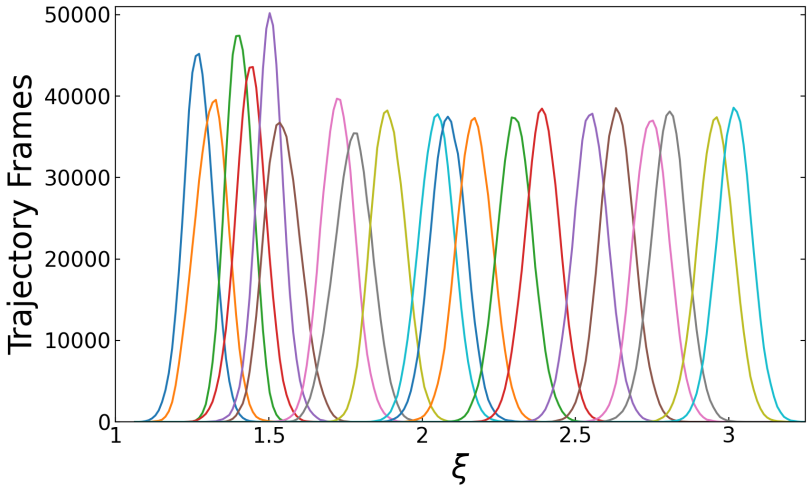


Figure 2.6 Illustration of the sampling distribution along the phase space through the use of multiple umbrella simulation windows at different ξ_0. The distribution of CV sampling from each of the N simulations are coloured differently.

The collection of N simulations with varying ξ_0 together provide $P_b(\xi)$ – the sampled probability distribution effected by the biased potential energy surface U_b. This sampled distribution can be related to the applied biased potential as

$$P_b(\xi) = \exp(-\beta U_b(\xi)) = \exp(-\beta(U(r) + U_{umb.}(\xi))) = P(\xi)\, e^{-\beta U_{umb.}(\xi)} \quad (2.55)$$

where $P(\xi)$ is the distribution obtained from an unbiased simulation and related to the Free-Energy as

$$F(\xi) = \frac{-1}{\beta} \ln P(\xi) \quad (2.56)$$

The Weighted Histogram Analysis Method (WHAM) [205] is used to calculate this $P(\xi)$ by combining the results of the multiple umbrella simulations [206] with varying ξ_0 as

$$P(\xi) = \sum_{i}^{N} p^i P^i(\xi) \quad (2.57)$$

where $P^i(\xi)$ is the unbiased distribution calculated from each window. p^i is the weight assigned to each umbrella window and is calculated self-consistently

using $\sum_i p^i = 1$ and

$$p^i = \frac{K}{\sum_i^N \int \exp\left(-\beta U^i_{umb.}\left(\xi\right)\right) d\xi} \tag{2.58}$$

Work in this thesis used *g_wham* [207] – the implementation of WHAM within the Gromacs MD simulation package [182] to analyze and calculate the free energy from Umbrella Sampling simulations.

2.4.4 Metadynamics

Metadynamics belongs to a class of algorithms similar to Umbrella Sampling that aims to promote sampling along a CV ξ through the application of an external biasing potential [200].

$$U_b\left(\xi\right) = U\left(r\right) + U_{metad}\left(\xi\right) \tag{2.59}$$

The external biasing potential $U_{metad}\left(\xi\right)$ is constructed as a sum of Gaussians along the CV space [208, 209] as

$$U_{metad}\left(\xi, t\right) = W \sum_{t'=\tau_G, 2\tau_G, \ldots} \exp\left(-\frac{\left(\xi - \xi_{t'}\right)^2}{2\delta^2}\right) \tag{2.60}$$

where τ_G is the time interval for gaussian deposition with height W and width δ. The gaussians are centred on the instantaneous CV value $\xi_{t'}$ and thus discourages the system from revisiting the sampled regions along the phase space. Figure 2.7 illustrates this functioning of Metadynamics where the addition of history-dependant gaussians raises the potential energy surface and promotes the transitioning of the system across states that are separated by a large energetic barrier.

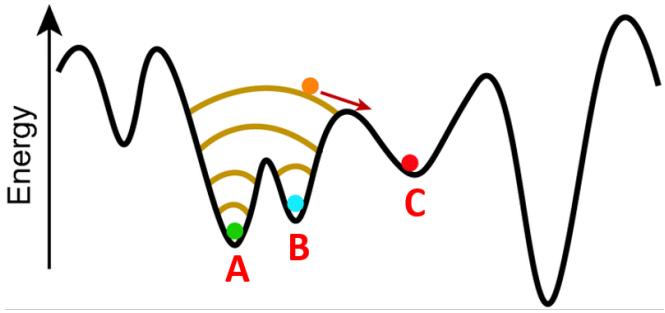


Figure 2.7 Illustration of the variation in the potential energy surface along the CV effected by addition of biasing gaussian potentials. The addition of the history dependant potential discourages the re-visitation of sampled minima to enable transitions across energetic barriers.

For metadynamics simulations along multiple collective variables, the function can be extended as

$$U_{metad} = W \sum_{t'=\tau_G,2\tau_G,\ldots} \prod_{i} \exp\left(-\frac{\left(\xi_i - \xi_{i_{t'}}\right)^2}{2\delta_i^2}\right) \tag{2.61}$$

where δ_i is the width of the gaussians along each CV ξ_i. The accuracy and effectiveness of the metadynamics algorithms are however sensitive to the choice of the gaussian parameters. The addition of larger gaussians (larger W, δ) at more frequent intervals (smaller τ_G) promotes a faster sampling of the phase space but can also cause errors in the reconstructed surface and push the system towards physically unfeasible regions. Conversely, smaller gaussians added infrequently promote accuracy but take longer to promote ergodicity, especially with deep energy wells. Simulations in this thesis thus utilized the Well-Tempered Metadynamics algorithm [200] where the external biasing potential U_{metad} is dependent on the histogram of states sampled by the system $N(\xi, t)$ instead of the instantaneous state at τ_G intervals. This external potential is constructed as

$$U_{metad}(\xi, t) = \Delta T \ln\left(1 + \frac{\omega N(\xi, t)}{\Delta T}\right) \tag{2.62}$$

where ω and ΔT have dimensions of energy rate and temperature respectively and $N(\xi, t)$ can be obtained from individual states δ_ξ sampled as

$$N(\xi, t) = \int_0^t \delta_\xi dt \tag{2.63}$$

The rate of change of the external potential $\dot{U}_{metad}(\xi, t)$ can be calculated as

$$\dot{U}_{metad}(\xi, t) = \frac{\omega \Delta T \delta_{\xi}}{\Delta T + \omega N(\xi, t)} = \omega \exp\left(\frac{U_{metad}(\xi, t)}{\Delta T}\right) \delta_{\xi} \tag{2.64}$$

A comparison of this expression of $\dot{U}_{metad}(\xi, t)$ with Equation 2.60 reveals that the well-tempered metadynamics essentially rescales the height of the gaussians from an initial value of ω in proportion to the sampling size $N(\xi, t)$. Thus, larger gaussians are added in the less-sampled regions of the phase space (small $N(\xi)$) and vice-versa for the well-sampled regions. The value of ΔT is the 'bias-factor' and determines the scaling of the gaussian heights with $N(\xi)$. Larger values of ΔT reduce this scaling dependency with the simulation reducing to standard metadynamics as $\Delta T \to \infty$ and vice-versa for smaller ΔT with the simulation reducing to unbiased MD as $\Delta T \to 0$.

2.4.5 Biased-Exchange Metadynamics

In metadynamics, the free-energy landscape of the system is reconstructed by gradually filling the local minima with gaussian hills (Figure 2.7). In most biomolecular systems however, this potential energy landscape is multidimensional with multiple CVs required to adequately describe the conformational transitions involved [210]. While the formalism of the deposited gaussians is extendable to multiple CVs (Equation 2.61), metadynamics simulations are typically unsuitable for the simultaneous use of greater than three CVs. Such high-dimensional metadynamics simulations are computationally inexpedient as the simulation time required to fill the potential energy landscape grows exponentially with the number of CVs [200].

The Bias-Exchange Metadynamics (BE-Metad) algorithm [211] circumvents this obstacle by simulating multiple replicas of the system in parallel with each being biased along different a CV ξ_i. At regular intervals, the configurations of two random replicas i and j are swapped with an acceptance probability of

$$\min\left\{1, \exp\left[\left(\frac{1}{k_B T}\right)\left(U^{j}_{metad}(r_j) + U^{i}_{metad}(r_i) - U^{j}_{metad}(r_i) - U^{i}_{metad}(r_j)\right)\right]\right\} \tag{2.65}$$

where U^{i}_{metad} is the history-dependant metadynamics gaussians deposited along CV ξ_i. The trajectory of each walker thus traverses the multidimensional energy landscape a single CV at a time with the diffusion along one CV space unencumbered by barriers along other CVs. In this thesis, an additional 'neutral' replica

with with no metadynamics biases ($U_{metad} = 0$) was included in the BE-Metad simulations.

The addition of the history dependant bias potentials however preclude a direct averaging of the system's characteristics as simulation time is without physical meaning. A number of algorithms have thus been developed to allow 'reweighting' of the trajectory frames and the subsequent calculation of the unbiased equilibrium ensembles. Work in this thesis primarily used the methodology of Bonomi et al. [212] to reconstruct the equilibrium ensemble. Given the inclusion of the multiple biasing potentials (U^i_{metad}), the biased probability distribution P^i_b along each replica can be calculated as

$$P^i_b(\xi_i) = \exp\left(-\beta\left(U(r) + U^i_{metad}(\xi_i)\right)\right) = P^i(\xi)\, e^{-\beta U^i_{metad}(\xi_i)} \tag{2.66}$$

where $P^i(\xi)$ is the unbiased distribution within each replica. The WHAM method (Equation 2.58) is then used to assign weights $w(t)$ for each frame within the multiple replicas and subsequently used to calculate the equilibrium average of any system characteristic O as

$$\langle O \rangle = \frac{\sum w_i(t)\, O_i}{\sum w_i(t)} \tag{2.67}$$

2.4.6 Parallel-Tempered Metadynamics

Metadynamics and its associated techniques are capable of significantly increasing sampling along the Collective Variables ξ. However, the selection of these CVs for use with metadynamics simulations is very often non-trivial [200, 213] with the energy barriers along unaccounted CVs affecting sampling efficiency [210, 214]. The Parallel-Tempered Metadynamics (PTMetaD) [215] algorithm circumvents this obstacle through the combination of the T-REMD method with metadynamics. To determine the conformational ensemble of a system at a temperature of interest T_0, a set of N replicas at temperatures $T_0 < T_1 < T_2 \ldots < T_N$ are simulated in parallel. Within each replica i, the metadynamics biases U^i_{metad} are evolved independently along a common set of CVs. At regular intervals, the system configurations between adjacent replicas i and j are swapped with an acceptance probability of

$$\min\left\{1, \exp\left[(\beta_i - \beta_j)(U_i - U_j) + \beta_i\left(U^i_{metad}(\xi_i) - U^i_{metad}(\xi_j)\right)\right.\right. \\ \left.\left. + \beta_j\left(U^j_{metad}(\xi_j) - U^j_{metad}(\xi_i)\right)\right.\right. \tag{2.68}$$

where β_i is the inverse thermodynamic temperature of replica i. The use of the high temperature replicas permits a traversal of the energetic barriers along slowly diffusing or unaccounted CVs.

However, similar to T-REMD the energy distributions of replicas i and j need to sufficiently overlap for exchanges to occur and in-turn necessitates the use of a large number of replicas. Parallel-Tempered Metadynamics in the Well-Tempered Metadynamics (PTMetaD-WTE) [216] was thus developed as an extension to the PTMetaD method to significantly reduce the number of necessary replicas. In this method, prior to the simulation with relevant CVs, a preliminary metadynamics simulation run is performed for each of the replicas using only their Potential Energy (U) as a CV. The bias from this preliminary simulation is retained in the PTMetaD simulations to significantly enhances the energetic fluctuations of each replica and thereby its overlap with adjacent replicas. Figure 2.8 illustrates the PTMetaD-WTE algorithm by comparing the overlap between eight replicas geometrically spaced between 298 - 400K with and without the WTE.

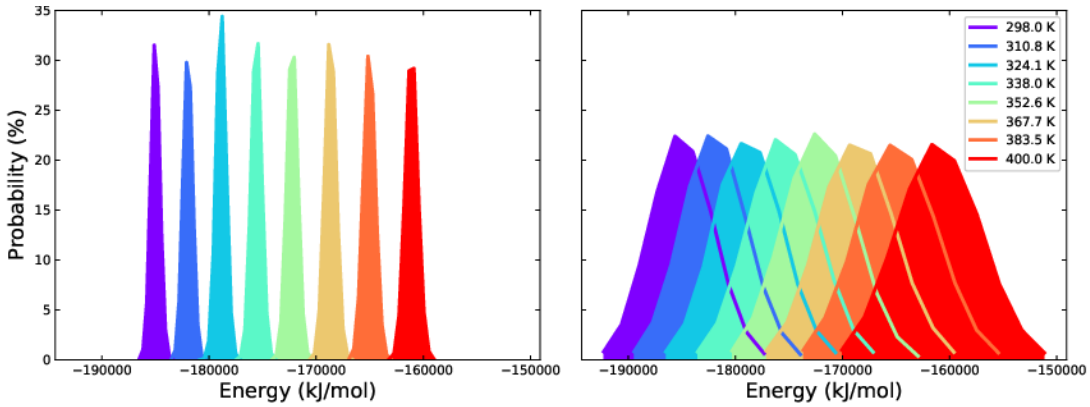


Figure 2.8 Illustration of the increase in the energetic overlaps between eight replicas without (**Left**) and with (**Right**) the enhanced potential energy fluctuations due to the well-tempered ensemble.

2.5 Docking

Molecular docking is a computational approach to rapidly approximate the preferred orientation of one molecule relative to a second when bound together to form a complex. While less accurate than Molecular Dynamics simulations and its associated enhanced sampling techniques, the docking approach is significantly more rapid with results obtainable within seconds on standard desktop computers [217]. Typical docking approaches are a two-step process where different probable conformations of the molecules are iteratively trialled

and a 'docking score' is used to identify the most favourable conformation [218]. With each atom having a total of six degrees of rotational and translational freedom, this results in a huge number of binding modes to be 'scored'. Thus, for computational expedience docking algorithms reduce the number of rotational bonds in the complex. This reduction is typically performed by considering the receptor (larger) structure to be completely rigid while the smaller 'docked' molecule is considered to be minimally flexible [219, 220].

The separation into a rigid 'receptor' and flexible 'docked' molecules has been extremely effective in leading to the development of multiple well-validated protocols for ligand docking including AutoDock [221], Rosetta [222], Gold [223] and MOE. In contrast, the docking of two larger biomolecules is significantly more challenging due to their increased conformational flexibility when interacting and the difficulty in post-docking clustering of results [224–226]. The protocols thus lead to a significant number of 'false-positive' predictions of the molecular interaction configurations [227, 228]. Docking in this thesis is thus performed using the High Ambiguity Driven protein-protein DOCKing (HADDOCK) protocol [229, 230] that allows the use of experimental information to drive the docking process and thereby significantly reduce false predictions.

Chemical shift perturbations in NMR or FRET measurements can identify residues that play a role in inter-molecular interactions and thus on the 'interaction-interface' of the two molecules. The HADDOCK protocol allows their inclusion within the docking protocol as 'active' residues. Contacts between these active residues of the two molecules are enforced through the use of Ambiguous Interaction Restraints (AIR). For each 'active' residue i in molecule A, the AIR ($d_{\mathrm{iAB}}^{\mathrm{eff}}$) is calculated as

$$d_{\mathrm{iAB}}^{\mathrm{eff}} = \min\left\{3.0, \left(\sum_{m_{iA}=1}^{N_{\mathrm{atoms}}}\sum_{k=1}^{N_{\mathrm{Bres}}}\sum_{n_{kB}=1}^{N_{\mathrm{atoms}}}\frac{1}{d_{m_{iA}n_{kB}}}\right)^{-1/6}\right\} \tag{2.69}$$

where $d_{m_{iA}n_{kB}}$ is the distance (in) between molecule A's m^{th} atom of active residue i and molecule B's n^{th} atom of active residue k. N_{atoms} represents the number of atoms in each of the active residues and N_{Bres} is the total number of active residues in molecule B. This usage of AIRs allows the docking protocol to iterate across all possible contact configurations between molecules A and B before scoring and selecting the best one. The scoring of configurations is carried out using the electrostatic and van-der-Waals interaction parameters from the OPLS-AA force field [162].

Chapter 3

Linker Histone H1 within nucleosome

Contents

This Chapter is based on the manuscript:

"*Emergence of chromatin hierarchical loops from protein disorder and nucleosome asymmetry.*" **Akshay Sridhar**, Stephen Farr, Guillem Portella, Tamar Schlick, Modesto Orozco, and Rosana Collepardo-Guevara. Proceedings of the National Academy of Sciences, 2020, 117, 7216-7224.

3.1 Introduction

Linker histones are 'non-core' histones that bind to the nucleosome at the DNA entry/exit site [9, 231] to interact primarily with linker DNA [232] (Figure 1.2). First identified as a component of the nucleosome by Simpson et al. [18] in 1978, the H1 protein has since been identified to play a critical role in regulating chromatin structure [233–235]. *In vitro*, the deletion of H1 has been correlated with dramatic chromatin decompaction and an increase in nuclear volume [236–238]. Similarly *in vivo*, the linker histone H1 is essential for the normal morphology of mitotic chromosomes and thus for the functioning of life itself [238–240]. This association of linker histones with chromatin compaction has lead to the protein being postulated as general repressors of genomic transcription [241]. In fact, even on the single nucleosomal scale, a comparison of the nucleosome crystallographic structures with/without H1 reveals that the H1 protein can compact linker DNA strands by close to ∼8 Å [9] (Figure 3.1).

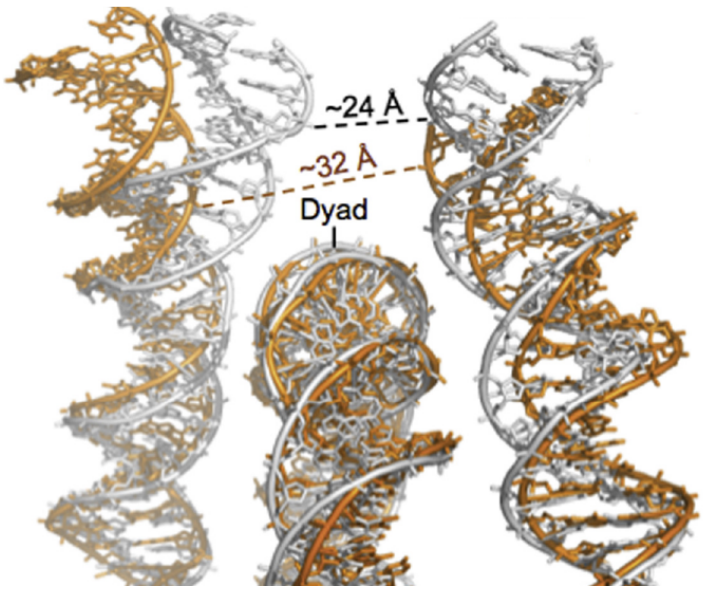


Figure 3.1 Illustration of the linker histone H1 capability to alter the trajectory of the entering/exiting linker DNA strands by screening electrostatic repulsion and thereby allowing compaction. The 601-nucleosomal structures with H1 (grey, 4QLC [9]) and without H1 (yellow, 1ZBB [31]) are compared by aligning base pairs (bp) around the dyad. The H1 protein is not visualized for clarity. Image from Zhou et al. [9].

However, more recent evidence has gradually unveiled a more nuanced role of H1 in genomic regulation [242] dependant on subtype, binding configuration and post-translational modifications, and thereby stirred intense research into its functioning [234, 243, 244].

3.1.1 H1 Structure

The linker histones are particularly rich in positively-charged Lysine residues and have a three-domain structure consisting of a ∼75 residue-long globular domain, an unstructured N-terminal domain (NTD) of up to 45 residues and a long unstructured C-terminal domain (CTD) of ∼100 residues [245, 246]. Figure 3.2 illustrates this tripartite structure of the protein with the unresolved disordered regions modelled onto the globular domain.

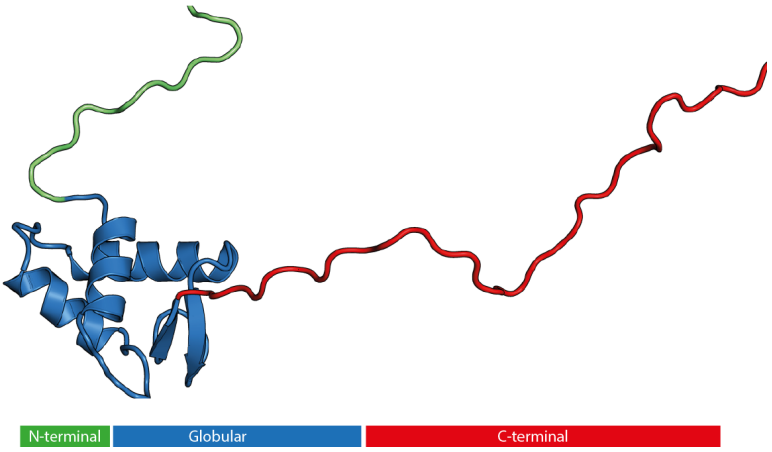


Figure 3.2 Three domain structure of linker histone H1 with the central globular domain (blue) from the crystallographic structure 5NL0 [231] and the two terminal domains (red, green) built using Modeller [194].

3.1.1.1 Globular Domain

The globular domain adopts a winged-helix configuration [17, 247] – a compact motif commonly observed in DNA binding proteins consisting of 3 α-helices, 2 linker 'wings' and 2 short β-strands [248]. While possessing a significantly lower positive-charge concentration than the two terminal domains [243], a range of studies have identified this globular domain to play the most significant role in the nucleosomal recognition of linker histones [249–251]. In fact, cryo-EM density maps of nucleosomes constituted with full-length linker histones and those without the terminal domains show significant density overlaps in the globular region [231, 232].

3.1.1.2 Carboxy-Terminal Domain

In contrast to the structured globular domain, the CTD is rich in basic residues while being completely devoid of hydrophobic residues. In fact, up to 50% of the CTD is composed of positively charged Lys/Arg residues crucial for mediating

interactions with the DNA backbone [252]. Although the winged-helix motif of the globular domain aids in nucleosomal recognition, the high positive charge content and long length of the CTD have led to the postulation of it being the primary determinant of sustained H1 binding to the nucleosome [124, 251, 253]. While this relative relevance of domains in mediating nucleosomal binding is an ongoing focus of research, a range of experiments have definitively identified the chromatin compaction capability of H1 to be dependent on the CTD [233]. In fact, chromatin segments constituted with CTD-less linker histones display compaction levels identical to segments without linker histones [232, 254].

3.1.1.3 Amino-Terminal Domain

The amino-terminal domain of H1 is the shortest linker histone segment and contains only 25% of the charge density and number of residues as the CTD [245]. Correspondingly, the NTD had been hypothesized to not contribute to the chromatin compaction capabilities of H1 [232, 254]. However, fluorescence recovery after photobleaching (FRAP) experiments later identified the NTD-less linker histones to display a considerably lower nucleosomal binding affinity [124, 255, 256]. Thus, linker histones without the NTD have been postulated to be capable of enabling maximal chromatin compaction but to require higher concentrations of H1 to achieve sufficient H1-nucleosome binding [257].

3.1.2 H1 Post-Translational Modifications

Similar to their core histone counterparts (Figure 1.7), linker histones are also highly enriched in post-translational modifications (PTM) that can modulate the protein's roles/capabilities to regulate chromatin structure [258, 259]. However, H1 interacts dynamically with chromatin [250, 260, 261] and the transiently unbound linker histones are thus significantly more accessible to modifying enzymes [253]. Thus, unlike within core histones, PTM sites within H1 are not exclusively found within the unstructured terminal domains but extend across all three domains [262–264]. The modifications H1 can exhibit include methylation [265, 266], phosphorylation [127] and ribosylation [267] within the NTD; acetylation [263], ubiquitylation [268] and citrullination [76] within the globular domain; and acetylation [263], methylation [269], ribosylation [270], formylation [73, 271] and phosphorylation [272] within the CTD.

Among the plethora of PTMs, the phosphorylation of Ser/Thr residues within the CTD is the best-studied modification whose physiological relevance

has been determined [272]. This phosphorylation is a conserved feature of all H1 subtypes and is preferentially performed on the **S/T**-P-X-K sited by Cyclin-Dependant Kinase (CDK) enzymes [273]. This modification has been identified as an important component of cell-division with phosphorylation levels gradually increasing as the cell cycle progresses, becoming maximal during the metaphase and dropping sharply thereafter during the telophase [274–276]. In fact, cells treated with phosphorylation inhibitors could be prevented from entry into the mitotic phase [277, 278]. *In vitro*, phosphorylation has been observed to have a cyclical effect with partial H1 phosphorylation disrupting DNA-binding and chromatin compaction while hyper-phosphorylation largely restoring the binding/compaction capabilities of H1 [279, 280]. While the structural mechanism of this phosphorylation-mediated modulation is yet to be determined, several studies have postulated this to occur through changes in the secondary-structure of the CTD IDR [279–281].

3.1.3 H1 Nucleosomal Binding

In addition to post-translational modifications, the functional variation in H1's genomic regulatory role is further diversified by its binding configuration within the nucleosome [282–284]. Two distinct models for this positioning of the globular domain have been proposed – the 'on-dyad' and 'off-dyad' configurations (Figure 3.3) [234, 285].

The 'on-dyad' model positions the globular domain on the nucleosomal dyad axis and in-contact with both linker DNA arms [232, 286]. This configuration is better validated with two crystallographic structures resolved recently – 4QLC [9] and 5NL0 [231]. Both structures utilized H1.0 orthologs and predicted a binding mode with three distinct interaction interfaces. Interactions with the nucleosomal DNA on the dyad are mediated primarily by the α2 helix within the globular domain's winged-helix motif. The two linker DNA arms interact with the globular domain through its α3 helix and L1 loop and are thus termed Linker-α3 and Linker L1 strands respectively (Figure 3.4).

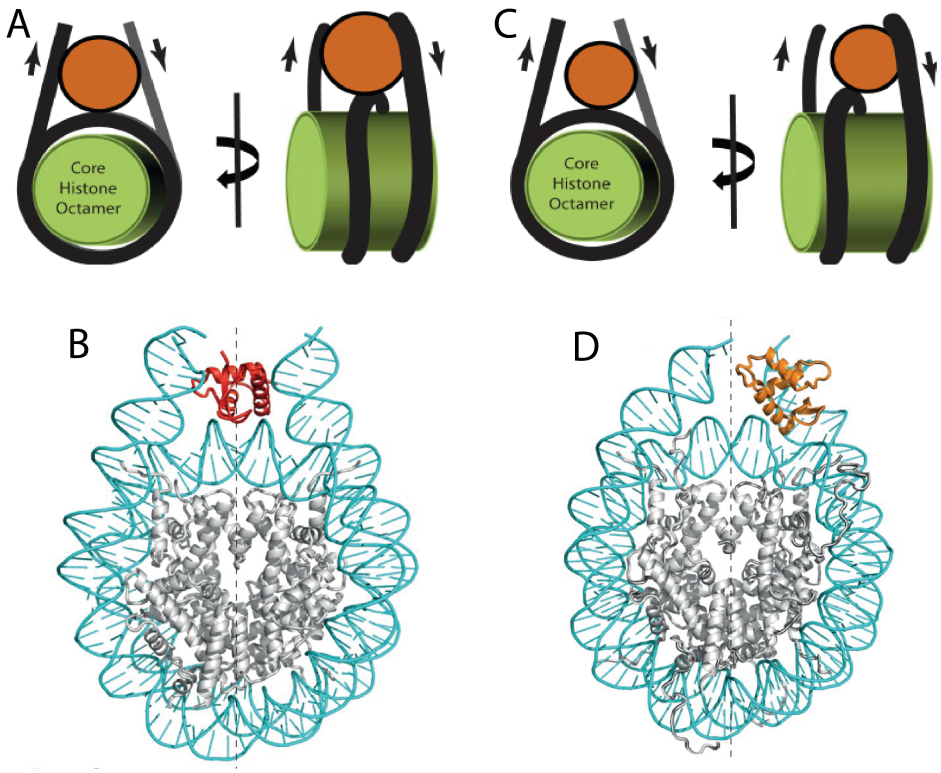


Figure 3.3 (A,B) Illustrations and structure of the on-dyad binding conformation. **(C,D)** Illustrations and structure of the off-dyad binding conformation. In the illustrations, the LH is shown in orange and the octamer core as a green cylinder. Within the structures, the LH is shown in red/orange and the dyad axis is illustrated as a dotted line. Images from Parseghian et al. [285] and Fyodorov et al. [234].

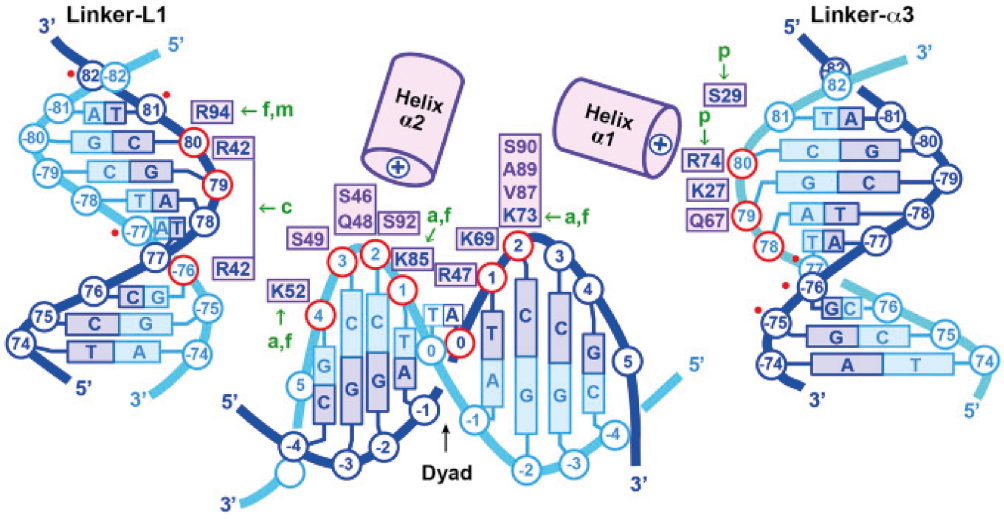


Figure 3.4 Illustration of the L1 linker, α3 linker and nucleosomal DNA contact interfaces of the globular domain in the on-dyad binding configuration. The residues interacting with DNA are annotated and the interacting base pairs are highlighted in red. The PTMs observed within the residues interacting with DNA are marked 'a' (acetylation), 'f' (formylation), 'm' (methylation), 'p' (phosphorylation) and 'c' (citrullination). Image from Bednar et al. [231].

In-contrast, the 'off-dyad' model positions the globular domain asymmetrically away from the dyad and in-contact with only a single linker DNA arm [287, 288]. This binding mode of linker histones has hitherto not been crystallographically validated and has thus even been postulated to be enthalpically less favourable [289]. This interaction model has thus far only been predicted using residue-restrained docking studies [287, 288, 290] or observed in low-resolution cryo-electron maps of nucleosome arrays [291]. Figure 3.5 illustrates the two distinct off-dyad configurations predicted from different experimental methodologies for different linker histone subtypes.

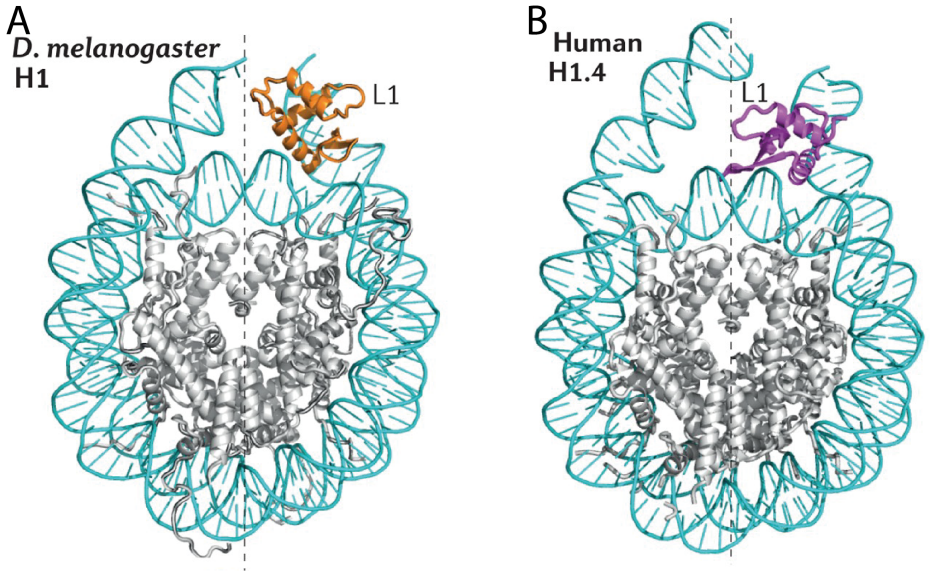


Figure 3.5 (A) The off-dyad conformation observed by Zhou et al. [288] using NMR restrained docking. **(B)** The off-dyad conformation observed by Song et al. [291] using cryo-EM microscopy. The orientations of the L1 loop in both conformations is added for clarity. The dyad axis is illustrated as a dotted line. Image from Fyodorov et al. [234].

3.2 Aims and Summary

The Linker Histone contributes significantly to the nucleosomal structure and in-turn the compaction of chromatin. However, significant open questions remain on its binding mode to the nucleosome; the role of the unstructured terminal domains and most importantly the mechanism of the protein's functioning. In this Chapter, I use a combination of enhanced sampling atomistic Molecular Dynamics simulations to analyze – (1) the orientations of the linker histone domains and (2) the impact of H1 binding on nucleosome structure. Further, the simulations analyze the role/impact of CTD phosphorylation and attempt to correlate the results to the PTM's biological role.

3.3 Methods

3.3.1 Model Building

3.3.1.1 211-bp Nucleosome

The atomic coordinates of the 211-bp nucleosome with two symmetric 32-bp DNA linker arms were extracted from one of the central nucleosomes of the tetranucleosome crystal structure (PDB ID: 1ZBB [31]). The eight histone proteins within this structure were then replaced by those from the high-resolution 1KX5 [16] nucleosome structure with resolved histone tails. The sequence of *h*H1.0 was obtained from Uniprot [292] (ID: P07305) and Modeller [194] was used to create a homology model of the linker histone's globular domain (GD) with *c*H5 (4QLC [9]) as the template. The unstructured terminal arms of *h*H1.0 were built in extended state using PyMOL [293], compacted using a short T-REMD simulation and subsequently attached to the homology modelled GD. Finally, this built structure was positioned onto the 211-bp nucleosome by overlaying its GD onto the on-dyad bound configuration observed for *c*H5 (4QLC). This condensed system was then solvated in a octahedral solvent box with an ionic concentration of 0.15 M Na/Cl for atomistic simulations.

3.3.1.2 Reduced Nucleosome

The reduced nucleosome system (Figure 3.6) was built for computational expedience to test the effects of phosphorylation on the conformations of CTD within the nucleosome. This was built by removing atoms from the 211-bp nucleosome model. Truncated $\alpha 3$ and L1 linkers were modelled as the terminal 32-bp at each end of the 211-bp nucleosome. The nucleosomal DNA segment interacting with H1 was modelled as a 30-bp strand (15-bp each side of the dyad). To maintain the nucleosomal configuration, the bottom two base pairs of the linker strands were restrained with force constants of 1000 kJ mol^{-1} nm^{-2}. To maintain the curvature of the nucleosomal DNA segment, the phosphorus atoms within them were restrained with similar force constants. The H3, H4, and H2B histone tails were also included in the model and their last two tail amino acids were restrained at their point of attachments to the histone core. This condensed system was then solvated in an octahedral solvent box with an ionic concentration of 0.15 M Na/Cl for atomistic simulations.

Figure 3.6 Visualization of the reduced nucleosome system used to model the phosphorylated CTDs at an atomistic scale. The LH globular domain is in black, NTD in yellow and CTD in cyan. The α3- and L1-linker DNA strands are in red and purple respectively. The core histone tails included in the system are shown in white.

3.3.1.3 C-terminal Domain

The final 98 residues of *h*H1.0 were considered to constitute the C-terminal domain (CTD). The conformation of the CTD previously attached to the globular domain was independently condensed in a 1 ns simulation in GBSA implicit solvent [294, 295]. This condensed system was then solvated in a dodecahedron solvent box with an ionic concentration of 0.15 M Na/Cl for atomistic simulations.

3.3.1.4 N-terminal Domain

The first 26 residues of *h*H1.0 was considered to constitute the N-terminal domain (NTD). The sequence was built in extended conformation using Avogadro [296]. The extended conformation was then condensed using an initial 1 ns simulation in GBSA implicit solvent [294, 295]. This condensed system was then solvated in a dodecahedron solvent box with an ionic concentration of 0.15 M Na/Cl for atomistic simulations.

3.3.2 Simulation Setup

Simulations were performed using Gromacs 5 [182] patched together with Plumed 2.3.0 [297, 298]. The van der Waals interactions were cut-off at 10 Å and electrostatic interactions were calculated using the PME method [186, 187] with a real space cut-off of 10 Å and a reciprocal grid of spacing 1.4 Å. Before production runs, the systems were energy minimized using the steepest descent algorithm and equilibrated for 1 ns each in the NVT and NPT ensembles. The temperatures were maintained using the Bussi thermostat [190] and pressure was maintained at 1 bar using the Parrinello-Rahman [299] barostat. The force fields used for each simulation together with their timescales and number of replicas are described in Table 3.2.

Phosphorylated serine and threonine residues were modelled using the Amber compatible parameters developed by Homeyer et al. [300] and were introduced into the structure using PyMOL [293]. Given that nuclear magnetic resonance (NMR) chemical shifts of the N-H protons in both H1 CTD phosphoserine and phosphothreonine residues show that they are mainly dianionic at pH > 6 [301], their charges were set to -2. Two distinct phosphorylation states of the CTD were modelled – partial and full phosphorylation. The partially phosphorylated state was modelled by introducing the post-translational modifications (PTM) in 30% of Ser/Thr residues evenly distributed throughout the CTD – T119, S124, T141, T153 and S171. The fully phosphorylated state was modelled by introducing the PTM at all Ser and Thr residues within the CTD.

3.3.3 Enhanced Sampling Methodology

3.3.3.1 Biased Exchange Metadynamics

Biased Exchange Metadynamics (BE-Metad) simulations were performed on the full-length linker histone H1.0 when bound to either the full length 211-bp or reduced nucleosomes. Five replicas (4 biased and 1 unbiased) were used in the BE-Metad simulations and the Collective Variable (CV, ξ) biases were applied to the LH's CTD (Table 3.1).

Table 3.1 The Metadynamics biasing CVs of the CTD within each of the Biased Exchange Metadynamics replicas.

Replica Number	Collective Variable
1	None (Unbiased)
2	ξ_α
3	ξ_β
4	ξ_{cont}
5	ξ_{rg}

The CVs ξ_α and ξ_β are designed to measure the extent of alpha-helix and beta-sheet secondary structural elements within the CTD respectively and are defined following the work of Pietrucci and Laio [302] as

$$\xi_{\alpha/\beta} = \sum \frac{1 - \left(\frac{\Delta RMSD}{R_0}\right)^n}{1 - \left(\frac{\Delta RMSD}{R_0}\right)^m} \tag{3.1}$$

where R_0, n and m are 0.08 nm, 8 and 12 respectively. $\Delta RMSD$ is root mean square difference of six residue segments between the instantaneous configuration and ideal alpha/beta conformations where the backbone and C_β atoms are included in the RMSD calculation.

The ξ_{rg} CV monitors the radius-of-gyration (Rg) of the Cα atoms of the 98 residues comprising the CTD and is defined as

$$\xi_{rg} = \left(\frac{\sum_{i=1}^{98} m_i |r_i - r_{com}|^2}{\sum_{i=1}^{98} m_i} \right)^{1/2} \tag{3.2}$$

where r_i and m_i are the position and mass of atom C_α atom i. r_{com} is the centre of mass of the CTD C_α atoms.

The ξ_{cont} CV measures the number of electrostatic contacts between the basic CTD residues and the DNA phosphate backbone and is defined as

$$\xi_{cont} = \sum_{i \in H^+} \sum_{j \in O^-} \frac{1 - \left(\frac{r_i - r_j}{R_0}\right)^n}{1 - \left(\frac{r_i - r_j}{R_0}\right)^m} \tag{3.3}$$

where R_0, n and m are 0.2 nm, 8 and 12 respectively. H^+ is the set of positively charged sidechain hydrogens in basic Arg/Lys residues and O^- is the set of negatively charged oxygen atoms within the phosphate backbone.

The metadynamics biasing gaussians were deposited every 500 trajectory steps (1 ps) and exchanges between the replica were attempted every 5000 trajectory steps (10 ps). A bias-factor (ΔT) of 8 was used together with an initial hill height of 1.2 kJ. The gaussian widths σ were calculated as the half of the standard deviations of the CVs within a short 5 ns trial unbiased simulation of each system.

3.3.3.2 Replica Exchange with Solute Tempering

Replica Exchange with Solute Tempering (REST2) simulations were performed on isolated CTD protein segments in varying phosphorylation states – WT, Partially Phosphorylated, Fully Phosphorylated. In all three cases, 16 replicas were used whose pseudo-temperature was geometrically scaled as

$$T_i = T_{min} \left(\frac{T_{max}}{T_{min}} \right)^{(i-1)/(N-1)} \tag{3.4}$$

where T_{min} and T_{max} are 300 and 450 K respectively and N is the number of replicas – 16. Exchanges between the replicas were attempted every 10 ps and the system was simulated for a total of 200 ns per replica. The initial 50 ns of the trajectory was discarded for equilibration and the remaining trajectory was used for analysis.

3.3.3.3 Temperature Replica Exchange

The Temperature Replica-Exchange MD (T-REMD) simulations were performed to sample the conformations of the isolated NTD. The distribution of temperatures were estimated using the predictor of Patriksson and Van-der-Spoel [303] and an acceptance probability of 20%. This resulted in 56 replicas between the temperatures of 300 and 450 K. The simulations were performed for 250 ns in each replica for a total sampling of 14 μs and exchanges between the replicas were attempted every 10 ps. The initial 50 ns were discarded for equilibration and the coordinates from the lowest temperature (300 K) replica were considered for analysis.

3.3.4 List of Simulations

The simulations performed in this chapter are listed in Table 3.2 below.

Table 3.2 List of atomistic simulations performed to investigate the functioning of H1 within the nucleosome.

System	Force Field	Time	Simulation
211-bp Nucleosome + WT LH	Amber99SB-ildn/parmbsc0	5 × 1 μs	BE-Metad
211-bp Nucleosome + WT LH	Amber03ws/parmbsc0	5 × 750 ns	BE-Metad
211-bp Nucleosome + WT LH	Charmm36M/Charmm36 DNA	5 × 500 ns	BE-Metad
Reduced Nucleosome + WT LH	Amber99SB-ildn/parmbsc0	5 × 1 μs	BE-Metad
Reduced Nucleosome + Partial Phos. LH	Amber99SB-ildn/parmbsc0	5 × 1 μs	BE-Metad
Reduced Nucleosome + Full Phos. LH	Amber99SB-ildn/parmbsc0	5 × 1 μs	BE-Metad
Reduced Nucleosome (No LH)	parmbsc0	500 ns	Unbiased MD
Isolated WT CTD	Amber99SB-ildn	16 × 200 ns	REST2
Isolated Partial Phos. CTD	Amber99SB-ildn	16 × 200 ns	REST2
Isolated Full Phos. CTD	Amber99SB-ildn	16 × 200 ns	REST2
Isolated NTD	Amber99SB-ildn	56 × 250 ns	T-REMD
211-bp Nucleosome + LH Globular Domain	Amber99SB-ildn/parmbsc0	500 ns	Unbiased MD
211-bp Nucleosome (No LH)	Amber99SB-ildn/parmbsc0	500 ns	Unbiased MD

3.4 Results

3.4.1 Domain Specific Interactions of H1

The binding mode of the linker histone globular domain atop the nucleosome remains a major open question [288]. While some H1 subtypes have been crystallized in the on-dyad configuration, these have used truncation mutants that lack the unstructured terminal domains [9, 231]. Thus, the Biased-Exchange Metadynamics simulations of the 211-bp nucleosome were first used to analyze the impact of the IDR's interactions on the globular domain. Towards this, the single-linkage method [304] implemented within Gromacs [182] was used to identify the heaviest weighted cluster of LH Cα atoms. Within this method, a frame 'j' was assigned to cluster 'i' if the Root-Mean-Square-Displacement (RMSD) between them was less than 2.5 Å. The weight of each cluster W_i was then calculated as

$$W_i = \sum_{j \in i} w_j \tag{3.5}$$

where w_j is the WHAM weight of each frame 'j' assigned to cluster 'i'. Subsequently, 20-bp (-10 to +10) around the dyad were used to align this clustered

structure to that of the on-dyad *Xenopus laevis* H1.0 (Figure 3.7A) resolved with a 197-bp nucleosome (PDB ID: 5NL0 [231]).

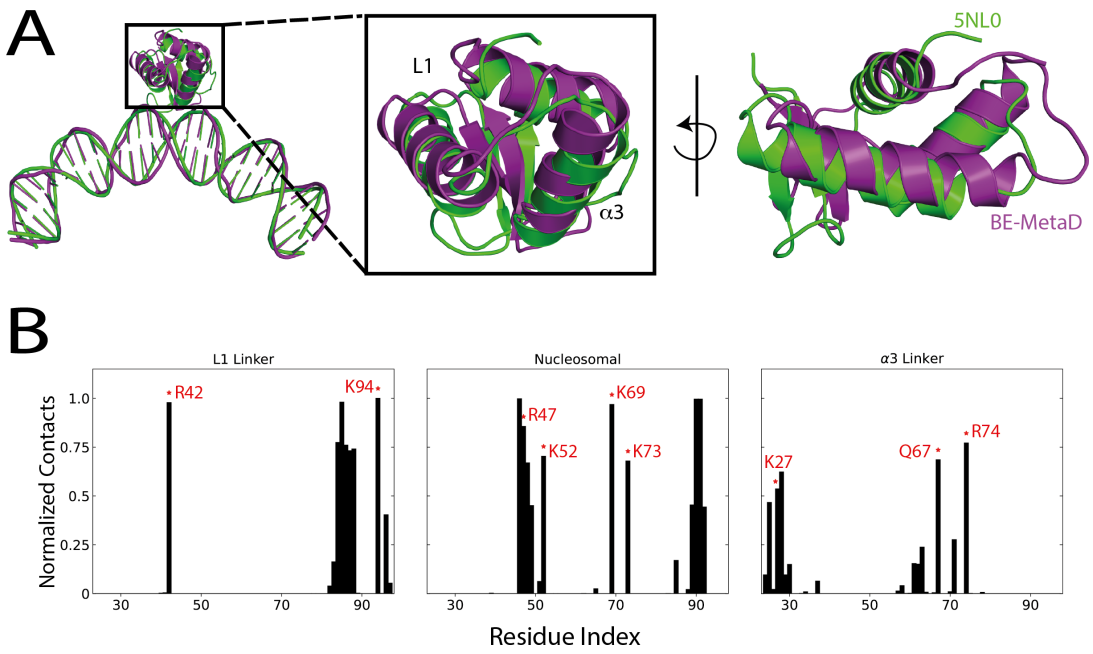


Figure 3.7 (A) The orientation of the H1.0 globular domain compared between the BE-Metad simulations with unstructured terminal domains (purple) and from the 5NL0 [231] crystallographic structure (green). The α3 helix and L1 linker of the LH are highlighted. **(B)** The per-residue interactions of the globular domain of the H1.0 linker histone with the L1, α3 and nucleosomal DNA. A contact was assumed if a non-hydrogen atom of the residue was within 3.2 Å of a non-hydrogen DNA atom. The contacts experimentally identified to mediate on-dyad nucleosomal binding are marked with a '*'.

Despite the inclusion of the terminal domains and the influence of their interactions, in five clusters that together encompass >50% of the conformational ensemble, the GD remained stable in the on-dyad conformation with only a negligible translocation perpendicular to the nucleosomal plane (Figure 3.7A). While this reduced diffusion of the GD away from the on-dyad configuration might be an artefact of inadequate sampling, more importantly, the orientations of the α3 and L1 linker histone segments remained unchanged to interact with their linker DNA strands. To further validate this, the normalized per-residue contacts of the globular domain with the two linker/nucleosomal DNA strands over the duration of the simulation were calculated. Significantly, the residues that mediated interactions in the MD simulations matched those identified by Isothermal Titration Calorimetry (ITC) [9] and FRAP [287] experiments to stabilize the LH's on-dyad configuration (Figure 3.7B).

Next, the open question of the localization of the unstructured terminal domains within the nucleosome was analyzed – an aspect only resolved thus

far through coarse statistics from FRET experiments [305, 306]. The N-terminal domain is short (26 residues) and is constrained by the on-dyad globular domain to 'face' the α3 linker. Thus, this disordered domain carrying six positively charged residues is restricted to interact with the first few base pairs (5-15 bp) of the α3 linker (Figure 3.8A). In contrast, the CTD while constrained by the on-dyad configuration to 'face' the L1 linker, is significantly longer at 84 residues. This long domain thus interacts with the first 23 bp of the L1 linker and subsequently extends to interact with α3 linker (Figure 3.8). Crucially, these results both corroborate and extend the cryo-EM observations of Bednar et al. [231] which suggested that the CTD interacts only with the L1-linker in a 197-bp nucleosome (25-bp linker arms).

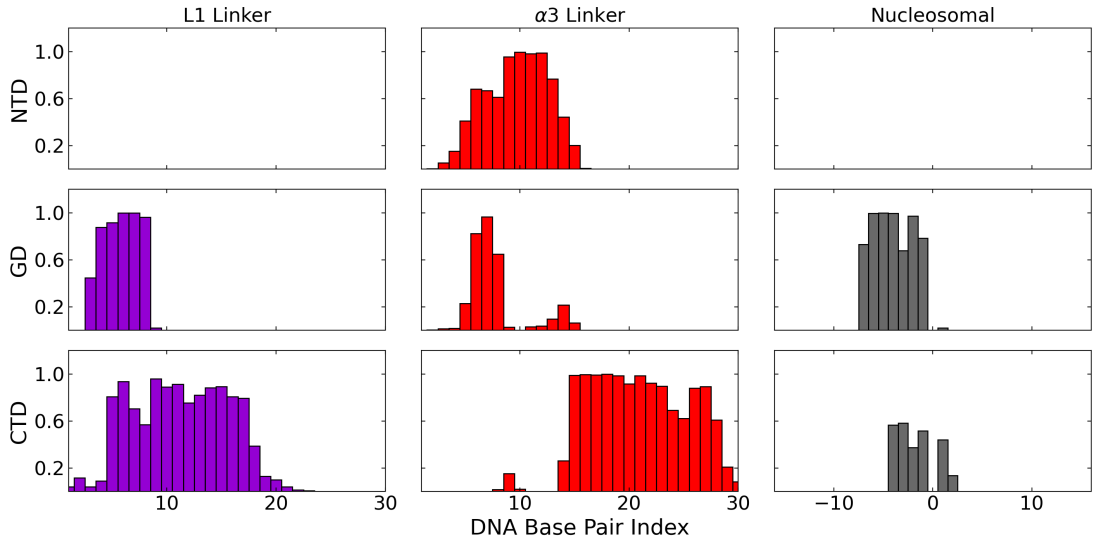


Figure 3.8 Patterns of H1-DNA interactions for the three H1 domains and the DNA regions within the nucleosome. A contact was assumed if a non-hydrogen atom of the base pair was within 3.2 Å of a non-hydrogen H1 atom of the domain. The contact frequencies are normalized to 1 using WHAM.

To validate if this 'bridging' of linker DNA strands by the CTD assists in nucleosome compaction, the results from the BE-Metad simulations were compared to those from unbiased simulations of the nucleosome without the LH and without the LH CTD (Table 3.3). The ability of the long CTD to interact with both linker strands impacts their configuration and brings them as close as $\sim$7 Å at the cross-over point. In-contrast, without the CTD, the linker strands remain independent of each other and at distances of > 2 nm apart.

Table 3.3 Mean minimum distance between the two linker DNA strand in systems with varying linker histone configurations. For the BE-Metad simulations with the whole length LH, the values are calculated from the unbiased ensemble to allow a direct comparison to the vanilla MD simulations.

System	Inter-Strand Distance (Å)
211-bp Nucleosome + WT LH	6.99 ± 1.8
211-bp Nucleosome + LH Globular Domain	22.45 ± 3.6
211-bp Nucleosome (No LH)	27.36 ± 2.7

Having validated the simulations' capability to reproduce the experimentally observed nucleosome compaction due to the LH [9] and specifically the CTD [232], the molecular mechanism of this functioning was next analyzed. Two different mechanisms of the CTD's functioning have thus far been proposed – 'sequence-dependant' and 'composition-dependant' models. The sequence-dependant model hypothesizes the nucleosome compaction capability to be focussed within functional subdomains of the CTD [232, 307, 308]. In contrast, the composition-dependant model hypothesizes the CTD's functioning dependant on the overall charge concentration of the domain and thereby the chromatin compaction capability uniformly distributed throughout [124, 255, 309].

The ensembles of the CTD obtained from the BE-Metad suggests that the long domain can be partitioned into four different structural regions that exhibit distinct patterns of interactions with DNA (Figure 3.9). The 'four' subdomains were termed 'Beta', 'Loop1', 'Loop2' and 'End'. Each of the four domains interact with at least two DNA segments and together effectively shield repulsion to aid compaction. The 'Beta' region involves the only stable secondary structural element present in the CTD that occurs in $\sim$80% of the frames analysed (Figure 3.11). This is a 22-residue anti-parallel β-sheet motif between residues $_{102}$KKSVAFKKTKK$_{112}$ and $_{140}$ATPVKKAKKKA$_{151}$ that is squeezed between the L1 and α3 linkers at their closest point to each other. The Loop 1 subregion is a 20-residue loop subtended by the β-motif and comprises residues $_{113}$EIKKVATPKKASKPKKAASKA$_{139}$ that interacts more strongly with the α3 linker (17-32 bp) and only transiently with the L1 linker (18-23 bp). Loop 2 is a 24-residue loop that interacts with both linker DNA arms but away from their closest points. The final 'End' region is formed by the last 19 CTD residues and interacts transiently with the nucleosomal DNA and the L1 linker. The atomistic resolution observations here both validate and elucidate the coarse results obtained from FRET experiments where Fang et al. [310] observed that a donor probe on residue 194 displayed maximal fluorescence intensity to an acceptor probe placed near the dyad. Similarly, Heidarsson et al. [126] observed

a donor probe on residue 194 to display maximal fluorescence intensity to acceptor probes placed at base pair 12 of both L1 and α3 linkers.

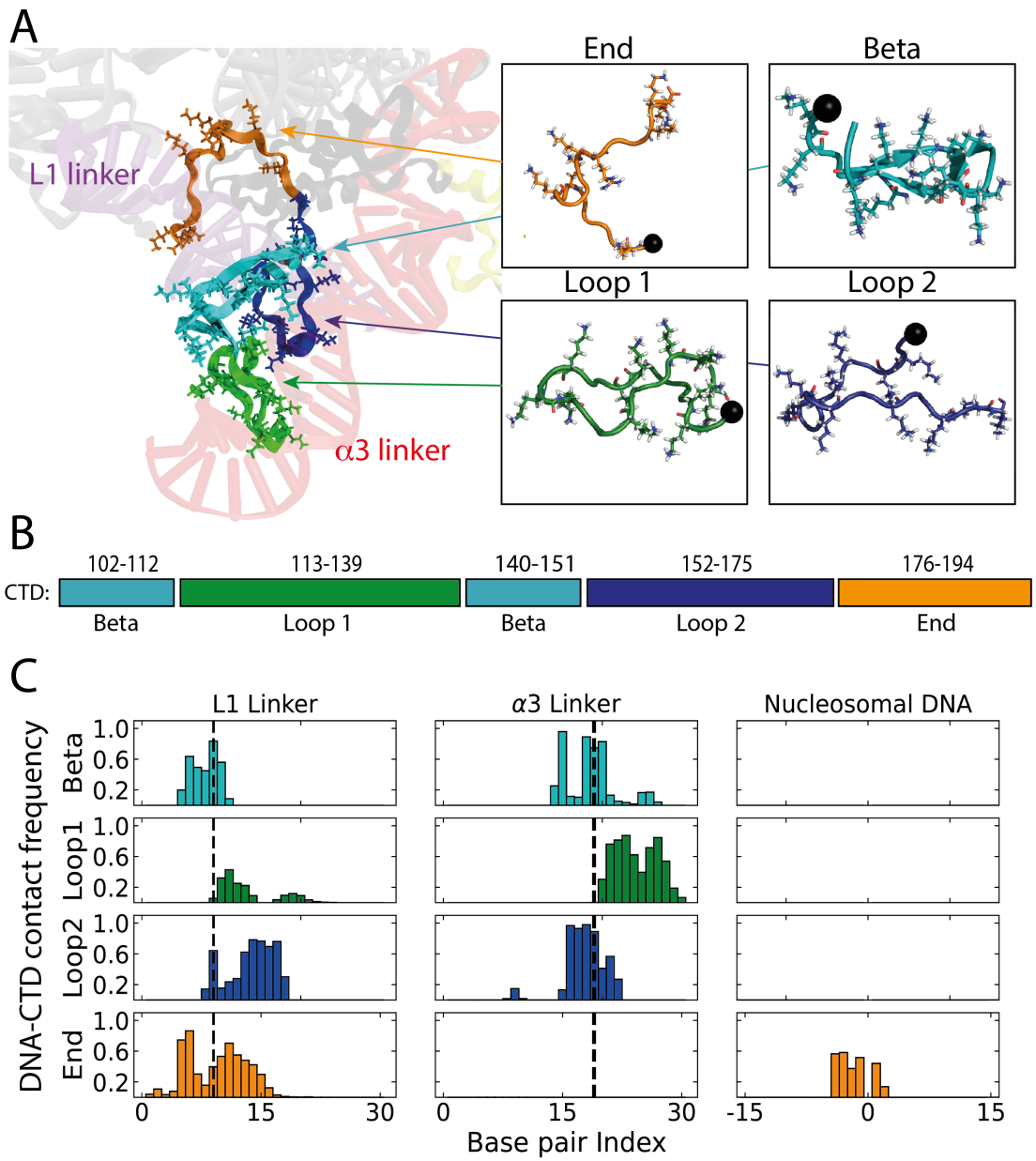


Figure 3.9 Partitioning of H1 CTD interaction patterns. **(A)** Structure of the CTD within the most populated cluster of the simulation illustrating the four subdomain regions and the basic residues within them. **(B)** The residue numbers of the four domains. **(C)** DNA interaction patterns of the four subdomains. A contact was assumed if a non-hydrogen atom of the base pair was within 3.2 Å of a non-hydrogen atom of the domain. The base pairs corresponding to the closest point between the L1 and α3 linkers are illustrated with a dotted line.

More importantly, the results reconcile the hypothesized 'sequence-dependant' and 'composition-dependant' models of CTD functioning. The four subdomains each containing ~10 basic R/K residues interact with two DNA segments (Fig-

ure 3.9) and thereby support the proposition that the nucleosomal compaction capability is distributed throughout the domain. However, Syed et al. [232] observed that the nucleosomal compaction of *h*H1.5 is focussed within its residues 120-130 which when aligned to the sequence of this work's *h*H1.0 matches the stable β-sheet motif (Figure 3.10).

```
H1.5   1 MSETAPAETATPAPVEKSPAKKKATKKAAGAGAAKRKATGPPVSELITKAVAASKERNGLSLAALKK  67
H1.0   1 -----MTENSTSAPAAK-PKRAKASKKSTDH---------PKYSDMIVAAIQAEKNRAGSSRQSIQK  52

H1.5  68 ALAAGGYDVEKN-NSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKAKKAGAAKAKK 133
H1.0  53 YIKS-HYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAK-----SDEPK-KSVAFKKTKK 112

H1.5 134 PAG--ATPKKAKKAAGAKKAVKKTPKKAKKPAAAGVKKVAKSPKKAKAAAKPKKATKSPAKPKAVKP 198
H1.0 113 EIKKVATPKKASK---PKKAASKAP--TKKPKATPVKKA-----KKKLAATPKKA-KKPKTVKAKPV 168

H1.5 199 KAAKPKAAKPKAAKPKAAKAKKAAAKKK                                        226
H1.0 169 KASKPKKAKP--VKPKAKSSAKRAGKKK                                        194
```

Figure 3.10 Alignment of the H1.0 and H1.5 linker histone sequences generated using Muscle [311, 312] and visualized using Jalview [313, 314]. The residue patches identified by Syed et al. [232] to dominate H1.5 nucleosome compaction and observed to form part of the β-motif here in H1.0 are highlighted in blue.

3.4.2 Divergent Disordered Terminal Region Behaviour

The two disordered amino- and carboxyl-terminal domains display distinct interaction patterns with DNA (Figure 3.8) and contribute to the compaction of linker strands (Table 3.3). However, the mechanism of this functioning of the domains has remained an open question with multiple conflicting experimental evidence [272]. Circular Dichroism of both the CTD and NTD in TriFlouro-Ethanol (TFE) [315–320] have suggested their capabilities to form significant α-helical conformations. Contrarily, neither cryo-EM [232, 291] or X-ray crystallography [9, 231] have visualized these 'structured' segments of the domains. Moreover, Borgia et al. [125] and Turner et al. [123] have recently identified the ability of the H1 CTD to form high-affinity complexes with anionic counterparts while remaining disordered. The structural characteristics of the domains within the nucleosome were thus next analyzed.

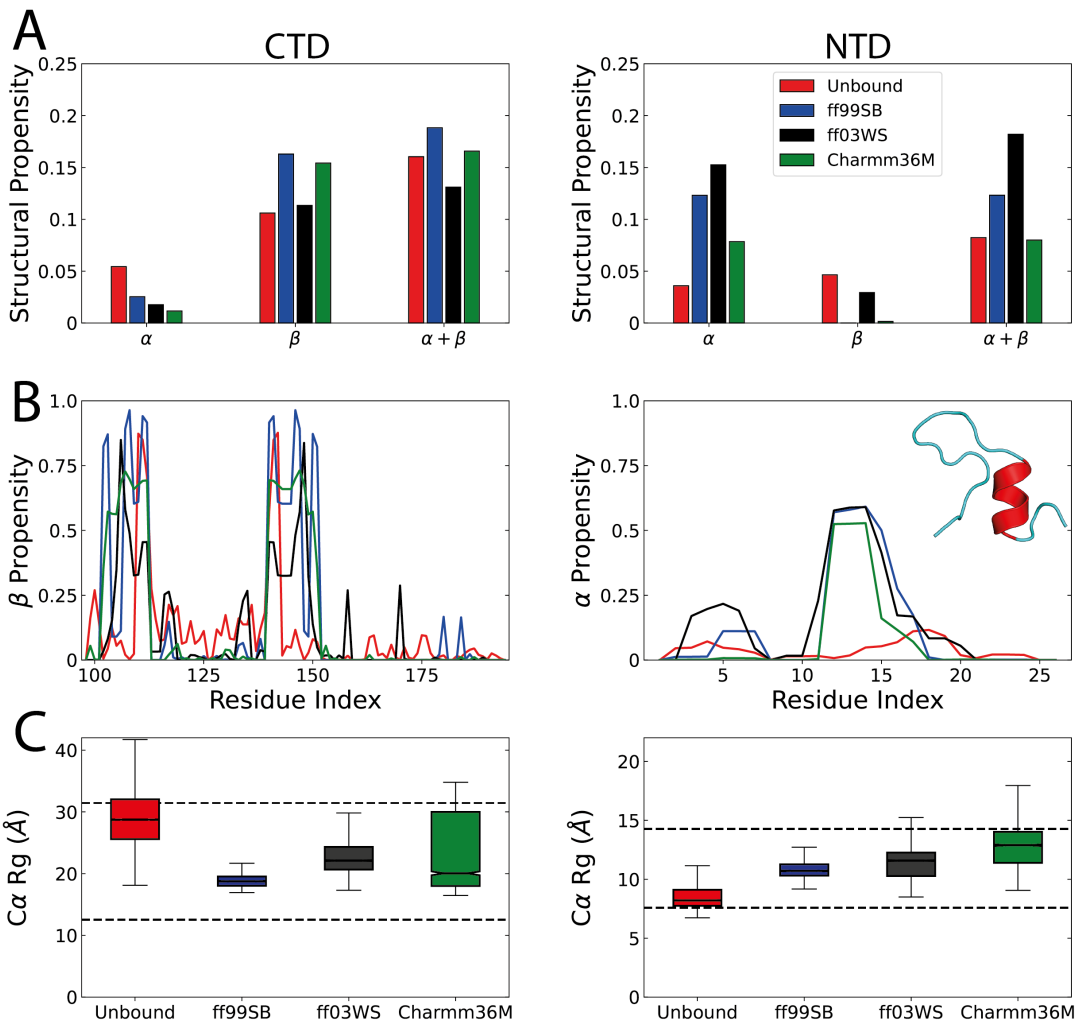


Figure 3.11 (A) Secondary structural content from simulations of the domains in the unbound and nucleosome-bound states for three different force fields. **(B)** Per-residue β and α secondary content for the CTD and NTD respectively in the unbound and nucleosome-bound states. Inset shows the clustered conformation of the NTD when nucleosome bound. **(C)** Cα Rg for the domains in the unbound and nucleosome-bound states. The boxes extend from the lower to upper quartile values and whiskers span from the 5th to the 95th percentile. The compaction for globular and coil proteins calculated from experimental regression (Equations 3.6, 3.7) are shown as dotted lines.

To quantify secondary structural content from Metadynamics simulations, the per-residue per-frame structure was first assigned using DSSP [321] and a simplified system where both alpha- and 3_{10}- structures are together labelled α. Subsequently, the WHAM weights were used to calculate the α/β for a residue 'i' using Equation 2.67 where the per-frame O_i was set to 1 if DSSP predicted a helical/beta structure or 0 otherwise.

The simulations suggest that independent of the force field considered, the total secondary structural content remains unchanged at ∼15% (Figure 3.11A). The CTD appears to retain its unstructured nature and thus contrary to prior hy-

pothesis [320], it's compaction of the nucleosome is not the result of a disorder-to-order transition. The only stable secondary element present in the CTD is a 22-residue anti-parallel beta-sheet motif between residues $_{102}$KKSVAFKKTKK$_{112}$ and $_{140}$ATPVKKAKKKA$_{151}$ (Figure 3.11B). While the span and extent of this β-structure varies with the force field used, significantly, this motif that dictates the orientation of the CTD (Figure 3.9) is even present in REST2 simulations of the isolated domain. The radius-of-gyration box plots further illustrate the parallels between the CTD's bound and unbound states. The 97-residue long IDR's compaction reduces minimally by $\sim$5 Å between the states and probably attributable to the structural restraints placed by the nucleosome [322]. In both cases, this compaction level lies between the theoretical values for globular [323] and random-coil [324] proteins of similar length 'N' calculated as

$$R_g^{glob} = 2.2N^{0.38} \tag{3.6}$$

$$R_g^{coil} = 2.02N^{0.60} \tag{3.7}$$

The behaviour of the 26-residue NTD is significantly different than that of the CTD in both the bound and unbound states. While remaining predominantly disordered when unbound, in-contrast to the CTD, the NTD adopts a collapsed conformation whose compaction is close to the value predicted for globular proteins of similar length (Figure 3.11C). Upon nucleosomal binding, the domain shows a slight decrease in compaction but a significant increase in α-helical content. Crucially, this disorder-to-order transition of the NTD is regardless of the force field considered (Figure 3.11B) and occurs within the $_{11}$AKPKRAKASK$_{20}$ segment of the NTD where basic residues are concentrated (Figure 3.10).

In addition to understanding the domains' roles in DNA compaction, their contrasting behaviours when isolated offers a molecular explanation for their experimentally observed contrasting roles in mediating H1-nucleosome interactions [251, 253]. Extended states of the CTD when isolated might allow it to initiate the process of nucleosome recognition through long-range electrostatic interactions [251] while the collapsed conformations of the NTD might preclude this [253].

3.4.3 Breaking the Nucleosome Symmetry

The three H1 domains have divergent structural characteristics when within the nucleosome and have distinct interaction patterns with the two entering

and exiting linker DNA strands (Figure 3.8). The interactions while aiding in compacting the L1 and α3 linkers, are asymmetric between the two DNA strands. The impact of these asymmetric interactions on the DNA strands were thus analyzed next. For this, the nucleosome was first projected onto the nucleosomal plane and the orientations of the strands were then compared (Figure 3.12). At first glance, the two DNA arms are curved, approach close to each other and criss-cross the nucleosomal dyad axis.

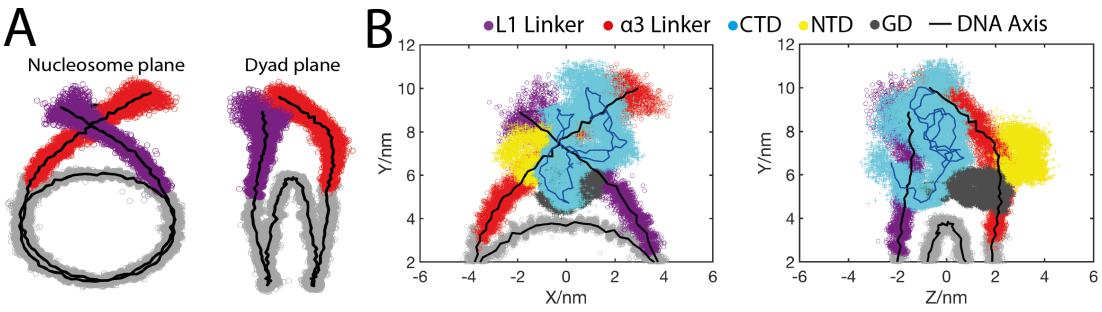


Figure 3.12 (A) Two dimensional projections of the nucleosome onto the nucleosome and dyad planes using every 100th frame of the unbiased ensemble. **(B)** Two dimensional projections of the nucleosome onto the nucleosome and dyad axes together with the positioning of the three H1 domains. The L1 and α3 linkers are shown as purple and red scatter plots respectively. The H1 CTD, NTD and GD domains are shown in cyan, yellow and grey respectively.

The criss-crossing of the DNA arms are however asymmetric with a higher degree of bending exhibited by the α3 versus the L1 linker strand. In-turn, this results in the DNA criss-crossing point occurring unevenly at bp 19 of the α3 linker in-contrast to bp 9 of the L1 linker (Figure 3.13). Crucially, this asymmetric crossing point correlates with the asymmetric CTD-DNA interactions (Figure 3.9) where the closest inter-strand point correlates with the β-segment's positioning. Figure 3.13 further validates the role of the disordered terminal domains in inducing the asymmetric nucleosome where control unbiased MD simulations of 211-bp nucleosomes with only the H1.0 GD show DNA linkers that approach each other symmetrically.

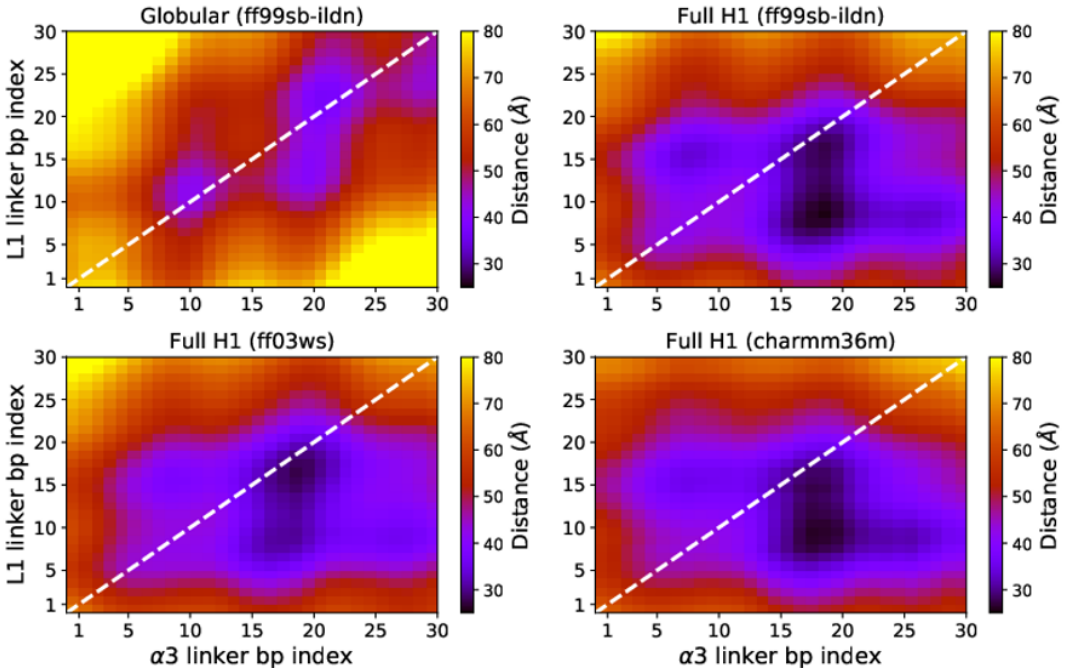


Figure 3.13 Inter-base pair distances between the two linker arms showing the formation of a symmetric (globular) or asymmetric (full H1.0) topology of linker DNA arms in simulations. The 1:1 bp indices for the two linker arms are shown as a white dotted line.

To quantify this asymmetric curvature of DNA strands, the procedure of Pasi et al. [325] was used to calculate the strands' dimensionless helical axis curvature. For this, Curves+ [326] was first used to fit the curvilinear helical axis of the DNA strand as a set of point U_i centred at each base pair i for each trajectory frame. For each base pair i, two local circles are then fitted across the points $\{U_{i-1}, U_i, U_{i+1}\}$ and $\{U_i, U_{i+1}, U_{i+2}\}$. The radii of these two circles are then calculated as r_i and r_{i+1} respectively. The curvature of base pair step C_i was then calculated as

$$C_i = 40 \cdot \sqrt{\left(\frac{1}{r_i} \times \frac{1}{r_{i+1}}\right)} \tag{3.8}$$

where the scaling factor of 40 normalizes the value to ensure the curvature of the 1KX5 [16] nucleosomal DNA is calculated as 1. Regardless of the force field, the α3 linker is significantly more curved than the L1 linker and in-fact as curved as DNA wrapped around the nucleosomal core (Table 3.4).

Table 3.4 Curvature of the $\alpha 3$ and L1 linker DNA arms for different force field systems calculated using the procedure of Pasi et al. [325].

System	$\alpha 3$ curvature	L1 curvature
ff99SB-ildn	1.13	0.78
ff03WS	1.13	0.86
Charmm36M	1.21	0.81
GD ff99SB-ildn	0.74	0.73

The breaking of the two-fold nucleosomal symmetry while seemingly negligible on the nucleosomal scale, can have a significant impact on chromatin structure at the nanoscale [231]. Specifically, the two nucleosomal arrangements proposed thus far – the zigzag and solenoidal models vary particularly in the interaction patterns of the cores (Figure 1.4). However, varying the relative orientations of the entering/exiting linker DNA strands would affect the positioning of adjacent nucleosomes and in-turn the formation of higher-order chromatin structures [284, 327].

3.4.4 CTD Phosphorylation within the Nucleosome

The Molecular Dynamics simulations demonstrate that electrostatic contacts dominate the interactions of the three H1 domains with DNA. Phosphorylation is a physiologically relevant post-translational modification capable of shifting the interaction free-energy of the histone-DNA interface by adding both steric bulk and negative charge [79, 80]. The simulations were thus next used to analyze the impact of this PTM on the CTD's structure and DNA compaction.

When unbound, the degree of the CTD's compaction increases monotonically as the degree of phosphorylation increases (Figure 3.14A) – a trend readily explained by the increasing number of favourable intra-protein contacts between the lysines and phosphorylated residues. Surprisingly, this monotonic effect of phosphorylation on the CTD is not retained when within the nucleosome. While the nucleosome-bound hyperphosphorylated CTD exhibits a similar distribution of R_g and R_{EE} as the WT, the nucleosome-bound partially phosphorylated CTD appears significantly decondensed (Figure 3.14B). The cyclical effect of phosphorylation levels on CTD compaction can plausibly be attributed to the asymmetric impact of the introduced negative charges. In the partially phosphorylated state, the basic lysine residues fluctuatingly interact with the phosphorylated serines and the DNA backbone to enhance the dy-

namic nature of CTD-DNA binding/unbinding events. However, in the fully phosphorylated state, the lysine-phosphorylated residue interactions dominate and thereby allow the reverting of the CTD to a condensed state similar to that of the WT.

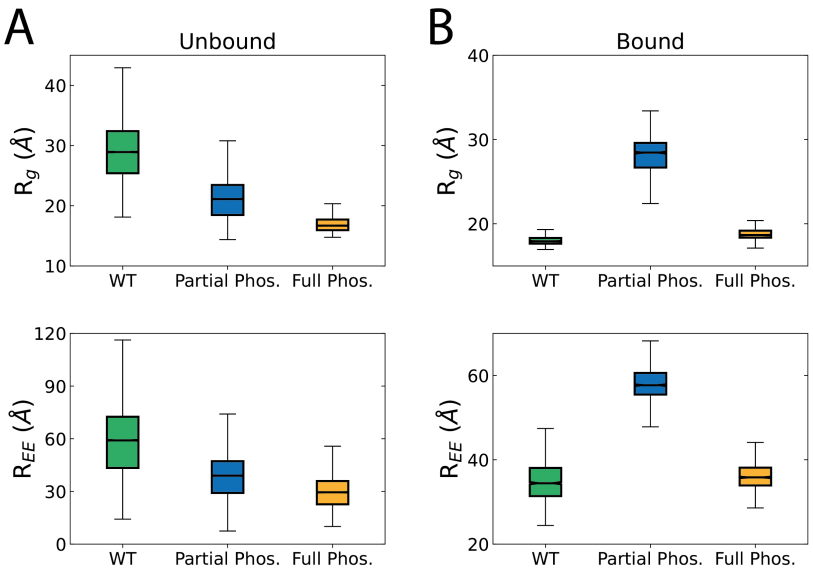


Figure 3.14 (A) Radius of gyration (Top) and End-to-End distance (Bottom) of the H1.0 CTD from REST2 simulations of the IDR when isolated in solution. **(B)** Radius of gyration (Top) and End-to-End distance (Bottom) of the H1.0 CTD from BE-Metad simulations of the IDR when within the reduced nucleosome. The boxes extend from the lower to upper quartile values and whiskers span from the 5th to the 95th percentile.

Crucially, this cyclical trend observed in CTD compaction is in-congruous with the impact of the phosphorylation states on chromatin compaction both *in vivo* [274–276] and *in vitro* [279, 280]. To validate if this difference in CTD compaction allows variation in the domain's capability to 'bridge' DNA, the mean minimum distance between the two linker strand in the three systems were compared (Table 3.5). The compact nature of CTD in WT and Fully phosphorylated states allows the domain to occupy the interstitial space between the linkers and thereby 'bridge' the strands. In-contrast, the partially phosphorylated CTD in an extended conformation preferentially interacts with the L1 linker strand (Figure 3.15) and thus inadequately shields the electrostatic repulsion between the linker strands. Thus, the introduction of 5 phosphorylated residues nearly doubles the mean inter-strand distance from 4.96 Å to 9.83 Å. Cyclically, the introduction of more PTMs in the fully phosphorylated state, reverts the inter-strand to 5.84 Å – nearly the same as that of the WT.

Table 3.5 Mean minimum distance between the two linker DNA strands in the reduced nucleosome systems with varying phosphorylation states. For the BE-Metad simulations, the values are calculated from the unbiased ensemble.

System	Inter-Strand Distance (Å)
Reduced Nucleosome + WT LH	4.96 ± 2.15
Reduced Nucleosome + Partial. Phosph. LH	9.83 ± 3.94
Reduced Nucleosome + Full Phosph. LH	5.84 ± 2.58
Reduced Nucleosome (No LH)	15.35 ± 2.35

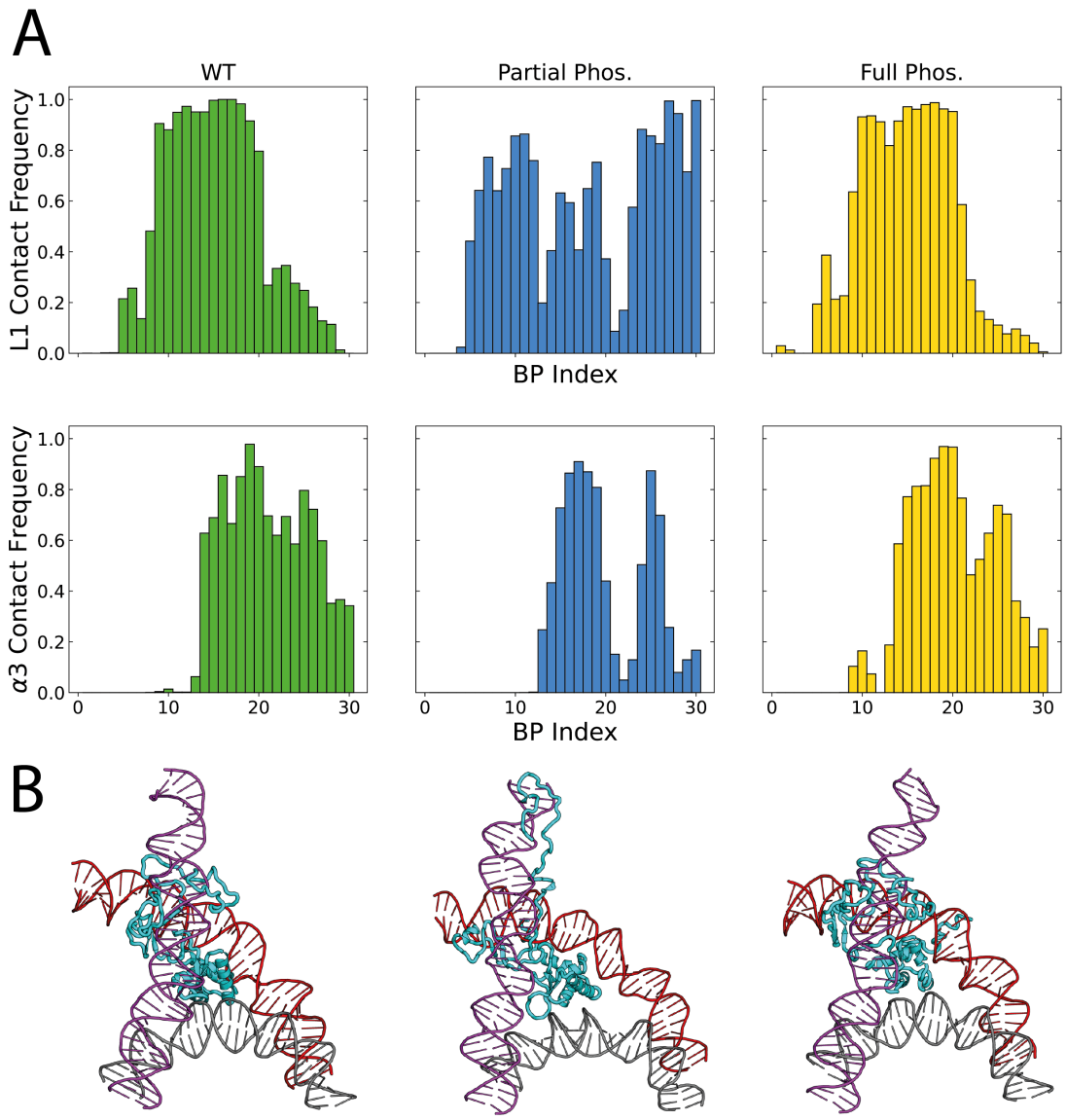


Figure 3.15 (A) Per-base pair interactions of the L1 (Top) and α3 (Bottom) linker DNA strands with the CTD at varying phosphorylation states. **(B)** Representative conformations of the CTD at varying phosphorylation states within the reduced nucleosome system. The α3 and L1 linkers are in red and purple colour respectively. The LH is in cyan.

3.5 Conclusion and Further Work

In this chapter, enhanced-sampling Molecular Dynamics simulations are used to characterize at atomistic resolution, the H1 linker histone protein when bound to the nucleosome. The modelling reveals that the protein's CTD remains flexible and highly disordered when bound to a nucleosome. These results offer evidence towards answering a long-standing open question of the domain's conformations within chromatin and are crucially consistent with recent breakthrough experiments demonstrating the CTD's 'fuzzy' nature when bound to free DNA [123] or anionic proteins [125].

Despite this disordered nature, the domain adopts a compact conformation that allows its residence between the nucleosome's linker arms. The fluctuating simultaneous interactions with both linker DNA permits the domain to bring the arms into spatial proximity albeit in an asymmetric manner. The entering and exiting DNA strands curve and criss-cross the dyad axis unevenly to lead to a loss of two-fold nucleosomal symmetry. The results are crucially consistent with Cryo-EM predictions [231] that hypothesized this loss of symmetry to affect positioning of adjacent nucleosomes and thereby even the structure of chromatin [322].

The predominance of electrostatic interactions in inducing these conformations of the positively charged CTD imply that the domain's compaction and flexibility can be readily affected by the introduction of post-translational modifications. Specifically, a PTM like phosphorylation that introduces both steric bulk and anionic charge can have a significant impact on conformations. Partial phosphorylation of the CTD resulted in extended conformations of the domain and thereby a reduced compaction of the linker arms. However, complete phosphorylation of the Ser/Thr residues surprisingly reversed the compaction of both the CTD and the linker arms. Crucially, this cyclical impact of phosphorylation observed in simulation is congruous to the PTM's role within the cell cycle [276].

The results of the atomistic simulations within this chapter while interesting themselves, offer an avenue to better predict oligonucleosome conformations at the nanoscale. With atomistic simulations of systems larger than a couple of nucleosomes inexpedient [328], a slew of mesoscale models have attempted to study the impact of many factors like DNA linker length [35], core histone tail PTMs [15], presence/absence of H1 [329, 330] and asymmetrical H1 binding [284] on systems upto 100 nucleosomes large. The atomistic simulation results

are readily mappable to the coarse-grained LH parameters within these models to allow their impact on nanoscale chromatin to be more accurately modelled.

Chapter 4

Subtype-dependant functioning of H1 Amino-Terminal Domains

Contents

This Chapter is based on the manuscript:

"*Protein disorder-to-order transition enhances the nucleosome-binding affinity of H1.*" **Akshay Sridhar**, Modesto Orozco, and Rosana Collepardo-Guevara. Nucleic Acids Research, 2020, 48, 5318-5331.

4.1 Introduction

The enhanced-sampling simulations discussed in Chapter 3 discern the structure and interactions of the three *h*H1.0 domains together with their roles in modulating the structure of nucleosomal DNA. The tripartite structure of the linker histone and its interactions are dramatically different than those of the core histones within the octamer (Section 1.1) and has even lead to the postulation that the two core- and linker-histone proteins have divergent evolutionary backgrounds [331–333]. The divergence between the core and linker histones is further accentuated by the multiplicity of subtype variants expressed in different eukaryotes and even within the same organism [244].

4.1.1 Subtypes

The four core histones while existing in variant forms [334], are limited in comparison to H1 where up to 11 distinct subtypes have been identified in humans [335–337]. Although the specific roles of the linker histone subtypes and the evolutionary basis for this multiplicity have not been completely understood [338, 285], the isoforms are expressed in a cell-cycle and tissue-dependant manner [233, 234, 244]. Table 4.1 summarizes the mammalian H1 subtypes of which 7 are somatic, 3 are sperm-specific and 1 is oocyte-specific.

Table 4.1 The 11 H1 subtypes, their characteristics and expression patterns. The nomenclature of isoforms follows the proposal of Talbert et al. [337] and the information compiled from Happel et al. [244] and the Uniprot Database [292].

H1 Subtype	Old Name	Length (residues)	Expression
H1.0	$H1^0$, H5	193	Somatic (Differentiated)
H1.1	H1a	214	Somatic
H1.2	H1c	212	Somatic
H1.3	H1d	220	Somatic
H1.4	H1e	218	Somatic
H1.5	H1b	225	Somatic
H1.6	H1t	206	Sperm
H1.7	H1T2	233	Sperm
H1.8	H1oo	345	Oocyte
H1.9	H1LS1	230	Sperm
H1.10	H1x	212	Somatic

The five H1.1-H1.5 subtypes are generally considered as replication dependant isoforms and are primarily expressed during the S phase of somatic cells. The sequences of the five isoforms are highly conserved within the globular domain with a homogeneity of upto 99% [339]. Despite this similarity, the subtypes regulate the expression of different subsets of genes [340] and interact differently with chromatin modulators like transcription factors [341] and heterochromatin proteins [342].

The H1.0 is the shortest linker histone subtype and displays a significant sequence divergence of up to 55% from the H1.1-H1.5 subtypes (Figure 4.1) [244, 339]. This differentiation of H1.0 is exemplified by its encoding gene being located in chromosome 22 unlike chromosome 6 for the core and other somatic linker histones [243]. The subtype is however significantly conserved across species. The distinctive H1.0 behaviour and its eukaryotic ubiquity have together allowed it to be the best studied LH subtype [343]. In fact, the only two LH-bound nucleosomal structures resolved thus far have both been with orthologs of H1.0 [9, 231]. The H1.0 isoform is predominantly found in tissues with low levels of cell proliferation [344, 345] and its expression in actively dividing tissues is kept at comparatively lower levels [346]. This localization has prompted several postulations of H1.0 being the most gene-repressive subtype [347] and thus its knockout having no effects on embryonic development [348].

The H1.10 (H1x) linker histone is the least characterized among the somatic subtypes and displays the greatest sequence dissimilarity to the ubiquitously expressed H1.1-H1.5 subtypes (Figure 4.1) [244]. Similar to the H1.0's distinctive genomic background, H1x is located on chromosome 3 and away from other somatic subtypes [349]. While the genomic significance of H1x has yet to be identified, experiments [350, 351] have postulated an alternative role for this subtype in organizing the alignment and segregation of chromatin during mitosis. This mitosis specific role of H1x has linked its localization to tumour cells [352] and genomic segments with cancer-related methylated patterns [353].

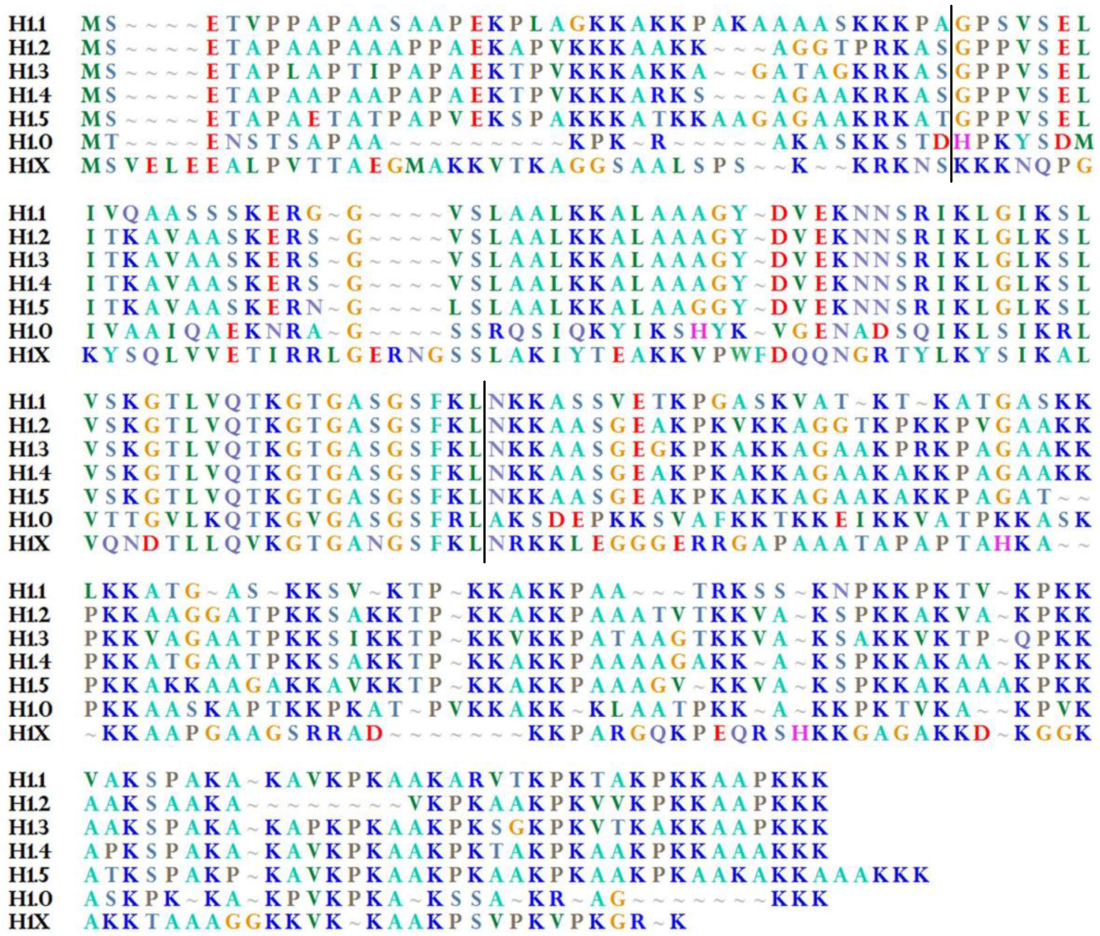


Figure 4.1 Sequence alignment of somatic human H1 subtypes H1.0-H1.5 and H1x (H1.10). The breakup of the sequence into amino, globular and carboxyl domains is shown by vertical lines. The basic Arg/Lys residues are shown in blue, acidic Glu/Asp residues in red and the non-polar residues in green. The Pro, Gly and Ala residues are shown in grey, yellow and cyan respectively. Image modified from Di Liegro et al. [243].

In an attempt to determine the structural mechanisms of the multi-faceted roles of H1 subtypes in cell function, several studies have attempted to determine the binding affinity and chromatin compaction capabilities of the isoforms [354]. However, the studies have insofar provided incomplete and contradictory results dependant on experimental methods and cell types. Talasz et al. [355] classified the in vitro mononucleosome binding affinity of mammalian subtypes as low (H1.5), intermediate (H1.1) and high (H1.2, H1.3, H1.4). However, these results differ from the cell-based FRAP results of Th'ng et al. [356] which classified the binding affinities as low (H1.1, H1.2), intermediate (H1.0, H1.3) and high (H1.4, H1.5). Table 4.2 summarizes the hitherto reported relative binding affinities of the somatic subtypes and their respective experimental methods.

Table 4.2 Summary of the hitherto reported nucleosomal binding affinities of the somatic linker histone subtypes. Some experiments only compared two specific subtypes and thus have no 'intermediate' ranking.

ID	Low	Intermediate	High	Exp. Method	Reference
1	H1.5	H1.1	H1.2,H1.3,H1.4	In vitro Gel Assay	Talasz et al. [355]
2	H1.1,H1.10	H1.0,H1.2,H1.3	H1.4,H1.5	In vitro Gel Assay	Clausell et al. [354]
3	H1.1	H1.2,H1.5	H1.0,H1.3,H1.4	In vitro Gel Assay	Orrego et al. [357]
4	H1.1,H1.2	H1.3,H1.4,H1.5	H1.0	FRAP	Flanagan et al. [253]
5	H1.1,H1.2	H1.0,H1.3	H1.4,H1.5	FRAP	Th'ng et al. [356]
6	H1.10	-	H1.0	FRAP	Okuwaki et al. [256]
7	H1.2	-	H1.0	FRAP	George et al. [339]
8	H1.2	-	H1.0	FRAP	Vyas and Brown [255]
9	H1.10	-	H1.2	FRAP	Takata et al. [351]

4.1.2 Amino Terminal Domain

The six primary somatic H1.0-H1.5 subtypes are significantly conserved within the structured globular domain (Figure 4.1) [339, 358]. The differential behaviour of the subtypes are thus hypothesized to stem from the heterogeneity within the unstructured terminal domains that vary in both length and sequence [255]. However, a slew of experiments that attempted to validate this hypothesis of the IDR's roles through the use of domain swap mutants achieved contrasting results. Vyas and Brown [255] observed that a swap of the NTD between the *m*H1.0 and *m*H1.2 resulted in a swap of their nucleosomal affinities while a similar swap of their CTDs had no effect. Contrarily, a swap of the H1.1 and H1.5 CTDs resulted in a swap of their binding characteristics while a NTD swap had no effect [253, 356].

The Carboxyl-terminal domain of the H1 subtypes are up to 100 residues long and homogeneously positively charged throughout. In-contrast, the H1 NTD sequences are both significantly shorter and heterogeneous with two distinct subregions [359]. The distal subregion located away from the globular domain is enriched in Ala, Pro and other non-aromatic hydrophobic residues. In contrast, the basic Arg/Lys residues crucial for electrostatic DNA interactions are concentrated within the subregion adjacent to the globular domain [359]. Figure 4.2 illustrates this subdivision into 'hydrophobic' and 'basic' subregions within the NTDs of the three hitherto best characterized H1.0, H1.1 and H1.2 subtypes.

Figure 4.2 Sequence comparison of the *h*H1.0, *h*H1.1 and *h*H1.2 linker histone NTDs illustrating their hydrophobic and basic subregions.

The NTD basic subregions of the three linker histone subtypes possess two important trends that correlate with the subtypes' nucleosomal affinities. Despite variance across experimental methods, most studies have classified the H1.0 somatic subtype as possessing the highest nucleosome binding affinity, followed by H1.2 and finally H1.1. This affinity trend of the three subtypes is consistent with the H1.0 subtype overexpression being associated with quiescent chromatin/repressed gene expression [360], and the H1.1 and H1.2 subtypes localizing in transcriptionally active euchromatic regions [356]. In a trend correlated to affinity, the H1.0 basic subregion possess the shortest length and highest concentration of positive residues; while the H1.1 basic subregion is both longest and has the lowest positive charge concentration (Table 4.3).

Table 4.3 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 R/K 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

4.2 Aims and Summary

The Linker Histones are the most divergent among the histone family of proteins. However, the specific roles of the subtypes and their individual mechanisms of functioning remain an open question. In this Chapter, I use a combination of enhanced sampling atomistic Molecular Dynamics simulations to – 1) analyze the amino-terminal domain's functional mechanism when bound to the nucleosome and 2) develop a hypothesis for the domain's role in differentiating the functioning of the H1 subtypes.

4.3 Methods

4.3.1 System and Simulation Setup

The sequences of the *h*H1 subtypes were obtained from the Uniprot database [292] for three isoforms – H1.0, H1.1 and H1.2. The NTD domains of the three subtypes were built in an extended conformation using Avogadro [296] without capping of the terminal residues. The linear configurations were condensed using an initial 5 ns simulation in GBSA implicit solvent [294, 295] and then used as the initial configuration for the simulations.

The simulations were performed using Gromacs 2016 [182] patched with Plumed 2.3.0 [297, 298] in explicit solvent and 0.15 M Na/Cl. The simulations used a combination of two IDP specific force fields: Charmm36M [168] and Amber99SB-ildn [166] with Helix-Coil transition specific corrections to the backbone dihedrals [361] and charge [362] (ff99sb*-ildn-q). The parameters of Aqvist [363] and Dang [364] were used to model ions within the Amber force field and those of Beglov and Roux [365–367] were used within the Charmm force field. The DNA strands were built 20-bp long in B-DNA form using the Nucleic Acids Builder module of Amber 16 [368] and modelled using the Charmm36 DNA [369] or parmbsc0 [370] parameters. The van der Waals interactions were cut-off at 10 Å and electrostatic interactions were calculated using the PME method [186, 187] with a real space cut-off of 10 Å and a reciprocal grid of spacing 1.4 Å. Before production runs, the systems were energy minimized using the steepest descent algorithm and equilibrated for 1 ns each in the NVT and NPT ensembles. The temperatures were maintained using the Bussi thermostat [190] and pressure was maintained at 1 bar using the Parrinello-Rahman [299] barostat. The force fields used for each simulation together with their lengths and number of replicas are described in Table 4.4.

4.3.2 Temperature Replica-Exchange Simulations

Temperature Replica-Exchange Simulations (T-REMD) simulations were performed starting from NTD configurations obtained through implicit solvent equilibration. The distribution of temperatures were estimated using the predictor of Patriksson and Van-der-Spoel [303] 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 simulations were performed for 250 ns in each replica and exchanges between the replicas were attempted every 10 ps. The initial 50 ns were

discarded for equilibration and the coordinates from the lowest temperature (300 K) replica were considered for analysis.

4.3.3 PTMetaD-WTE Simulations

Parallel-Tempered Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) simulations were performed for two Lysine charge states of the basic-subregion configurations obtained from implicit solvent equilibration. The eight simulation temperatures were geometrically distributed as

$$T_i = T_{min} \left(\frac{T_{max}}{T_{min}} \right)^{(i-1)/(N-1)} \tag{4.1}$$

where T_{min} and T_{max} are 300 and 450 K respectively and N is the number of replicas – 8. The initial WTE metadynamics run used gaussians of height 2.5 kJ, width 500 kJ, bias factor of 50 and a deposition rate of 0.5 ps. The bias from this preliminary run was kept fixed in the subsequent production run to ensure sufficient replica exchanges. The production run used a 2-dimensional metadynamics potential along the alpha helicity (S_α, Equation 3.1) and radius-of-gyration (S_{rg}, Equation 3.2) CVs. The metadynamics biasing gaussians were deposited every 500 trajectory steps (1 ps) and exchanges between the replica were attempted every 5000 trajectory steps (10 ps). A bias-factor of 8 was used together with an initial hill height of 1.2 kJ. The gaussian widths σ were calculated as the half of the standard deviations of the CVs within a short 5 ns trial unbiased simulation of each system.

4.3.4 Biased Exchange Metadynamics

Biased-Exchange Metadynamics (BE-Metad) simulations were used to verify the independence of the results to the choice of force field/ion/water parameter set (Table 4.4) or the choice of collective variables. The H1.0 subtype NTD was used for these validation simulations and were initiated from the GBSA implicit solvent equilibrated conformation used for the T-REMD simulations. The BE-Metad simulations utilized four walkers with three biased and 1 unbiased replicas. The three metadynamic potentials biased the α-helical content (Equation 3.1), β-sheet content (Equation 3.1) and the Cα atom radius of gyration (Equation 3.2). The simulations were performed for 200 ns per replica to generate an accumulated sampling of 800 ns. The simulations used a bias-factor of 8 together with an initial hill height of 1.2 kJ. Biasing gaussians were added every

500 trajectory steps and exchanges between replicas were attempted every 10 ps.

4.3.5 DNA-Protein Metadynamics

The H1.0 subtype was used to examine the effects of DNA binding on the conformational preferences of the NTD. The crystallographic nucleosome structure of H1.0 in the on-dyad configuration (PDB ID: 5NL0 [231]) was obtained and Modeller [194] was used to add the NTD basic subregion to this structure in an extended configuration. The relative orientations of this built IDR and one of the linker DNA arms (20 bp) was used as the initial configuration (Figure 4.3). A 2-dimensional metadynamics potential was used that biased the number Lysine sidechain/Phosphate backbone interactions (S_{cont}, Equation 3.3) and the IDP's α-helicity (S_{α}, Equation 3.1). The biasing potentials were deposited every 10 ps and the simulation was run for 600 ns. To assist convergence and preclude the sampling of irrelevant states, the lowest terminal base pair (bp 20) and the backbone of the IDP's carboxy-terminal residue were both restrained using a force constant of 750 kJ mol^{-1} nm^{-2}.

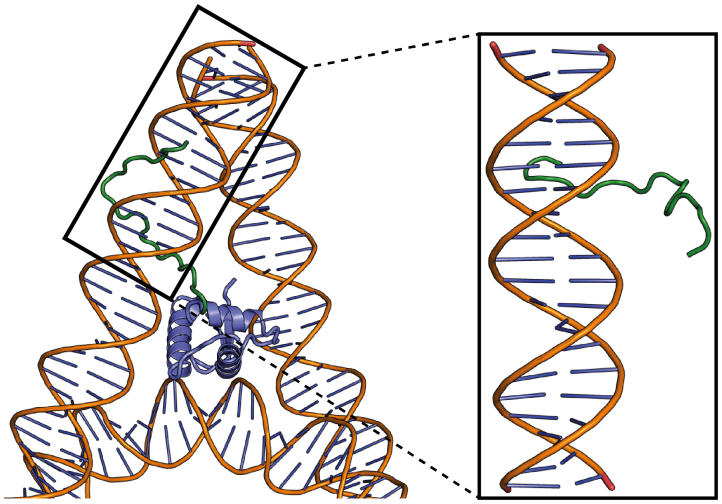


Figure 4.3 Constructing the initial configuration for metadynamics simulations. **Left:** The NTD basic subregion (green) was built in an extended random coil configuration onto the on-dyad H1 globular domain (PDB ID: 5NL0 [231]). **Right:** A 20-bp linker arm and the constructed NTD subregion were then together considered the initial configuration.

4.3.6 Docking and PMF Calculations

Docking was performed using the HADDOCK 2.2 webserver [229, 230] to determine the orientation of the conformations obtained from PTMetaD-WTE

simulations when bound to DNA. The B-DNA for docking was built as a 20-bp long strand using the Nucleic Acids Builder module of Amber16 [368]. To allow for all possible orientations of the IDP within the DNA, the central 11-bp (one-turn) were considered the 'active' residues for docking. The His residues were considered neutral for docking and no segment within the IDP was considered flexible.

The Potential of Mean Force (PMF) calculations were initiated from the docking configurations and used a protocol similar to that of Wieczor and Czub [371]. The helical axis of the DNA strand was aligned along the z-axis and the radial distance between the central DNA bp and the Helix Cα 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 CV using 20 windows of 0.1 nm (Figure 4.4). Each window was then simulated for 100 ns using an umbrella biasing potential of 750 kJ mol^{-1} nm^{-2} along the CV. To assist convergence, the diffusion of the DNA strands were reduced [372, 373] using positional restraints of 500 kJ mol^{-1} nm^{-2} on the two terminal bp at each end (bp 1,2,19,20). The initial 25 ns of the umbrella sampling trajectories were discarded and the rest of the trajectory was used for analysis. PMF profiles were generated using the Weighted Histogram analysis method as implemented in the *g_wham* [207] module of Gromacs [182].

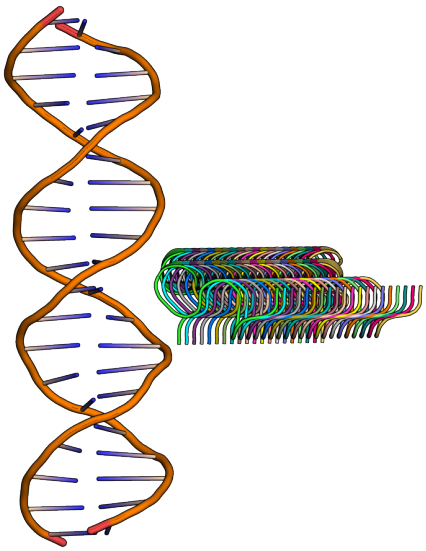


Figure 4.4 Illustration of the Umbrella Sampling windows for the DNA-H1 NTD basic subregion PMF calculations. The protein within each of the 20 windows is shown together with the DNA position.

4.3.7 List of Simulations

The simulations performed in this chapter are listed in Table 4.4 below.

Table 4.4 List of atomistic simulations performed to investigate the functioning of H1 amino-terminal domain.

System	Lysine State	Force Field	Time	Simulation
H1.0 Entire NTD	Charged	ff99sb*-ildn-q	56 × 250 ns	T-REMD
H1.1 Entire NTD	Charged	ff99sb*-ildn-q	96 × 250 ns	T-REMD
H1.2 Entire NTD	Charged	ff99sb*-ildn-q	72 × 250 ns	T-REMD
H1.0 Entire NTD	Charged	Charmm36M	56 × 250 ns	T-REMD
H1.1 Entire NTD	Charged	Charmm36M	96 × 250 ns	T-REMD
H1.2 Entire NTD	Charged	Charmm36M	72 × 250 ns	T-REMD
H1.0 Basic Region	Charged	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.1 Basic Region	Charged	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.2 Basic Region	Charged	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.0 Basic Region	Neutral	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.1 Basic Region	Neutral	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.2 Basic Region	Neutral	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.0 Entire NTD	Neutral	ff99sb*-ildn-q	8 × 400 ns	PTMetaD-WTE
H1.0 Entire NTD	Neutral	ff03ws	4 × 200 ns	BE-Metad
H1.0 Entire NTD	Neutral	ff99sb*-ildn-q	4 × 200 ns	BE-Metad
H1.0 Basic Region + ds-DNA	Charged	Charmm36M/Charmm36 DNA	600 ns	Metadynamics
Docked H1.0 + ds-DNA	Charged	Charmm36M/Charmm36 DNA	400 ns	Unbiased MD
Docked H1.0 + ds-DNA	Charged	ff99sb*-ildn-q/parmbsc0	400 ns	Unbiased MD
Docked H1.1 + ds-DNA	Charged	Charmm36M/Charmm36 DNA	400 ns	Unbiased MD
Docked H1.1 + ds-DNA	Charged	ff99sb*-ildn-q/parmbsc0	400 ns	Unbiased MD
Docked H1.2 + ds-DNA	Charged	Charmm36M/Charmm36 DNA	400 ns	Unbiased MD
Docked H1.2 + ds-DNA	Charged	ff99sb*-ildn-q/parmbsc0	400 ns	Unbiased MD
Docked H1.0 + ds-DNA	Charged	Charmm36M/Charmm36 DNA	20 × 100 ns	Umbrella Sampling
Docked H1.0 + ds-DNA	Charged	ff14SB/parmbsc1/cufix	20 × 100 ns	Umbrella Sampling
Docked H1.2 + ds-DNA	Charged	Charmm36M/Charmm36 DNA	20 × 100 ns	Umbrella Sampling

4.4 Results

4.4.1 Amino Terminal Domains in Isolation

The structural ensemble of the *h*H1.0 NTD when in solution is predominantly disordered with limited secondary structural content (Figure 3.11B). The NTDs of the *h*H1.1 and *h*H1.2 vary significantly from *h*H1.0 in sequence, length and charge concentration (Table 4.3). T-REMD simulations were first used to analyze if these sequence differences are sufficient to induce different structural behaviours, such as, flickering secondary structural elements [15] or stable secondary structural folds [374]. Temperature replicas ranging from 300 to 450 K were considered and secondary structures were assigned using the DSSP algorithm [321] implemented within the MDTraj [375] Python library.

For all three subtypes, the H1 NTDs when isolated in solution, emerge as a highly unstructured domain displaying only transient secondary structural elements. Irrespective of the force field used (Amberff99sb*-ildn-q [166, 361, 362]/Charmm36M [168]) the conformational ensembles include only a small fraction (<15%) of residues with α-helical or β-strand structural elements.

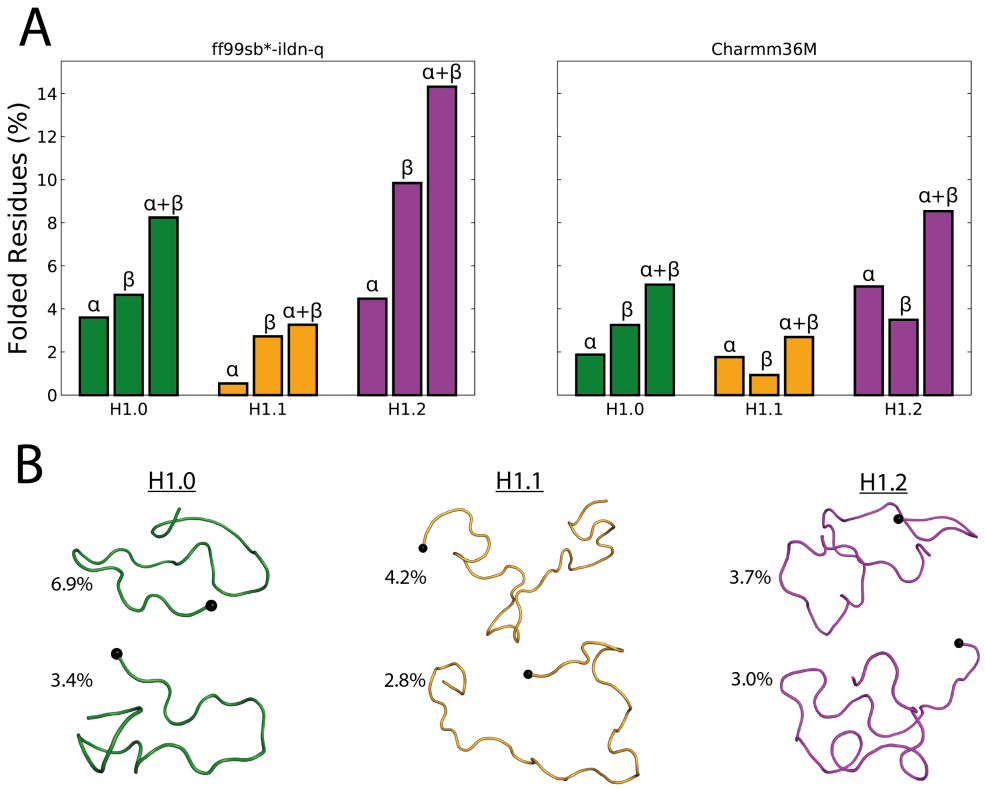


Figure 4.5 (A) Ensemble averages of the fraction of secondary structures in each of the three NTDs. **(B)** Illustrations of the highest populated clusters of the three NTDs. The final residue of the NTDs (point to attachment to the globular domain) is shown as a back sphere.

To validate the predominance of random coil conformations within the structural ensemble, the Cα atoms of the NTDs were clustered using the single-linkage method [304] and a distance cut-off of 2 Å. The clusters illustrate the structural heterogeneity resulting from the predominance of random-coil configurations with the highest populated cluster encompassing merely $\sim$7% of the trajectory frames (Figure 4.5B). To further explore the functional implications of this structural heterogeneity of the NTDs, their size distributions were calculated and compared to the mean radius of gyration (R_g) of globular (Equation 3.6) and random-coil proteins (Equation 3.7). Despite the structural heterogeneity evidenced by the wide distributions of R_g, most of the conformations adopted by the NTD are compact with R_g values closer to that predicted for a globular protein of similar length (Figure 4.6).

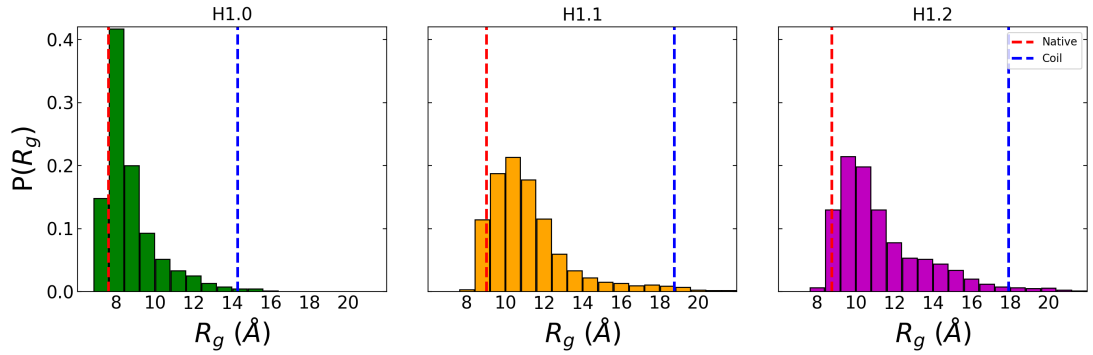


Figure 4.6 Distributions of the radius of gyration (R_g) of the NTD Cα atoms from T-REMD simulations using the ff99sb*-ildn-q force field. The theoretically predicted regressions for the random-coil and globular protein are shown in blue and red respectively.

The consistent behaviour of all the NTD subtypes is in-contrast to that observed from REST2 simulations of the H1.0 CTD (Figure 3.11C) where the IDR preferentially adopts extended conformations. This difference in compaction between the two unstructured domains might plausibly be attributable to the uneven distribution of charge density within them. The 97-residue H1.0 CTD is uniformly positively charged, while the NTD is split into 'hydrophobic' and 'basic' subregions (Figure 4.2). To test this hypothesis, the R_g of the NTDs' two subregions were recalculated and compared to theoretical values (Figure 4.7). The basic subregion with higher charge density adopts extended conformations and the hydrophobic subregion predominantly adopts collapsed conformations that contribute to the comparatively compact overall NTD R_g observed in Figure 4.6. The results offer a probable molecular role for the unique distribution of charged/uncharged residues within the NTD. Extended states of the CTD (Figure 3.11C) might allow it to initiate the process of nucleosome recognition through long-range electrostatics. In-contrast, the NTD's hydrophobic subdomain directs the IDR to adopt a collapsed conformation that could preclude the basic charges within it from doing so [251, 253].

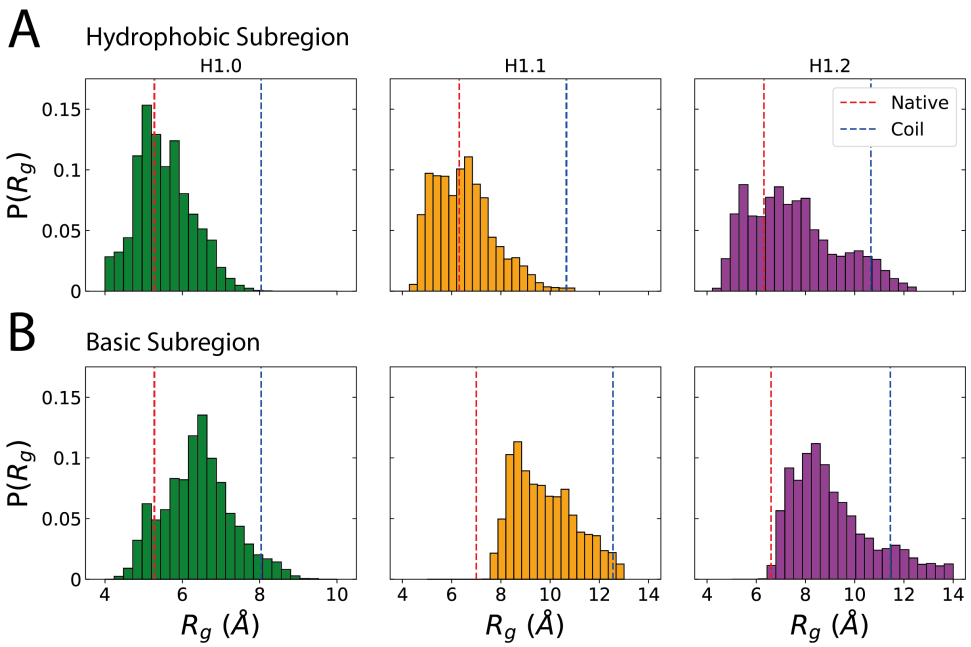


Figure 4.7 Distributions of the radius of gyration (R_g) of the H1 NTD hydrophobic (Top) and basic (Bottom) subregions. The theoretically predicted regressions for the random-coil and globular protein are shown in blue and red respectively.

4.4.2 Amphiphilic Helicity of the H1 NTD

The disordered nature of the H1 NTDs in solution is in stark contrast to the behaviour of the H1.0 NTD observed when bound to the nucleosome (Section 3.4.2). In the simulations discussed in Chapter 3, secondary structural biasing potentials within the BE-Metad simulations of the 211-bp nucleosome were only applied to the CTD. However, despite this, the basic subregion of the NTD condensed into a structure with significant helical propensity. PTMetaD-WTE simulations were thus used to study this structural transformation of basic subregions upon their binding to DNA. However, characterizing an IDR's folding pathway upon DNA binding is computationally inexpedient and can necessitate upto 50 μs of sampling [376]. The binding of the H1 NTD basic subregions to DNA, transforms its structural behaviour due to the screening of the electrostatic repulsion among its numerous positively charged residues enabled by the interactions with the phosphate backbone [320, 377, 378]. The PTMetaD-WTE simulations of the three subtypes' basic subregions were thus performed in two conditions – 1) with standard charged Lys (WT) and 2) with neutralized Lys sidechains as an approximation to the NTD's neutralization by DNA.

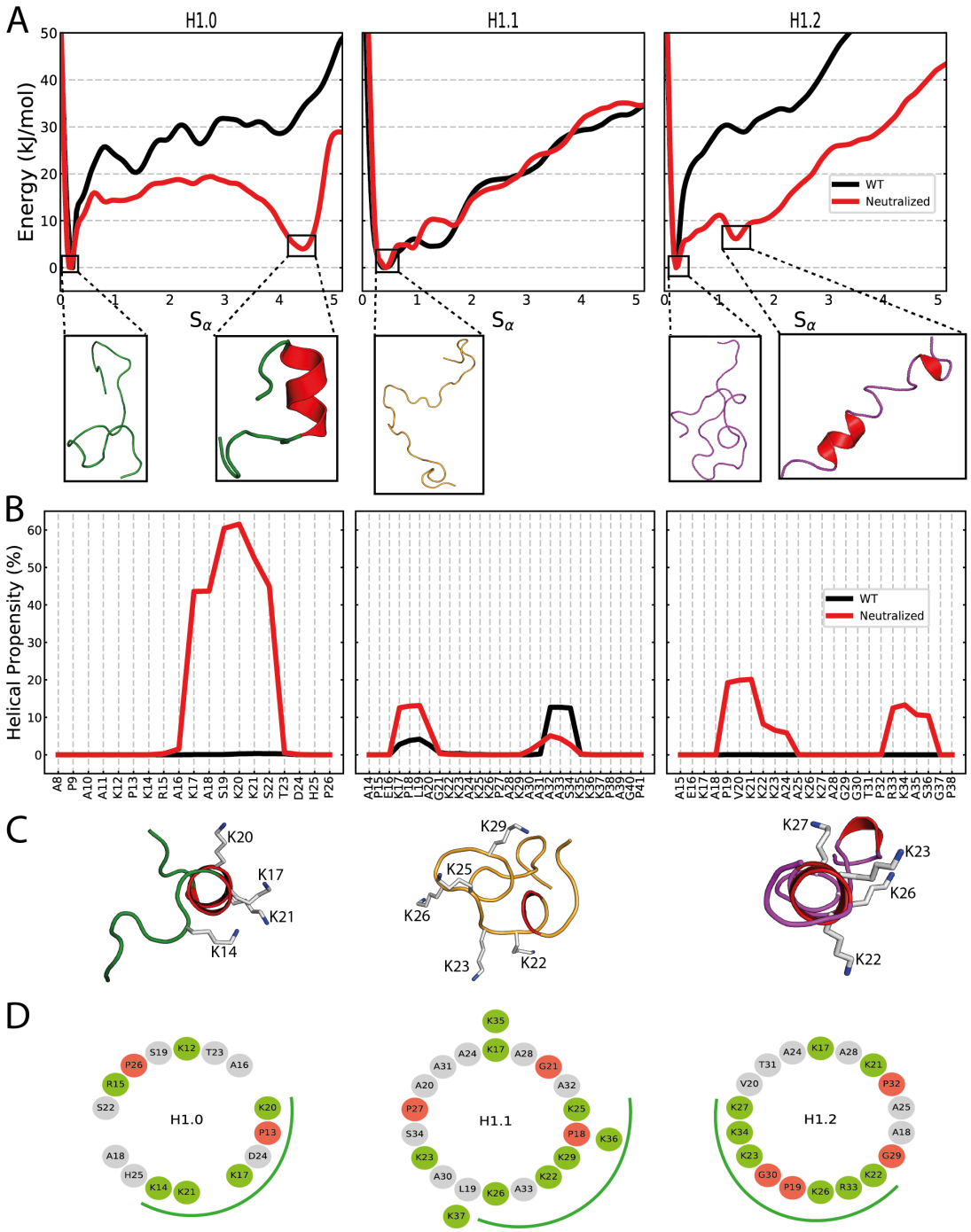


Figure 4.8 Inducible amphipathic helicities of the H1 NTDs. **(A)** The one-dimensional free energies along the Sα CV for the WT NTDs (black) and with neutralized Lys sidechains (red). Representative conformations illustrate the free energy minima with the α-helical motifs in red. **(B)** The reweighted per-residue helical content from PTMetaD-WE simulations of the WT NTDs (black) and with neutralized Lys sidechains (red). **(C)** Top-down illustrations of the induced NTD conformations from simulations with the neutralized Lys sidechains. The α-helical motifs are shown in red and specific residues within the positive 'face' are illustrated. **(D)** Helical wheel projections of the basic subregions of the three H1 NTDs. Basic Arg/Lys residues are in green and the helix-breaking Pro/Gly residues are in red. The 'basic' face of each subtype is illustrated with a green curve.

For the WT NTD, independent of the subtype, the free-energy profile along the Sα CV shows a single minimum at a low helical value of $\sim$ 0.1 (Figure 4.8A). This profile is in agreement with the predominantly unstructured ensembles observed in the T-REMD simulations of the N-terminal domains (Figure 4.5A). Upon charge neutralization, all three variants preserve the global minimum at Sα values of $\sim$0.1, signalling favourable unstructured conformations in all cases. However, interestingly, the neutralized NTDs of the H1.0 and H1.2 subtypes exhibited a second low energy minima at a higher value of Sα, corresponding to conformations with helical secondary structural content. To validate the secondary minima and to disregard the existence of other unaccounted minima, the complete two-dimensional free energy surface along the two metadynamics CVs were plotted and compared between the two Lysine charge states (Figure 4.9).

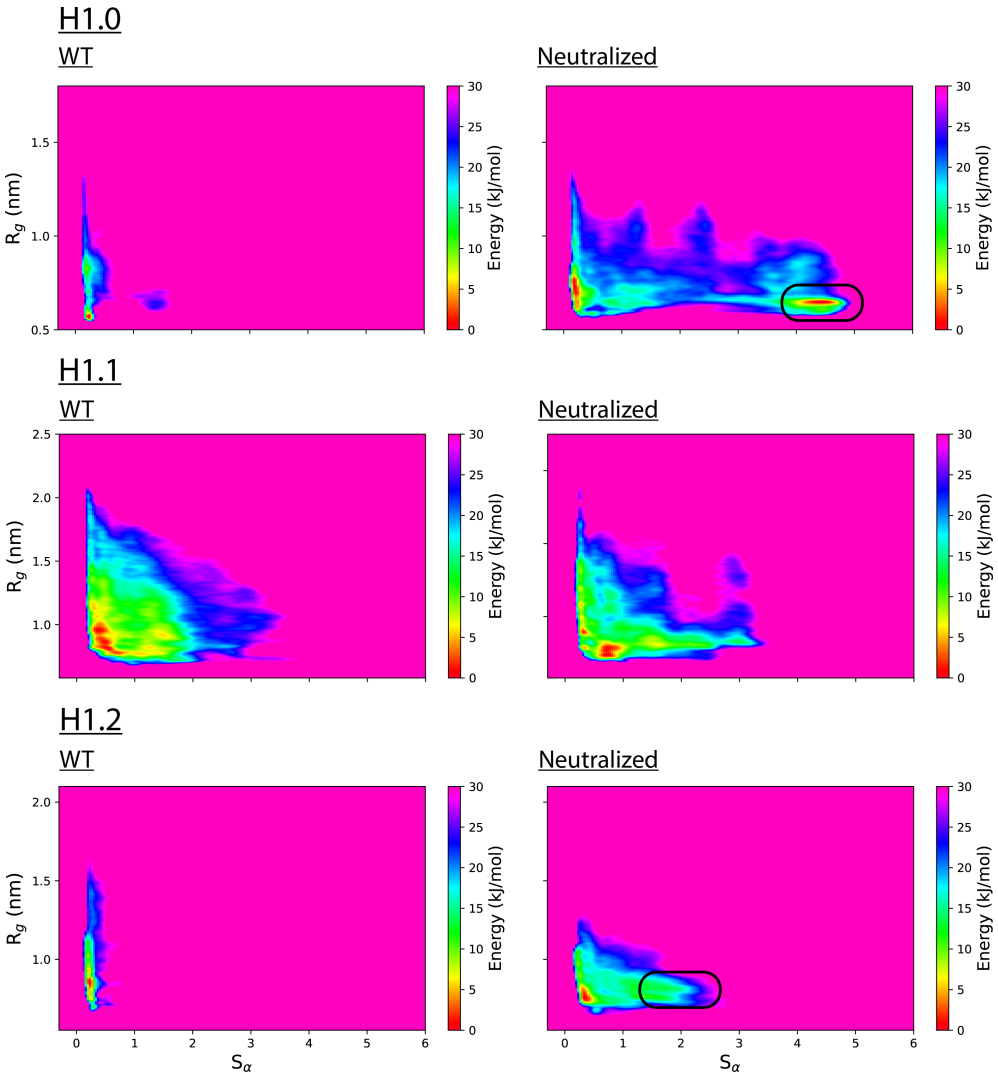


Figure 4.9 Free Energy surfaces calculated from the 300 K replica of the NTD basic subregions along two CVs S_α and S_{rg}. The energy basins 'induced' upon Lysine sidechain charge neutralization are highlighted with a box.

The minimum of the H1.0 NTD at large Sα values (Sα=4.4) is nearly as stable as that of the unstructured conformation (ΔE = 3.6 kJ/mol). The helical propensity of the neutralized H1.0 NTD estimated using the WHAM reweighted probability distributions shows that this is a single seven residue-long helix spanning the region $_{16}$AKASKKS$_{22}$ (Figure 4.8B). Further inspection of this motif shows that its formation would concentrate the positively charged residues K14, K17, K20 and K21 on the same 'face' of the helix to create a charged face that should favour interactions with DNA. In fact, helical wheel projections down the axis of an ideal helix (3.6 residues per turn, 100° between residues) reveals that the sequence of the H1.0 NTD is ideal for clustering charges along one face (Figure 4.8D) – a common motif within many proteins that undergo a disordered-to-ordered transition upon binding to oppositely charged counterparts [374, 379, 380].

Although the H1.2 NTD with uncharged Lysine sidechains also exhibits a second minimum corresponding to a partial helical conformation, this occurs at a much lower Sα value of 1.3 and with a higher free energy in-comparison to the unstructured conformation (ΔE = 5.8 kJ/mol). In line with the lower Sα 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 4.8B). An inspection of the H1.2 NTD sequence explains the IDR's incapability to form a single long helix: each of the two short helices start with a 'helix-initiating' Proline residue and are separated by a helix-breaking double Gly (G29/G30) motif [381]. In fact, Vila et al. [318] observed a similar helix-Gly-Gly-helix motif within the *m*H1.4 NTD (>95% sequence similarity to *h*H1.2) when within the helix-stabilizing solvent TriFlouro-Ethanol (TFE). Despite helicity in the H1.2 NTD case being shorter-lived than in the H1.0 subtype, the transient partial folding would still enable a greater concentration of positive charges along a 'face' for DNA interactions than when in a random coil conformation.

In contrast to the two other H1 NTDs, the H1.1 subtype exhibits a one-dimensional free-energy surface along Sα that is invariant with the Lysine charge states. The per-residue helicity plot (Figure 4.8B) confirms this with only two small uncharged segments with a propensity of <10%.

The differences in the observed folding propensities for the three subtypes correlate with the fraction of positively charged residues and the lack of helix-breaking residues each system possesses (Table 4.3). The H1.0 NTD with the highest folding propensity has the highest concentration of positive residues and only one 'helix-destabilizing' Pro/Gly residue [382]. Vice-versa, the H1.1 NTD subtype that exhibits the weakest folding propensity has the lowest pos-

itive charge concentration and is interspersed with four helix-destabilizing residues [382]. The helical-wheel analysis of Figure 4.8D further illustrates this trend with the H1.0 and H1.2 subtypes that display helical propensity also exhibiting significant built-in amphiphilicity in contrast to H1.1. Crucially, these trends in folding propensity and charge concentration also significantly correlate with the experimentally observed binding affinities of the subtypes (Table 4.2).

4.4.3 Disorder-to-Order transition

The PTMetaD-WTE simulations predicts a subtype-dependant propensity for the amino-terminal domains to condense into conformations with helical secondary structural content. The conformational ensemble of disordered proteins are however extremely sensitive to the choice of force field, water model and ion parameters [383, 169]. BE-Metad simulations of the H1.0 NTD were thus performed to validate the helical condensation of the IDR observed with neutralized Lysine charge states. The simulations used two distinct combinations of force field, ion and water parameters – 1) ff03WS protein [167], TIP4P-2005 water [177], Aqvist & Dang ions [363, 364], 2) ff99sb*-ildn-q [166, 361, 362], TIP3P water, Joung & Cheatham ions [384].

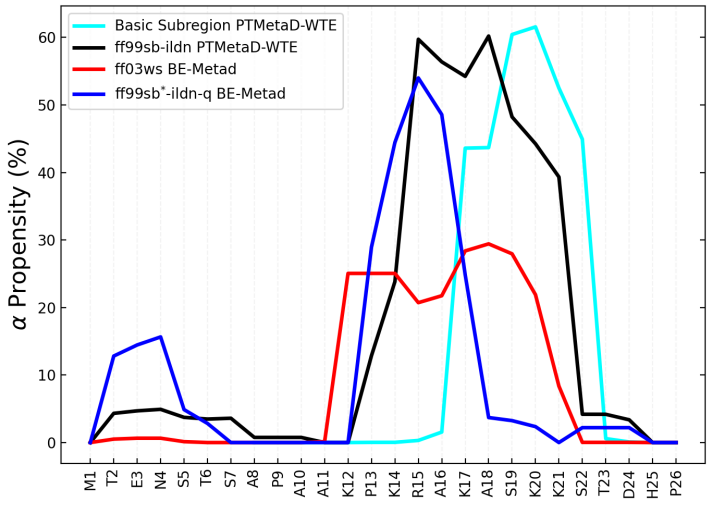


Figure 4.10 The per-residue helical propensity of the H1.0 NTD from BE-Metad (with neutral Lys) of the whole NTD using varying force field parameters. The per-residue helicity from the PTMetaD-WTE simulations of the whole length NTD (black) and basic-subregion (cyan) are plotted for comparison.

Regardless of the choice of parameter set, the H1.0 NTD displays significant helical secondary structural content (Figure 4.10). However, the span and

extent of this helicity varied with the system and force field. The inclusion of the entire NTD in-contrast to only the basic subregion within the PTMetaD-WTE simulations, slightly shifted the span of the helix away from the point of attachment to the globular domain. Most crucially however, the four K14, K17, K20 and K21 residues that form the positive 'face' (Figure 4.8C) are helical within all force field combinations.

The metadynamics simulations thus far predict the helical propensities of the NTD when the reduced electrostatic repulsion between their Lysine residues are modelled using neutral sidechains. To validate the stability of this disorder-to-order transition when the reduction in repulsion is instead caused by binding to the phosphate backbone, additional metadynamics simulations were performed of the H1.0 NTD (with standard charged Lys) in contact with a 20-bp ds-DNA strand (Figure 4.3). The simulations predict that upon DNA binding, the H1.0 NTD transitions from the unstructured configuration (S_α = 0, Figure 4.11) to a helical state. The global minimum of the system occurs at S_α ~5.8 and S_{cont} ~12.5 which corresponds to the helical peptide bound to DNA (Figure 4.11). The resulting helix spans a slightly larger region ($_{13}$PKRAKASKKST$_{23}$) and most importantly includes the four-residue K14, K17, K20 and K21 positive 'face' that governs interactions with DNA.

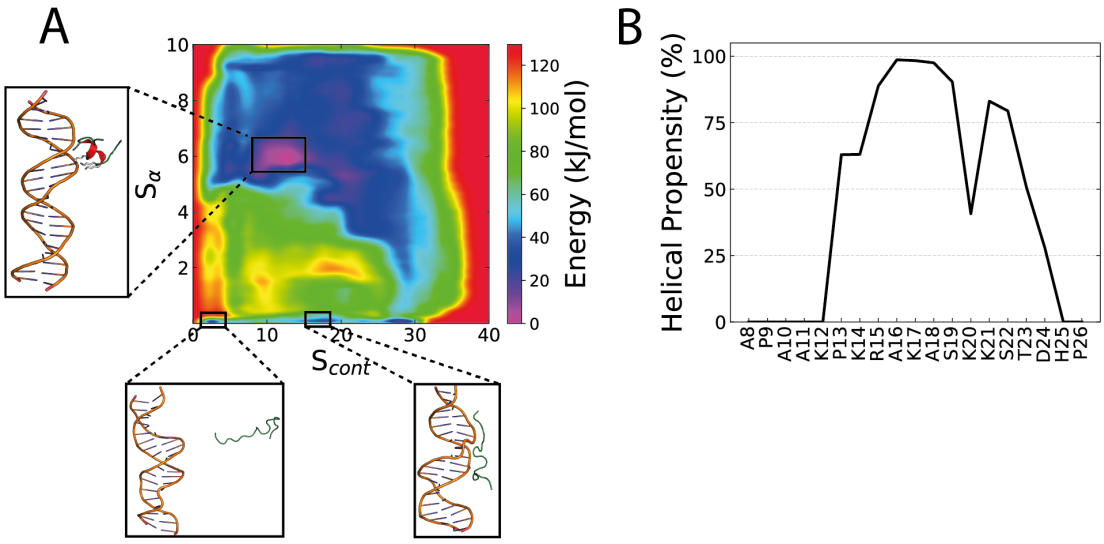


Figure 4.11 DNA induced secondary structure within the basic subregion of the H1.0 NTD **(A)** Two-dimensional free energy surface of the NTD when interacting with DNA along two CVs – S_α and S_{cont}. Three representative configurations within the free energy surface are illustrated. **(B)** The reweighted per-residue helical content of the H1.0 NTD basic subregion when interacting with DNA.

4.4.4 Subtype specific DNA Affinity

The combination of metadynamics simulations predicted a propensity for the H1 NTD subtypes to condense into conformations with helical secondary structural content. Crucially, this propensity correlates with the experimentally observed nucleosomal affinities of the three H1.0, H1.1 and H1.2 linker histone subtypes. Additional MD simulations were thus performed to quantify if these variant-specific helical propensities indeed translate to variable DNA-binding propensities. The dynamics of peptides bound to DNA are however particularly slow due to the significant number of simultaneous electrostatic interactions – a trend accentuated if the preferred binding modes of the protein are within the DNA grooves [376]. HADDOCK [229, 230] was thus used to rigid body dock NTD configurations representative of either the secondary minima (for H1.0 and H1.2) or the first minimum (for H1.1) from the PTMetaD-WTE simulations to a 20-bp ds-DNA segment.

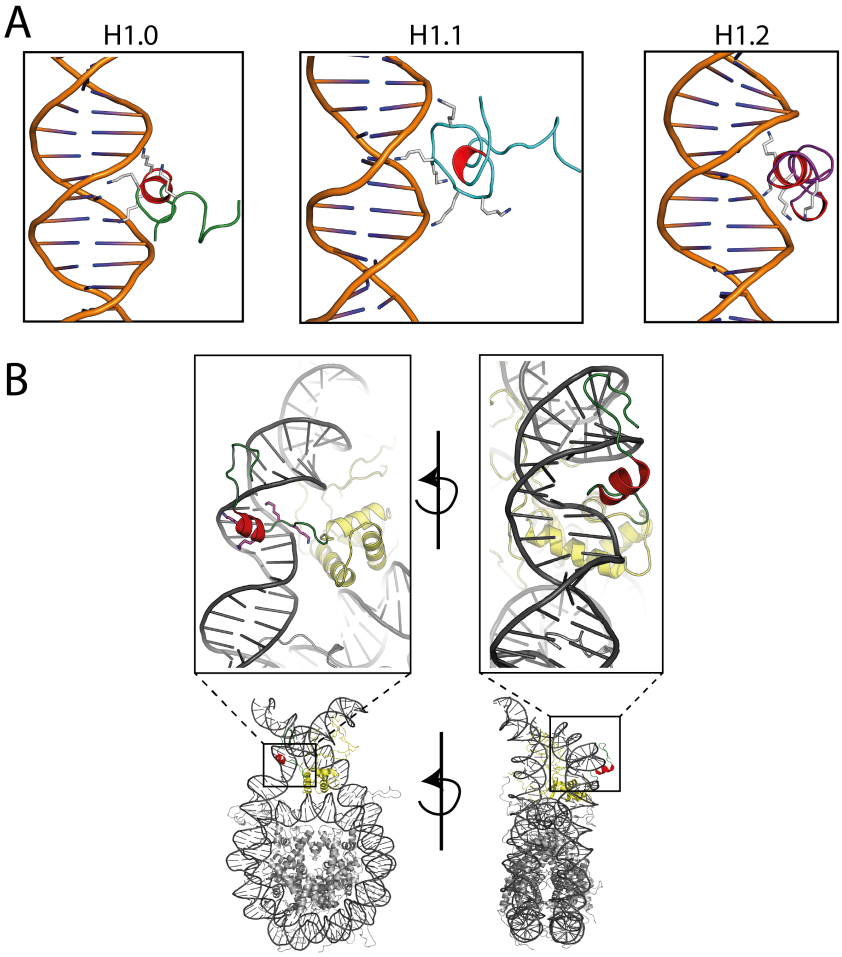


Figure 4.12 Orientations of H1 NTD conformations within the DNA grooves. **(A)** Lowest energy docked conformations of the induced helical conformationswith a 20-bp ds-DNA strand. **(B)** Illustrations of the most populated cluster of the NTD from BE-Metad simulations of the full-length H1.0 within a 211-bp nucleosome.

Figure 4.12A shows the lowest energy clusters obtained from HADDOCK docking of the three subtypes. For the H1.0 and H1.2 subtypes, docking predicts the helix region to fit snugly within the DNA major groove to facilitate interactions between the four Lys residues within the basic face and DNA. Such a binding of helices is amongst the most common DNA-protein interaction motifs [385] with such interfaces present within zinc-finger [386] and leucine zipper proteins [387]. Crucially, this docking prediction is consistent with the positioning of the H1.0 basic subregion within the DNA major groove observed from simulations of the full-length H1 within a 211-bp nucleosome (Chapter 3, Figure 4.12B). In contrast to this, the interactions within the H1.1 docked NTD configuration are non-specific with only three basic residues (all unstructured) in contact around the DNA minor groove.

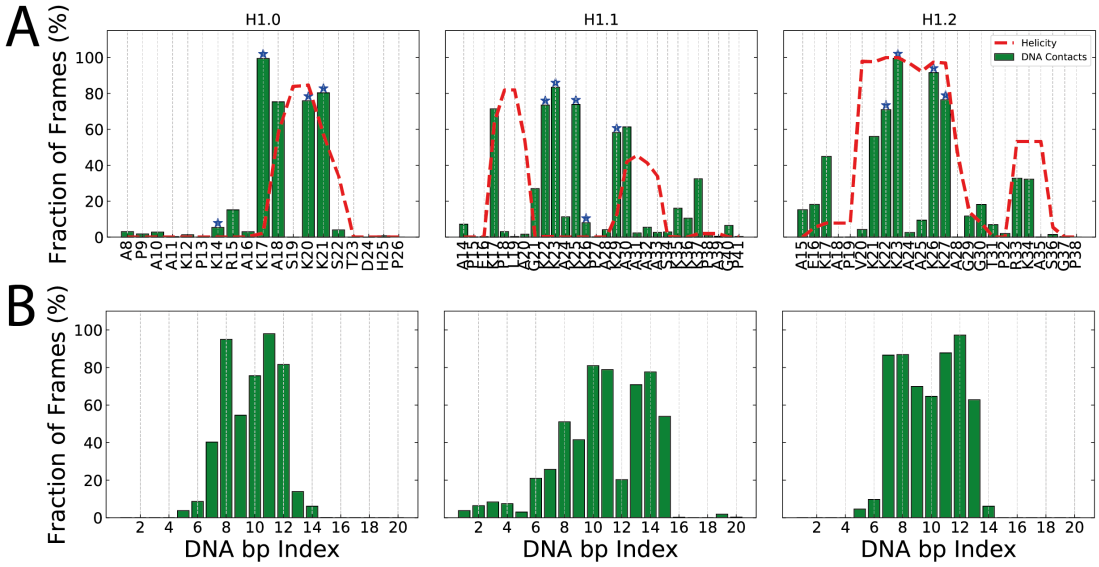


Figure 4.13 (A) Per-residue interactions of the H1 NTDs with DNA from unbiased MD simulations initiated from the docked conformations. A contact was assumed if a non-hydrogen atom of the residue was within 3.2 Å of a non-hydrogen DNA atom. The average per-residue helicity is shown in red. The Lys residues within the basic 'face' of each NTD are marked with a '*'. **(B)** Interactions per base pair of the DNA strand with the three LH subtypes within the unbiased MD simulation.

To assess the viability of the docked conformations and to investigate the implications of NTD differential folding across subtypes on DNA interactions, unbiased MD simulations were initiated from the HADDOCK predicted conformations. When interacting with the DNA strand, the helical configuration of the H1.0 NTD is stable across the 400 ns MD simulation with residues within the basic face (K17, K20, K21) accounting for the majority of interactions with the backbone (Figure 4.13A). The DNA-bound H1.2 NTD also retains its helical conformation with the basic residues within the $_{19}$PVKKKAAKKA$_{28}$ segment

directing interactions with DNA. Here, the four K22, K23, K26 and K27 residues fall on the same side of the helix to make two contiguous helical turns that interact with the major groove. The secondary $_{33}$RKAS$_{36}$ with two basic residues however is short and interacts only transiently with DNA. Contrary to the two helical H1.0/H1.2 subtypes, within the H1.1 NTD, DNA interactions are primarily mediated by the Lys-rich unstructured $_{21}$GKKAKKPA$_{28}$ loop. This governing of interactions by a flexible segment results in transient binding with a reduced number of DNA-bound states ($\sim$83% of simulation frames) than in the H1.0 and H1.2 cases (100%), where rather than focussing on the major groove, the NTD promiscuously binds/unbinds the entire DNA strand (Figure 4.13B).

To quantify the impact of this observed subtype-specific interaction patterns and helical propensities on their DNA binding affinities, umbrella sampling simulations were used to independently model NTD binding to a 20-bp ds-DNA strand. From the umbrella sampling simulations, the potential of mean force (PMF) was calculated along the radial protein-DNA distance orthogonal to the DNA helical axis (Figure 4.4). The PMFs reveal that while the subtype with highest helical propensity, the H1.0 NTD, is stabilized by $\sim$20 kJ/mol upon DNA binding, the longer H1.2 NTD with lower helical propensity and positive charge concentration is stabilized by only $\sim$15 kJ/mol. However, within molecular mechanics force fields, the interactions between oppositely charged amine and phosphate groups have not been explicitly calibrated [388]. The PMF calculations were thus repeated with the Amberff14SB [389] and parmbsc1 [390] force fields together with the non-bonded (cufix) corrections to guanidinium-phosphate [391] and amine-phosphate [388] interactions. While the inclusion of the non-bonded interaction corrections modified the free-energy profile at short distances, the depth of the resulting well remained unchanged at $\sim$20 kJ/mol.

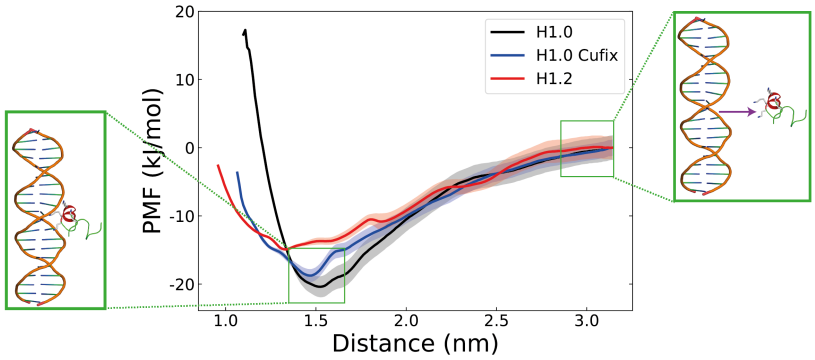


Figure 4.14 PMF profiles for the binding of H1.0 (black/blue) and H1.2 (red) with the DNA major groove. The standard deviations of the calculated PMFs estimated from bootstrapping are shown as shaded regions. The energy profiles are shifted so that the energy at large distances is set to zero.

The results suggest that the disordered-to-helical transition of the H1 NTD allows the concentration of positive charge to promote interactions with the DNA phosphate backbone. Remarkably, these trends in helical propensities and thereby charge concentration vary congruously with the experimentally observed binding affinities of the subtypes (Table 4.2). The results thus allow the formulation of a hypothesis regarding the structural mechanism of the NTD's role in contributing to the differential nucleosomal binding of the subtypes. The higher nucleosomal affinity of H1.0 might be in-part due to the NTD's propensity to form a stable amphipathic helix that 'glues' the LH to DNA [254]. Conversely, the longer and less-positive H1.1 NTD is disordered and thus its weaker fluctuating interactions inadequately anchor the LH to nucleosomal DNA.

4.5 Conclusion and Further Work

In this chapter, a combination of microsecond-long T-REMD, PTMetaD-WTE and BE-Metad simulations are used to sample the conformational landscape of the NTD of H1 subtypes H1.0, H1.1 and H1.2. When free in solution, the simulations predict all three NTDs to adopt primarily unstructured but partially collapsed conformations. However, the reduction of inter-Lysine sidechain repulsion within the NTDs either artificially by increasing pH or physiologically by binding to DNA, leads the domain to undergo a disorder-to-order transition. The ordered NTD structures in the H1.0 and H1.2 subtypes are amphipathic helices that orient the basic residues along one 'face' ideally suited for DNA interactions. Interestingly, the proportionality of this helical propensity to the subtypes' nucleosomal affinity offers a structural hypothesis for the domain's

role – 'anchoring' the LH to the nucleosomal DNA. Crucially, this computationally predicted structural mechanism correlates with the experimentally observed effects of NTD domain swap mutants where H1.0/H1.2 exchanges resulted in an exchange of binding characteristics [255].

The results of the atomistic simulations within this chapter while offering an interesting testable hypothesis of LH functioning, also provide an avenue for investigating the impact of other charge-reducing post-translational modifications like Lysine acetylation [15, 156] or Arginine citrullination [76] on IDP structure and function.

Chapter 5

Heterochromatin Protein HP1 within chromatin

Contents

5.1 Introduction

The condensation of DNA from ∼2 m in stretched form to fit within the nucleus is only one half of chromatin's role. The three-dimensional spatial organization of DNA and DNA-associated proteins serves a second and possibly more important role in allowing the capability to exert complex levels of control over gene expression [44–46]. The chromatin structure thus adds an additional layer of complexity above the genome code and thereby permits the transcription of different proteins depending on cell lineages/cycles [47–50]. This intimate link between genome structure and gene regulation is suggested to exist across all phases of the hierarchical chromatin structure [53, 392].

At the local level, interactions of DNA with histones within the nucleosome render them inaccessible to proteins that read sequence. Local nucleosomal rearrangements (Figure 5.1A) are thus necessary for the exposure of the binding sites and the initiation of transcription [393, 394]. At the intermediate scale, gene promoters and enhancers are often linearly separated by thousands of base pairs and chromatin structural 'looping' is sometimes necessary to bring them in physical proximity to interact [395–397] (Figure 5.1B). At the megabase scale, chromosomes have recently been identified to spatially segregate into 'TAD's – Topologically Associated Domains that are capable of both excluding and enriching the concentration of regulatory proteins within the domains [398–400] (Figure 5.1C).

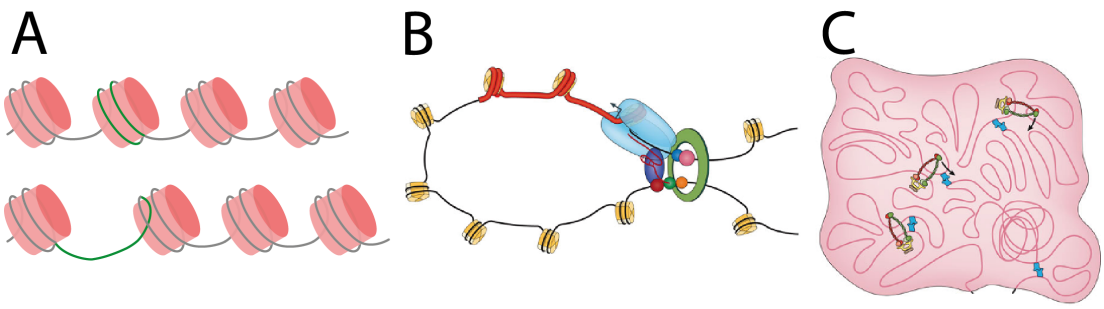


Figure 5.1 Intrinsic role of chromatin structure in gene expression. **(A)** Local nucleosomal rearrangements are necessary to permit exposure of DNA (green) to transcription factors. **(B)** Chromatin looping is necessary to allow physical proximity of the gene (red), promoter (blue) and enhancer (purple). **(C)** A combination of chromatin looping to form Topologically Associated Domains (TAD) capable of including/excluding other proteins. Image from Bonev et al. [53].

The intrinsic relationship between genome structure and gene expression is exemplified by the two structurally and functionally distinguishable regions of chromatin – euchromatin and heterochromatin [401, 402]. Euchromatin regions

of the genome are less-condensed, gene-rich and more easily transcribed in-contrast to heterochromatin regions that are compact, gene-poor and transcriptionally repressed [401, 403]. Figure 5.2 illustrates this difference between the two chromatin regions where the tightly packed heterochromatin stains more intensely than the loosely packed euchromatin [404, 405]. Post-translational modifications of histones lie at the core of distinguishing between the two chromatin states with euchromatic regions particularly enriched in acetylated H3/H4 tails [15, 401, 406] while hypo-acetylation and H3K9 methylation are particular indicators of heterochromatic regions [399, 402, 407].

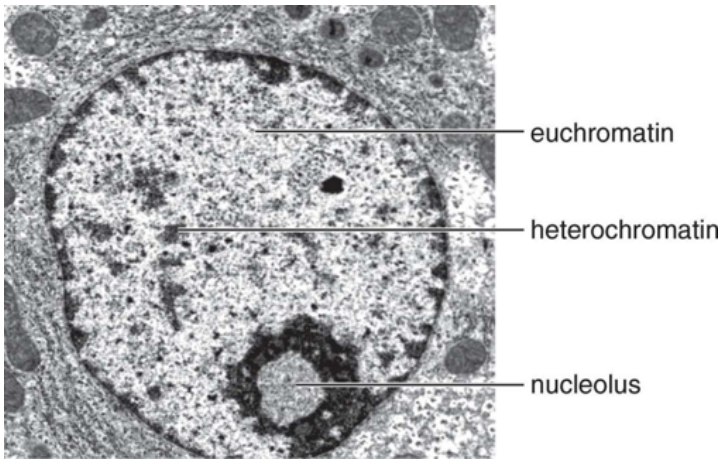


Figure 5.2 Differential compaction of eu- and heterochromatin regions of the nucleus revealed by their lighter and darker staining respectively. Image from Fazary et al. [404].

5.1.1 Heterochromatin Protein 1

The structural mechanisms of the correlation between histone post-translational modifications and the chromatin states remains a major open question in the field of genomic biology [402]. However, the Heterochromatin Protein 1 (HP1) is a non-histone architectural protein that has been identified to lie at the core of the formation of H3K9-mediated heterochromatic regions [408]. The importance of HP1 to gene regulation is illustrated by the presence of its homologs across the eukaryotic spectrum including Drosophila [409], yeast [410] and up to mammals [411]. The homologs across the organisms share a conserved penta-partite structure composed of a combination of structured and disordered domains. Figure 5.3 illustrates this structure of HP1 together with the positioning of the five domains – N-terminal domain (NTD), chromodomain (CD), hinge domain, chromoshadow domain (CSD) and the C-terminal domain (CTD). In vivo however, the protein has been identified to primarily exist as a homodimer caused by the association of the chromoshadow domains [412] (Figure 5.3D).

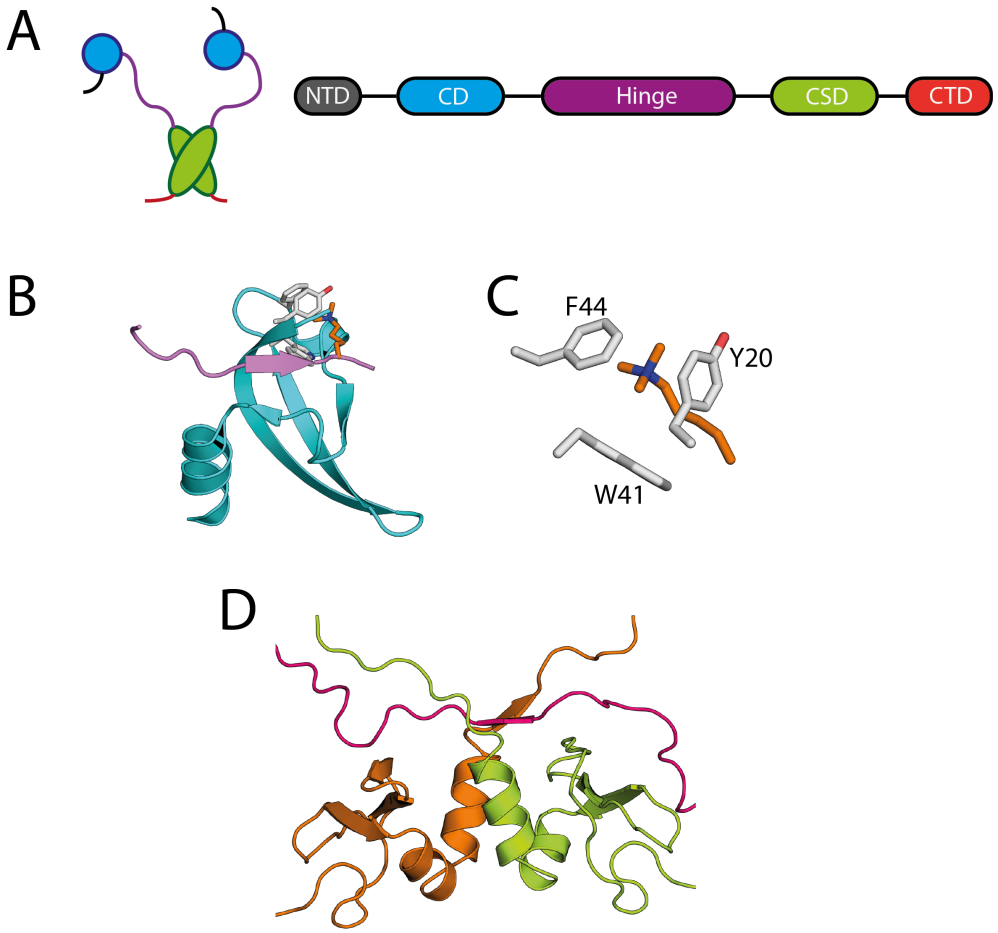


Figure 5.3 Penta-partite structure of Heterochromatin Protein 1. **(A)** Illustration of the five NTD, CD, Hinge, CSD and CTD domains together with its CSD dimerization interface. **(B)** H3 tail bound configuration for the chromodomain illustrated using the *h*HP1α structure 3FDT [413]. The CD is in blue and the H3 tail in purple. **(C)** The formation of the hydrophobic cage within the CD to 'capture' the methylated Lysine illustrated using the *h*HP1α structure 3FDT. **(D)** The twin dimerization and H3-binding functional roles of the CSD illustrated using the *m*HP1α structure 1SZ4 [414]. CSD from two different HP1 molecules are coloured orange and green. The H3 tail is in red.

5.1.1.1 Chromo-Domain

The HP1 chromodomain adopts an OB-fold motif – a compact structured motif of $\sim$50 residues found within eukaryotic nucleic acid binding proteins [415] composed of a three-stranded β-sheet pressed against an α-helix [416]. However, structural evidence for the nucleic acid interactions of the HP1 chromodomain are lacking [417] in-comparison to its primary role of recognizing and binding the K9 methylated H3 tail [408, 418]. This recognition of the methylation mark occurs through a network of aromatic residues within the CD that together form a hydrophobic cage that binds through cation-π interactions [144]. The cage-motif in-turn allows the normally disordered [15] ${}_{6}\text{TARK}_{\text{met}}\text{S}_{10}$ segment of

the H3 tail to adopt an antiparallel β-strand configuration within the two CD β segments (Figure 5.3B).

5.1.1.2 Chromoshadow Domain

The chromoshadow-domain is the second structured domain within HP1 and as its name suggests, structurally 'shadows' the chromodomain albeit with an additional short α-helix [419]. Functionally, the domain is implicated in two important facets of HP1 – dimerization and nucleosome binding. Figure 5.3D illustrates the most common mode of dimerization where two CSD monomers interact through hydrophobic residues within their α-helices [412]. This dimerization interface offers a secondary H3 binding site for the recognition of the tail's hydrophobic PxVxL motif [408].

5.1.1.3 N-terminal, Hinge and C-terminal Domains

The N-terminal, hinge and C-terminal domains are all unstructured and have amino acid compositions archetypical of disordered proteins – rich in charged residues while being devoid of large hydrophobic ones [420]. This intrinsic disorder within the three domains that together make-up nearly 60% of the protein has hindered the deciphering of the protein's overall structural conformation [418, 421].

The positioning of the domains and their amino acid compositions however provide significant insights into the domains' roles in HP1 functioning. The point of attachment of the NTD to the CD is situated adjacent to the hydrophobic cage and possesses a string of acidic residues (Figure 5.4A). The NTD has thus been hypothesized to cooperate with the CD in mediating interactions with the predominantly positively charged H3 histone tail [408]. In contrast, no definitive physiological role has yet been identified for the disordered C-terminal domain that extends from the CSD [421].

The hinge domain of $\sim$70 residues connecting the chromo- and chromoshadow domains is the longest and least evolutionarily conserved domain of HP1 [421]. The domain possesses a multitude of positively charged patches (Figure 5.4A) that have been demonstrated to mediate the protein's binding to both DNA and RNA [422, 423]. Physiologically, this hinge region has been implicated in the nuclear localization of HP1 [424] and importantly for heterochromatin formation, been identified to recruit both histone deacetylases [425] and the inner centromere protein [426].

5.1.2 HP1 Subtypes and Post-translational Modifications

The association of HP1 with the formation and maintenance of heterochromatic structures has lead to the protein being postulated as general repressors of genomic transcription. However, recent evidence has unveiled a role for HP1 in a range of nuclear functions including DNA repair [427, 428], RNA splicing [429, 430], telomere maintenance [431, 432] and quite remarkably even transcription activation [433, 434]. This versatile functional makeup of HP1 arises mainly from **(1)** the presence of multiple HP1 paralogs with domain-specific variation [408] and **(2)** the extensive network of post-translational modifications throughout the pentapartite structure of the protein [435].

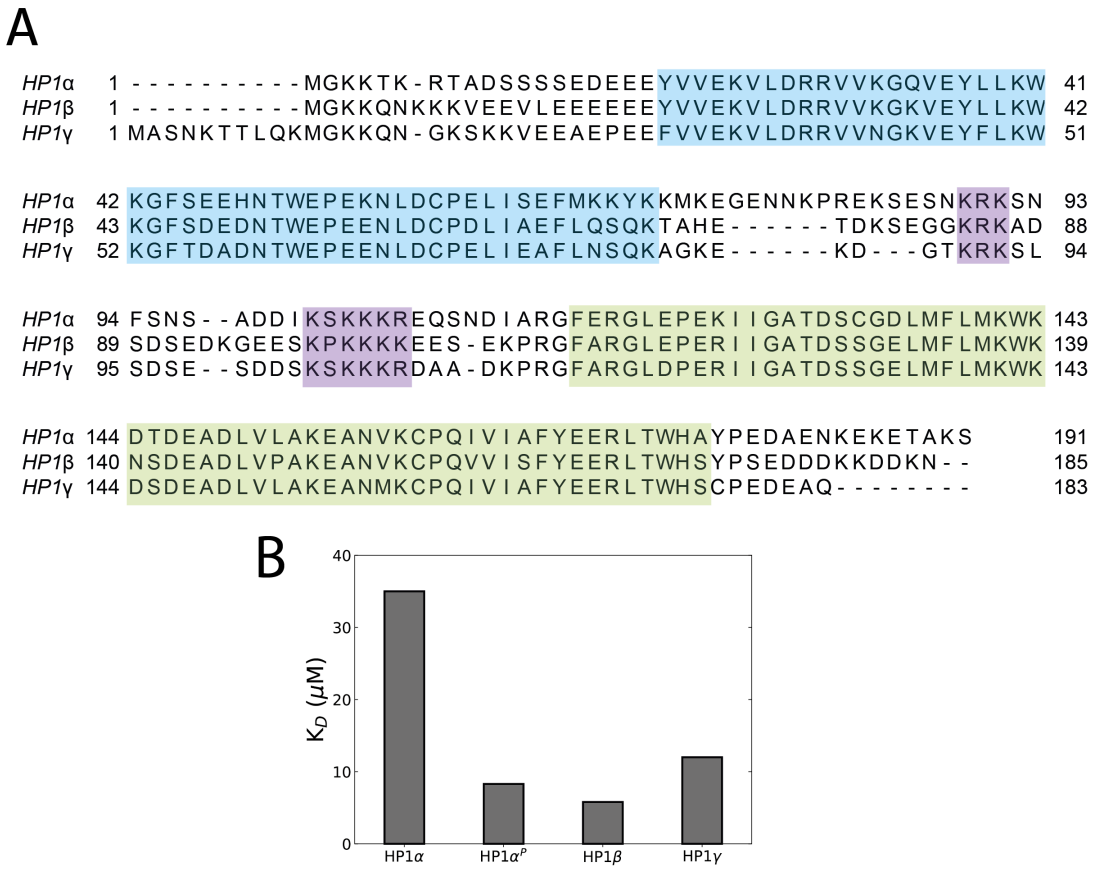


Figure 5.4 The subtypes of HP1. **(A)** Aligment of the human HP1α, HP1β and HP1γ sequences generated using Muscle [311, 312] and visualized using Jalview [313, 314]. The two structured chromo- and chromoshadow domains are illustrated in blue and green respectively. The positively charged patches within the hinge regions identified to mediate nucleic acid binding [436] are shown in purple. **(B)** The dynamic range of affinities of the subtypes for $H3K9_{met}$ histone tail illustrated using their dissociation constants K_D as calculated in the ITC experiments of Hiragami-Hamada et al. [437].

In mammals, three distinct HP1 paralogs have thus far been identified – the α, β and γ forms that are encoded by the CBX5, CBX1 and CBX3 genes

respectively [408]. The HP1α and HP1β paralogs primarily associate with the heterochromatic regions of the genome and help mediate transcriptional gene silencing [438]. In contrast, HP1γ localizes to euchromatin and plays a role in transcription activation/elongation [439]. Corresponding to this variance in cytological distribution and function [409], the three subtypes have been identified to possess distinct binding affinities for both DNA [440] and the nucleosome [437, 441]. Depending on the subtype, this affinity for the H3K9$_{\text{met}}$ peptide can span an incredibly broad range from 5 to 40 μM (Figure 5.4B). However, the three subtypes share a significantly high sequence homology within the two structured chromo- and chromoshadow domains (Figure 5.4A). The HP1 protein has thus often been compared to the Linker Histone H1 where the different biological functions of the subtypes stem from subtle sequence variations within the disordered regions [442].

Further analogous to the linker histone H1, the Heterochromatin Protein 1 subtypes can undergo an extensive network of post-translational modifications throughout their structure [435] that include phosphorylation [443, 444], acetylation [435], methylation [410], ubiquitination [445], SUMOylation [446] and citrullination [447]. Among this plethora of PTMs, the phosphorylation of the subtypes by the Casein Kinase 2 (CK2) enzyme are the best-studied modifications whose physiological relevance has been determined [448]. The subtypes however vary significantly in the number of 'phospho-acceptor' sites and the PTM thus differently regulates both their activity and localization. For example, the phosphorylation of the hinge residues within HP1γ inhibits gene transcription [410] while a similar modification of the HP1β hinge residues has no effect [448].

Additional PTM sites of particular significance lie within the N-terminal domain of the HP1α subtype (Figure 5.5) – a set of four consecutive Serine residues that are not conserved within the HP1β and HP1γ subtypes. The poly-phosphorylation of these four residues has been identified to modify the protein's structural conformation by dissuading intra-dimeric interactions [449] while simultaneously promoting multimeric interactions of the protein [450]. Most significantly, this PTM has been identified to promote the H3K9$_{\text{met}}$ affinity of HP1α nearly 40-fold (Figure 5.4B) [437, 441, 451] – a factor that probably contributes to nearly all *h*HP1α within heterochromatin to be phosphorylated at these sites [437, 452].

5.2 Aims and Summary

The multiplicity of subtypes and their post-translational modifications together combine to allow the Heterochromatin Protein 1 to be both cytologically and functionally promiscuous. However, significant open questions remain on the molecular mechanisms underlying the very preliminary step of HP1 functioning – its recognition and binding of the K9 methylated H3 histone tail. In this Chapter, I use Biased-Exchange Metadynamics simulations to analyze the structural ensembles of the disordered NTDs of HP1 and determine their roles in underpinning the differential nucleosomal affinities of the subtypes. Further, the simulations analyze the significant contribution of HP1α NTD phosphorylation in promoting the subtype's nucleosomal affinity.

5.3 Methods

5.3.1 System and Simulation Setup

The sequences of the *h*HP1α, *h*HP1β, *h*HP1γ and histone H3 were obtained from the Uniprot Database [292]. The corresponding H3-bound HP1 chromodomain structures of the three subtypes (PDB IDs: 3FDT [413], 6D07 [453], 3TZD [159]) were obtained from the PDB [454]. PyMOL [293] was then used to build the unresolved HP1 N-terminal domains and H3-tail segments onto the structures in an extended conformation. The MD simulations of the subtypes were performed using Gromacs 2016 [182] patched with Plumed 2.3.0 [297, 298] in explicit solvent and 0.15 M Na/Cl. Each simulation was repeated using two different force fields - Amberff99SB-ildn [166] and Charmm36M [168, 455]. Ions were modelled using the parameters of Joung and Cheatham [384] with the Amber force field and the parameters of Beglov and Roux [365] with the Charmm force field. The van der Waals interactions were cut-off at 10 Å and electrostatic interactions were calculated using the PME method [186, 187] with a real space cut-off of 10 Å and a reciprocal grid of spacing 1.4 Å. Before production runs, the systems were energy minimized using the steepest descent algorithm and equilibrated for 1 ns each in the NVT and NPT ensembles. The temperatures were maintained at 300 K using the Bussi thermostat [190] and pressure was maintained at 1 bar using the Parrinello-Rahman [299] barostat.

Mutations were introduced into the structures using PyMOL [293] and phosphorylations were introduced in the built HP1α NTD using the PyTM plugin [456]. With Amber, the methylated Lysines were modelled using the parameters

of Papamokos et al. [457] and phosphorylated Serines were modelled using the parameters of Steinbrecher et al. [458]. With Charmm, the methylated Lysines were modelled using the parameters of Grauffel et al. [459] and the phosphorylated Serines using the parameters of Feng et al. [460].

5.3.2 Enhanced Sampling Methodology

Biased-Exchange Metadynamics (BE-Metad) [211] simulations were used throughout to enhance sampling of the conformational states of the disordered regions. The Collective variables within the simulations took two functional forms – the extent of electrostatic contacts (S_{cont}) (Eq. 3.3) and the radius of gyration $\left(S_{\text{rg}}\right)$ (Eq. 3.2). The number of replicas within the BE-Metad varied with the nature of the simulations. The simulations of the HP1 (N-terminal & CD) when bound to the H3 tail utilized five walkers with four biased and 1 unbiased replicas. Two metadynamic potentials biased the Cα atom radius of gyration of the H3 and the NTD each. The two other metadynamic potentials biased the electrostatic contacts – one intra-molecular between the acidic/basic residues within the NTD and the other inter-molecular between the NTD's acidic and the H3's basic residues. The simulations of the *apo*-HP1 (N-terminal & CD) utilized three walkers with two biased and 1 unbiased replicas. Here, the two metadynamic potentials biased the NTD's Rg and its intra-molecular electrostatic interactions.

The simulations were performed for 200 ns per replica to generate an accumulated sampling of 1μs in the *holo* and 600 ns in the *apo* states of HP1. The simulations used a bias-factor of 8 together with an initial hill height of 1.2 kJ. Biasing gaussians were added every 500 trajectory steps and exchanges between replicas were attempted every 10 ps. The complete list of simulations performed together with the force fields used and number of replicas are described in Table 5.1.

5.3.3 IDP Conformation Clustering

The hierarchical clustering methodology of Baul et al. [461] was used to characterize the structural ensemble of the disordered regions and identify representative conformations. For this, all trajectory frames within the unbiased ensemble were first described using a pairwise distance metric D_{ij} defined as

$$D_{ij} = \left(\frac{1}{N_{pairs}} \sum_{a,b} |\left(r^i_{a,b} - r^j_{a,b}\right)|^2 \right)^{1/2} \tag{5.1}$$

where $r^i_{a,b}$ and $r^j_{a,b}$ are the pairwise distances between the C_α atoms 'a' and 'b' in trajectory frames i and j respectively. In the *apo* simulations without the bound the H3, C_α atoms within the N-terminal domain of HP1 were considered for the calculation of D_{ij}. In the simulations with the bound H3 tail, C_α atoms of the HP1 NTD were considered together with those of the H3 tail following the $K9_{met}$ for the calculation of D_{ij}.

Subsequently, the python *scipy* module [462] was used to cluster the D_{ij} of conformations using the Ward variation minimization criterion and the clusters were visualized as dendrograms. The number of representative clusters were calculated using the elbow method [463] where given the set of decreasing inter-cluster distances $\{d_1, d_2, \ldots, d_N\}$, the acceleration of clustering was calculated as $\{d_3 - 2d_2 + d_1, \cdots, d_N - 2d_{N-1} + d_{N-2}\}$ which in-turn was used to calculate the number of clusters k_{clust} as

$$k_{clust} = N + 2 - \underset{i \in [3,N]}{\operatorname{argmax}} \{d_i - 2d_{i-1} + d_{i-2}\} \tag{5.2}$$

The conformation with the lowest C_α RMSD with all other members of the cluster was then chosen as the representative conformation.

5.3.4 List of Simulations

The simulations performed in this chapter are listed in Table 5.1 below.

Table 5.1 List of atomistic simulations performed to investigate the functioning of $H3K9_{met}$ histone tail affinity of HP1 subtypes.

System	Force Field	Time
HP1α	Amberff99SB-ildn	3 × 200 ns
HP1α	Charmm36M	3 × 200 ns
Phos. HP1α	Amberff99SB-ildn	3 × 200 ns
Phos. HP1α	Charmm36M	3 × 200 ns
HP1α + H3	Amberff99SB-ildn	5 × 200 ns
HP1α + H3	Charmm36M	5 × 200 ns
Phos. HP1α + H3	Amberff99SB-ildn	5 × 200 ns
Phos. HP1α + H3	Charmm36M	5 × 200 ns
HP1β	Amberff99SB-ildn	3 × 200 ns
HP1β	Charmm36M	3 × 200 ns
HP1β + H3	Amberff99SB-ildn	5 × 200 ns
HP1β + H3	Charmm36M	5 × 200 ns
HP1γ	Amberff99SB-ildn	3 × 200 ns
HP1γ	Charmm36M	3 × 200 ns
HP1γ + H3	Amberff99SB-ildn	5 × 200 ns
HP1γ + H3	Charmm36M	5 × 200 ns
HP1β D49H,E56K + H3	Charmm36M	5 × 200 ns
HP1γ D58H,E65K + H3	Charmm36M	5 × 200 ns

5.4 Results

5.4.1 HP1α Phosphorylation

The amino acids sequences of the HP1 NTDs display several distinguishable features of IDPs – low hydrophobicity, high charge and low sequence complexity [95]. However, an analysis of the NTD sequences reveals a notable additional characteristic, the sequestration of charged residues within two distinct subregions. The subregion adjacent to the globular CD is enriched in negatively charged residues while the distal subregion is enriched in basic Arg/Lys residues (Figure 5.5). Most significantly, in HP1α, the four Serine residues (S11-S14) phosphorylated in vivo [464] bridge the positive and negative subregions. Biased-Exchange Metadynamics (BE-Metad) were thus first used to analyze the impact of this NTD PTM on the domain's conformational preferences. While the CVs were only applied to the NTD, the entire CD was included to appropriately consider the effects of the structural restraints it places on the IDR's mobility.

Figure 5.5 Alignment of the NTDs of the HP1α, HP1β and HP1γ sequences generated using Muscle [311, 312] and visualized using Jalview [313, 314]. The positively charged residues are shaded green and the negative residues are in red.

Over the course of the simulation, the structure of the chromodomain was unaffected by the phosphorylation state of the NTD with its Cα root mean square fluctuation (RMSF) displaying negligible differences (Figure 5.6A). However, the conformational preferences of the disordered NTD were significantly affected by the PTM where, regardless of the force field, the phosphorylated IDR displayed a significant increase in both the average and fluctuations of the end-to-end distances (R_{EE}). To identify the molecular basis for this decrease in compaction with PTM, the HP1α NTD ensembles from the BE-Metad simulations were hierarchically clustered. In both the WT and phosphorylated states, the conformational ensembles partitioned into two distinct clusters whose representative conformations are illustrated in Figure 5.6D.

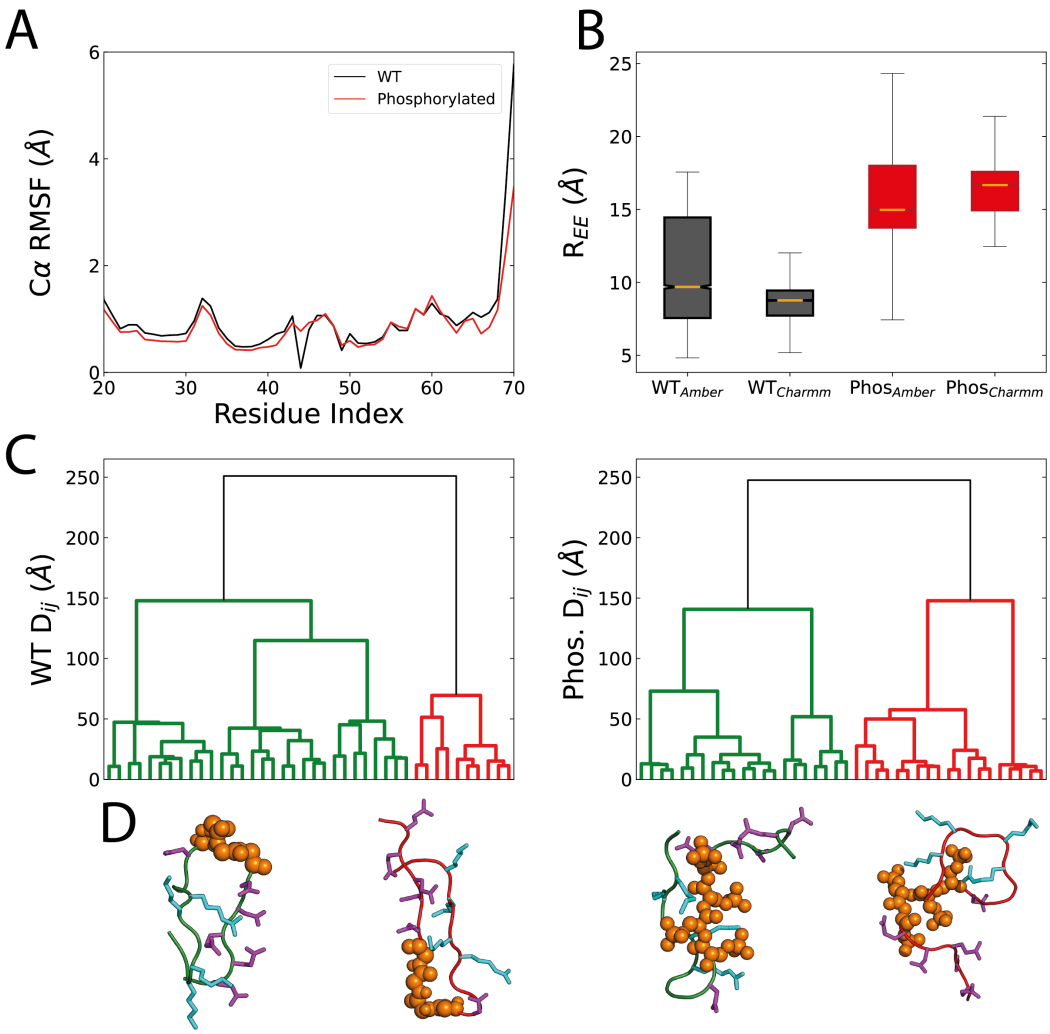


Figure 5.6 Impact of phosphorylation on HP1α conformations. **(A)** Cα RMSF comparison of the HP1α chromodomain between the WT (black) and phosphorylated (red) states of the NTD. **(B)** Cα R_{EE} for the HP1α NTD in WT and phosphorylated states. The boxes extend from the lower to upper quartile values and whiskers span from the 5th to the 95th percentile. **(C)** Hierarchical clustering of the NTD IDR visualized as a dendrograms in the WT (Left) and phosphorylated (Right) states. The two clusters within each phosphorylation states are in green and red. **(D)** Representative conformations of the NTD IDR. The backbone is coloured in green/red to correspond to their clustering in (C). The sidechains of the acidic residues are illustrated in purple and the basic residues in cyan. The sidechains of the serine/phosphoserine residues are shown in orange.

Regardless of the phosphorylation state, the NTD conformational ensemble was dictated by the formation of intra-IDR salt-bridges albeit between different residues. A glance of the representative conformations suggested that in the WT, the NTD primarily adopts a hairpin configuration bent around the four S11-S14 Serine residues to form salt bridges between the basic and acidic subregions. Contrarily, when phosphorylated, the NTD adopts a tighter turn around the T9/A10 residues to form salt-bridges between the basic subregion and the even

more acidic (-2 charged) phosphoserine residues. To confirm this, the methodology of Winogradoff et al. [156] was used where the IDR was separated into three segments ($_3$KKTKR$_7$, $_{11}$SSSS$_{14}$, $_{15}$EDEEE$_{19}$) and the number of intersegment contacts were separately calculated in the WT and phosphorylated states. Subsequently, the relative segmental contacts was calculated as a ratio $\frac{n_{Phos}}{n_{WT}}$. Thus, a relative contact frequency <1 indicates a decrease in interactions upon PTM in comparison to the WT and vice-versa. Upon phosphorylation, contacts between the basic and acidic subregions reduced to a third of the WT while those between the basic subregion and Serines increased 8-fold (Figure 5.7A). The Cα RMSF of the NTDs (Figure 5.7B) further confirms this conformational change with a leftward shift in the peak suggesting a transitioning of the flexible turn from the Serine segment to the T9/A10 residues.

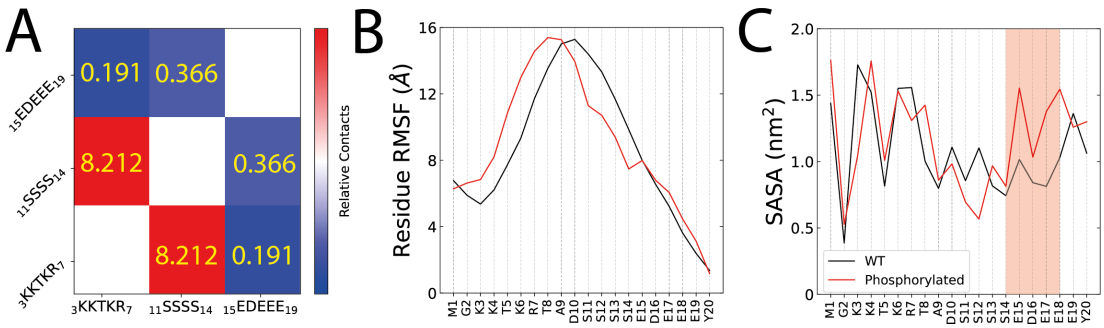


Figure 5.7 Impact of phosphorylation on HP1α NTD conformations. **(A)** Relative contact maps between three different NTD subsegments. The ratios range from blue to red and the absolute values are enumerated in yellow. A contact between subsegments was assumed if a non-hydrogen atom of one was within 3.2 Å of a non-hydrogen atom of another. **(B)** Cα RMSF comparison of the HP1α NTD between the WT (black) and phosphorylated (red) states. **(C)** Solvent accessible surface area (SASA) of the HP1α NTD calculated using a probe of radius 0.14 nm. The acidic patch with enhanced SASA is shaded red.

The change in intra-NTD salt-bridge patterns most significantly affects the interactions of the acidic subsegment with its contacts with both the basic and Serine patches decreasing considerably. To validate this, the per-residue solvent accessible surface area (SASA) was compared between the WT/phosphorylated states (Figure 5.7C) and observed notably increased accessibility for the acidic patch signifying its reduced protein interactions. Crucially, this result offers a plausible hypothesis for the structural mechanism for the NTD phosphorylation's role in promoting H3 binding [437, 441, 451]; where the acidic patch now 'unconfined' from interacting intra-molecularly with the basic patch can now transition to bind the H3 tail.

5.4.2 HP1α NTD phosphorylation and H3 binding

The phosphorylation-induced changes in the conformation of the NTD offered a plausible hypothesis to the PTM's role in promoting H3 binding. To test this hypothesis, the BE-Metad simulations were repeated albeit with the H3 tail bound to the CD in its crystallographic conformation. However, while only the $_{1}$ARTKQTARK$_{met}$S$_{10}$ segment of the tail has been crystallized, the full-length H3 tail was built and modelled. Over the course of the simulation, regardless of the phosphorylation state, the H3 peptide remained bound to HP1α with the K9met residue residing within the hydrophobic cage and the tail retaining it position between the two CD β-strands to form a three-stranded β-sheet (Figure 5.8A). The conformations of the HP1α NTD were strongly dependant on the binding of the H3 peptide with the mean RMSF fluctuation of the domain's Cα atoms reducing considerably by over 4 Å in the phosphorylated case (Figure 5.8B).

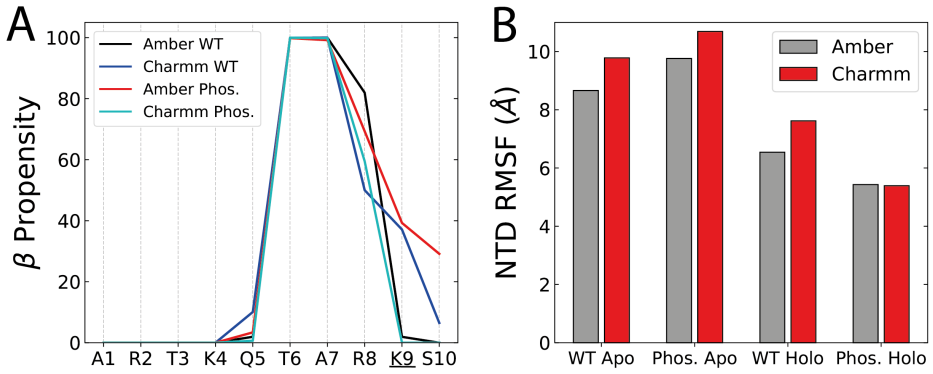


Figure 5.8 H3K9$_{met}$ peptide within the HP1α chromodomain. **(A)** Per-residue β-secondary structural content for the H3 peptide from BE-Metad simulations when within the HP1α CD. The secondary structures were assigned using DSSP [321] and the methylated K9 residue is highlighted. **(B)** Mean RMSF fluctuations for the HP1α NTD Cα atoms compared across phosphorylation and H3 bound/unbound states.

Figure 5.9A plots the hierarchical clustering of the HP1α-H3 IDRs where the ensembles partition into two distinct clusters for both the WT and phosphorylated states. However, in the WT, the clusters only differ in the orientations of the IDRs relative to the globular chromodomain. In both WT clusters, the NTD adopts a configuration unchanged from the *apo*-state with salt-bridges between the acidic/basic patches still dominating. The acidic patch thus preferentially forms intra-molecular interactions in-lieu of binding to the multitude of basic residues within the H3 tail. The WHAM-weighted number of contacts between HP1 and the H3 tail plotted in Figure 5.9C quantifies this conforma-

tional preference where inter-IDR salt-bridges were present in fewer than 8% of frames.

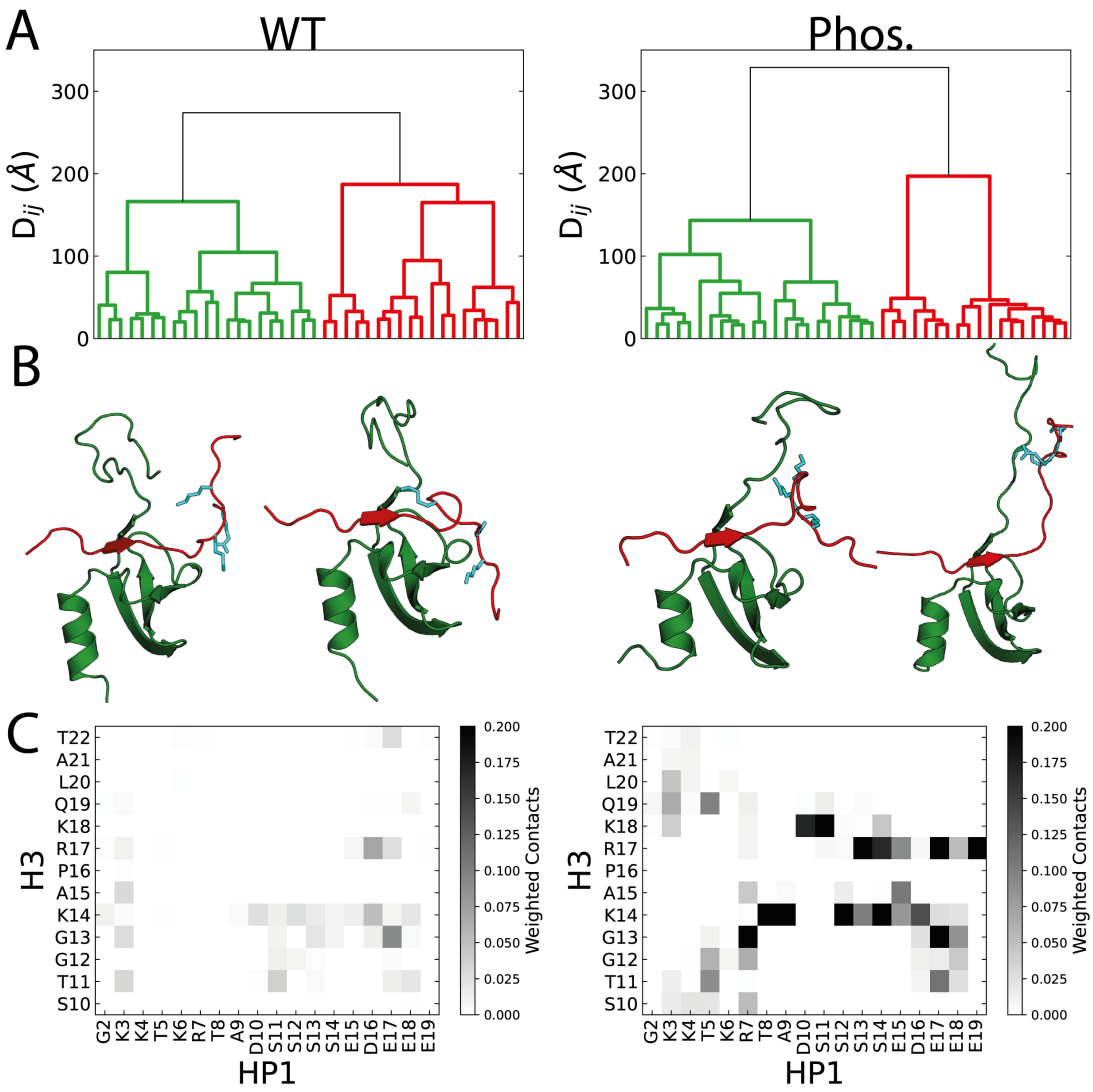


Figure 5.9 Interactions of the H3K9$_{\text{met}}$ peptide with the HP1α NTD. **(A)** Hierarchical clustering of the NTD/H3 IDRs visualized as a dendrograms in the WT (Left) and phosphorylated (Right) states. The two cluster within each phosphorylation states are in green and red. **(B)** Representative conformations for each of the two clusters within the WT/phosphorylated states of HP1α. The globular chromodomains were aligned to allow easier comparison between the clusters and the basic residues within the extended H3 tail are illustrated in cyan. **(C)** Normalized number of contacts between residues of the H3 tail and HP1α NTD. A contact was assumed if a non-hydrogen atom of one residue was within 3.2 Å of a non-hydrogen atom of the other. The colour-spectrum was kept consistent between the WT/phosphorylated states to allow comparison of the interaction frequencies.

Conversely, in the case of the phosphorylated HP1α, the two clustered conformations are significantly different. In the first conformation representative of nearly three-fourths (76.4%) of the ensemble, the phosphorylated NTD retains the partially bent conformation from the *apo* state. Here, the intra-NTD salt-

bridges between the positive subregion and the phosphorylated S11-S14 patch in turn allows the 'free' $_{15}$EDEEE$_{19}$ segment to facilitate interactions with three basic residues within the extended H3 tail – K14, R17 and K18. In the second less populated cluster (23.6%), the NTD adopts a fully extended conformation that permits both the acidic and phosphoserine residues to form interactions with basic H3 tail. The WHAM-weighted contacts (Figure 5.9C) quantifies this impact where inter-IDR salt bridges between the H3 tail and NTD are observed in upto 50% of frames.

The results thus hypothesize an 'intra-molecular inhibition' model for the nucleosomal binding of HP1α. In the WT, the acidic residues adjacent to the aromatic cage are precluded from assisting H3 binding by the basic residues within the NTD. The acidic residues are 'liberated' upon phosphorylation to bind the extended H3 tail. Crucially, this model provides a molecular mechanism for two disparate sets of experimental data investigating HP1-H3 interactions. Firstly, mutagenesis experiments [451] predicted HP1α truncation mutants without the NTD basic subsegment to possess a similar H3K9$_{\text{met}}$ peptide affinity as the phosphorylated protein. Further, the model suggests a significant physiological role for the K14/R17/K18 basic residues of the H3 tail situated away from the hydrophobic cage – three residues whose Ala swap mutations were sufficient to abolish HP1 binding [144, 437].

5.4.3 HP1β and HP1γ NTD configurations

The NTDs of the HP1β and HP1γ display a similar sequestration of basic and acidic residues towards opposite ends of the disordered amino-terminal domain (Figure 5.5). However, they vary significantly in both sequence and length in comparison to HP1α with a relative sequence identity as low as 25% [408]. More importantly, the two isoforms lack the phosphorylatable Serine residues within the NTD that potentiate H3 biding (Figure 5.9C) [437, 441, 451]. However, the two subtypes still display upto a six-fold enhanced affinity for the H3K9$_{\text{met}}$ tail in-comparison to the WT HP1α. BE-Metad simulations of the subtypes in *apo* states were thus first used to analyze the impact of these sequence differences on the amino-terminal disordered region.

The NTD of HP1β while similar in length to HP1α, possesses a sequence of negatively charged Glu residues in-place of the phosphorylatable Serines (Figure 5.5). Thus, unlike the HP1α subtype, there exists a limited separation of the basic and acidic subsegments. The primary cluster obtained from the hierarchical clustering of the BE-Metad trajectory reveals the effects of this lack

of hairpin 'turn' residues with the NTD adopting a collapsed conformation that allows promiscuous electrostatic interactions between the two subregions. However, the secondary cluster that accounts for ∼35% of the frames within the unbiased ensemble provided an intriguing result with the NTD folding over the $_{10}$VEEV$_{13}$ segment to allow the extended $_{15}$EEEEEE$_{20}$ Glu patch to form salt-bridges with the triple lysine $_{7}$KKK$_{9}$ motif. In-turn, the K3 and K4 residues are positioned to interact with the globular chromodomain. The WHAM-weighted inter-domain contacts plotted in Figure 5.10B validates the prediction of the hierarchical clustering where inter-domain interactions are primarily focussed around the K3/K4 residues positioned adjacent to the CD in the secondary cluster.

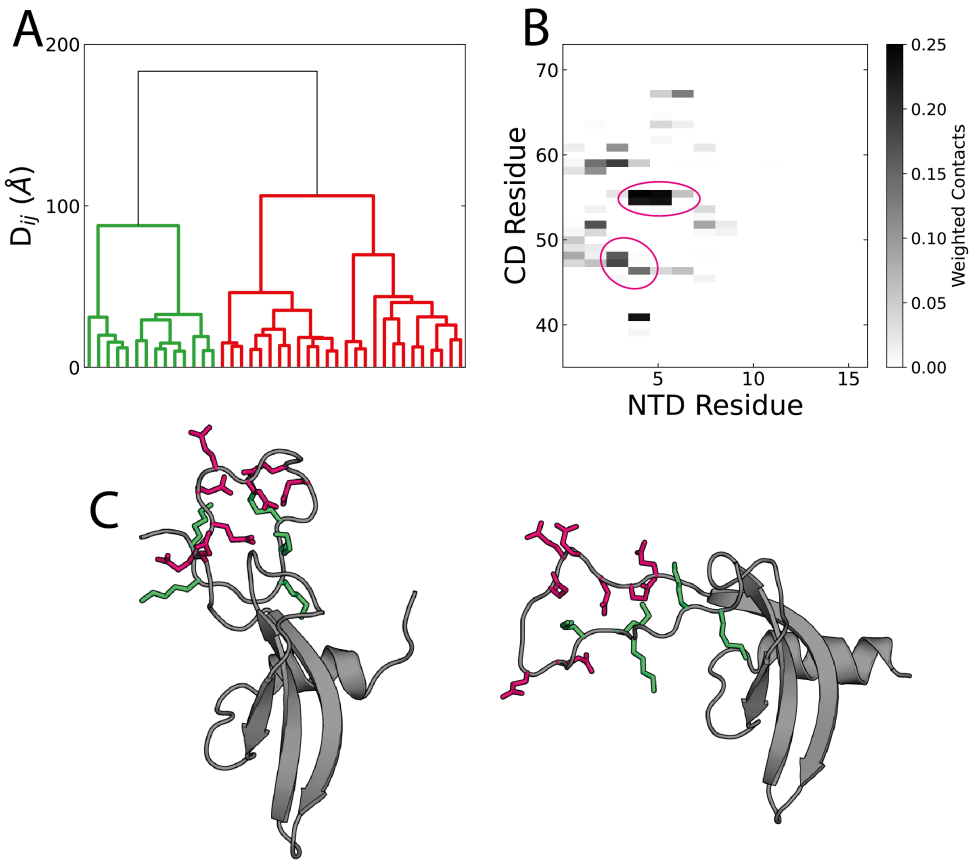


Figure 5.10 HP1β NTD conformations. **(A)** Hierarchical clustering of the HP1β NTD visualized as a dendrogram. The two clusters are in green and red. **(B)** Normalized number of contacts between residues of the NTD and CD. A contact was assumed if a non-hydrogen atom of one residue was within 3.2 Å of a non-hydrogen atom of the other. Circles highlight the $_{47}$DED$_{49}$ and $_{53}$EPEE$_{56}$ acidic interaction sites within the CD. **(C)** Representative conformations for each of the two clusters of the HP1β NTD. The CDs were aligned to allow easier comparison between the clusters and the basic/acidic residues within the NTD are illustrated in green/red respectively.

Crucially, the positive K3/K4 residues predominantly interacted with two acidic patches within the CD – $_{47}$DED$_{49}$ and $_{53}$EPEE$_{56}$. A comparison of the CD

sequences reveals that the two patches correspond to the only sites divergent in electrostatic characteristics [437] between the HP1β and the HP1α subtypes (Figure 5.11A). The HP1γ subtype retains this divergence in electrostatic sites within the CD and is further differentiated from the HP1α isoform in the length of the disordered NTD where the basic subsegment is upto 20 residues long and has two more positively charged residues. In fact, HP1γ is the only subtype with a net positive charge within the NTD. The hierarchical clustering reveals the effects of these sequence differences with a complete absence of the hairpin conformations observed within the NTDs of the α/β subtypes. Inter-domain NTD-CD interactions dominate the conformational ensemble with the representative configurations of the two clusters only differing in the relative orientations of the NTD (Figure 5.11D) that in-turn dictate the formation of either the K14-D58 or the K5-E65 salt-bridges (Figure 5.11C) – two cross-domain interactions that are together present in ∼47% of the frames.

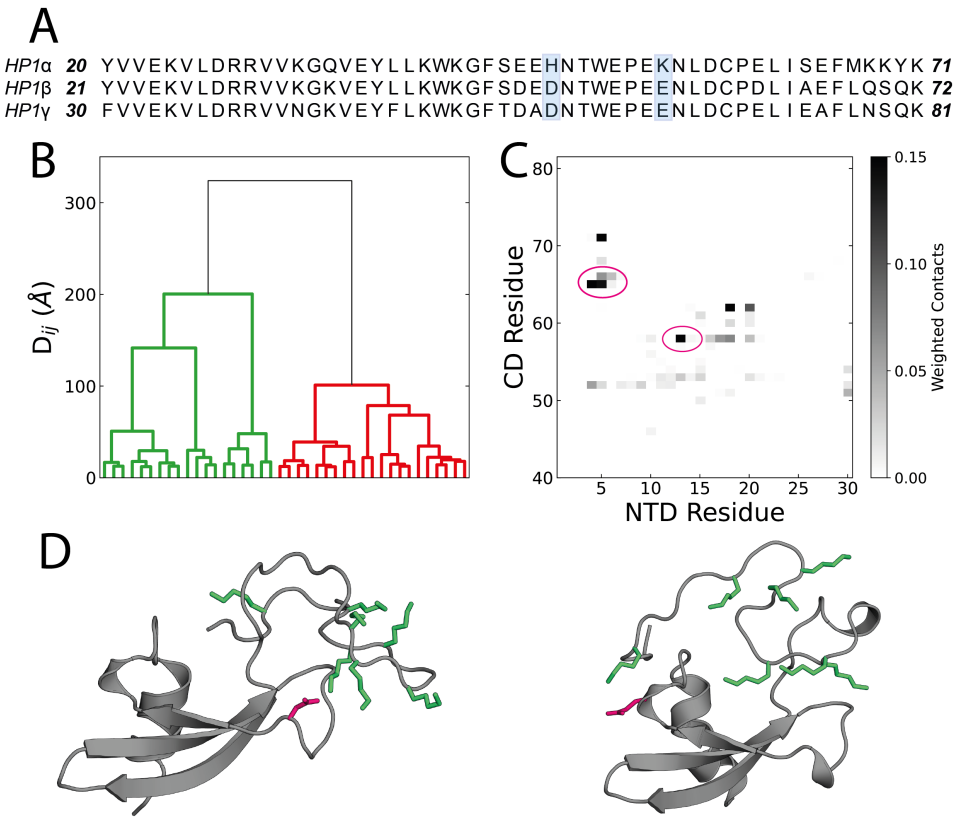


Figure 5.11 (A) Alignments of the CD sequences of the HP1 subtypes. The two sites with divergent electrostatic characteristics are highlighted in blue. **(B)** Hierarchical clustering of the HP1γ NTD visualized as a dendrogram. The two clusters are in green and red. **(C)** Normalized number of contacts between residues of the NTD and CD. A contact was assumed if a non-hydrogen atom of one residue was within 3.2 Å of a non-hydrogen atom of the other. Circles highlight the $_{56}$DAD$_{58}$ and $_{62}$EPEE$_{65}$ acidic interaction sites within the CD. **(D)** Representative conformations for each of the two clusters of the HP1γ NTD. The D58 (Left) and E65 (Right) residues identified to interact with the NTD in (B) are illustrated in red and the NTD basic residues in green.

The consistent nature of these cross-domain NTD-CD interactions within the HP1β/γ subtypes are in stark contrast to the HP1α subtype where the conformational ensemble is dominated by intra-NTD salt-bridges in both WT and phosphorylated states. While these increased inter-domain interactions observed within the BE-Metad simulations do not allow a hypothesis for the increased H3K9$_{\text{met}}$ affinity of the subtypes, they demonstrate the significance of the cooperation between the two adjacent disordered/ordered domains. In fact, a slew of modelling studies that have been unsuccessful in deciphering the molecular mechanism for this differential affinity of subtypes have only considered either the globular CD [465, 466] or the IDR [457] but not both.

5.4.4 Chromodomain residue polymorphisms impact H3

The impact of this chromodomain induced HP1 β/γ NTD conformation on H3 binding was next assessed through BE-Metad simulations. Similar to the HP1α case, while only the $_{1}$ARTKQTARK$_{\text{met}}$S$_{10}$ segment of H3 has been crystallographically resolved [159, 453], the full-length H3 tail was built and modelled. Over the course of the simulation, in both subtypes, the H3 peptide remained bound to HP1 to retain its position within the chromodomain three-stranded β-structure (Figure 5.12A). However, the binding of the H3 tail had contrasting effects on the flexibility of the subtypes' NTD. In the HP1β subtype, H3 binding leads to a significant drop in the mean RMSF of the disordered domain while having no effect in the case of HP1γ (Figure 5.12B).

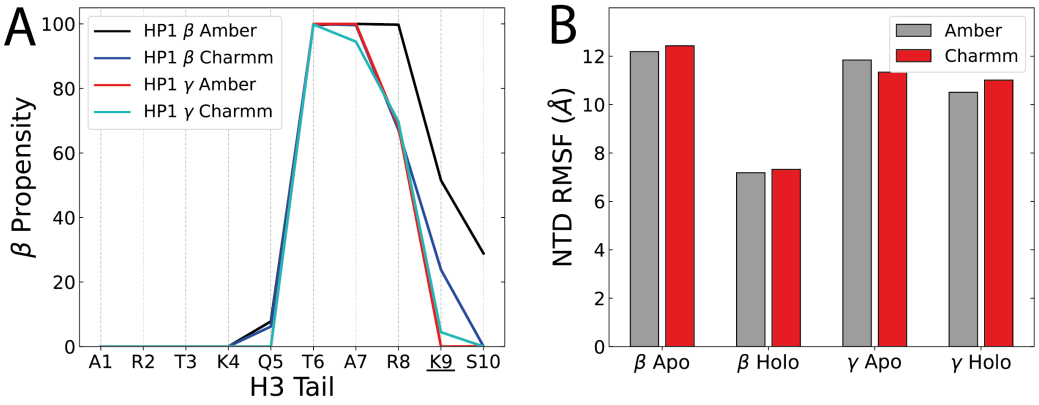


Figure 5.12 H3K9$_{\text{met}}$ peptide within the HP1β/γ chromodomain. **(A)** Per-residue β-secondary structural content for the H3 peptide from BE-Metad simulations when within the HP1β/γ CD. The secondary structures were assigned using DSSP [321] and the methylated K9 residue is highlighted. **(B)** Mean RMSF fluctuations for the HP1β/γ NTD Cα atoms compared across H3 bound/unbound states.

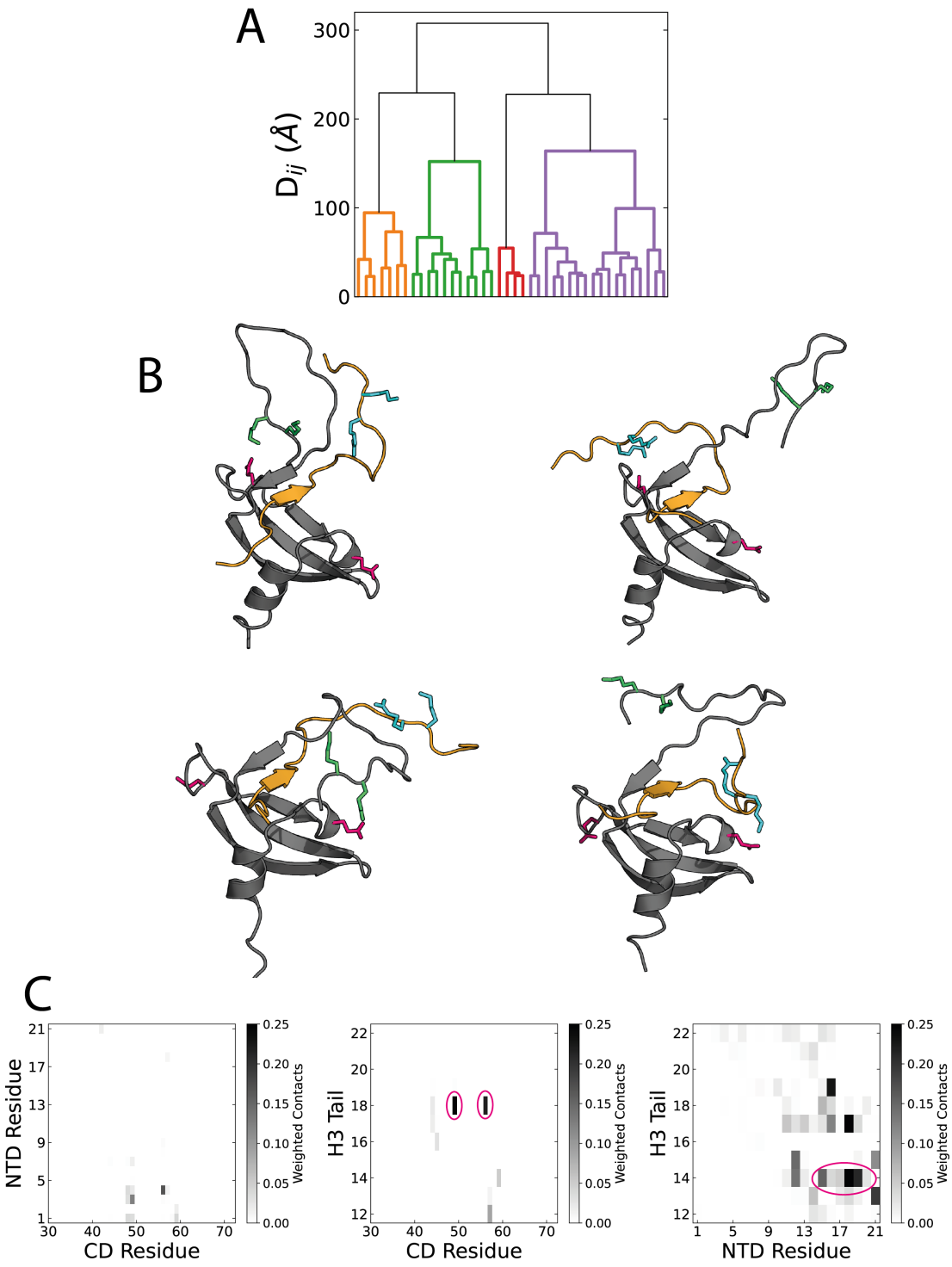


Figure 5.13 H3K9$_{met}$ peptide within HP1β. **(A)** Hierarchical clustering of the NTD/H3 IDRs visualized as a dendrogram with the four clusters identified coloured differently. **(B)** Representative conformations for each of the four clusters identified in (A). The HP1β is in grey and the H3 peptide is in orange. The D49/E56 chromodomain residues are shown in red. The basic K3/K4 residues of the NTD are in green and the R17/K18 residues of the H3 tail are in cyan. **(C)** Normalized number of contacts between residues of the H3 tail, NTD and the chromodomain. A contact was assumed if a non-hydrogen atom of one residue was within 3.2 Å of a non-hydrogen atom of the other. Circles highlight salt-bridges between H3 tail basic residues and acidic residues of the HP1β NTD/CD.

Figure 5.13A plots the hierarchical clustering of the HP1β-H3 IDRs where the ensemble partitions into four distinct clusters. The representative clusters illustrate the molecular mechanism of the drop in the NTD flexibility where there is a complete absence of the primary collapsed conformation with promiscuous electrostatic interactions observed in the H3-unbound state (Figure 5.10C). Instead, the four clusters are primarily formed by the basic K3/K4 residues of the NTD and R17/K18 residues of the H3 tail alternatively interacting with the two acidic patches within the chromodomain. Crucially, these interactions positioned the two IDRs adjacently to allow the formation of sustained salt-bridges between K14 of H3 and E18/E19 of the NTD that were together present in nearly 42% of the trajectory frames (Figure 5.10C).

In the case of HP1γ, the conformational ensemble partitioned into two clusters with two distinct configurations of the H3 tail that differed in their interactions with the E65 residue within the chromodomain. Within the primary cluster, R17/K18 residues at the distal end of the H3 tail form salt-bridges with the E65 residue and in-turn position the H3K14 adjacent to the HP1γ NTD acidic patch. However, both IDRs formed negligible interactions with the other polymorphic D58 residue of the CD possibly stemming from the reduced electronegativity of this HP1γ $_{56}$DAD$_{58}$ patch in-contrast to $_{47}$DED$_{49}$ in HP1β.

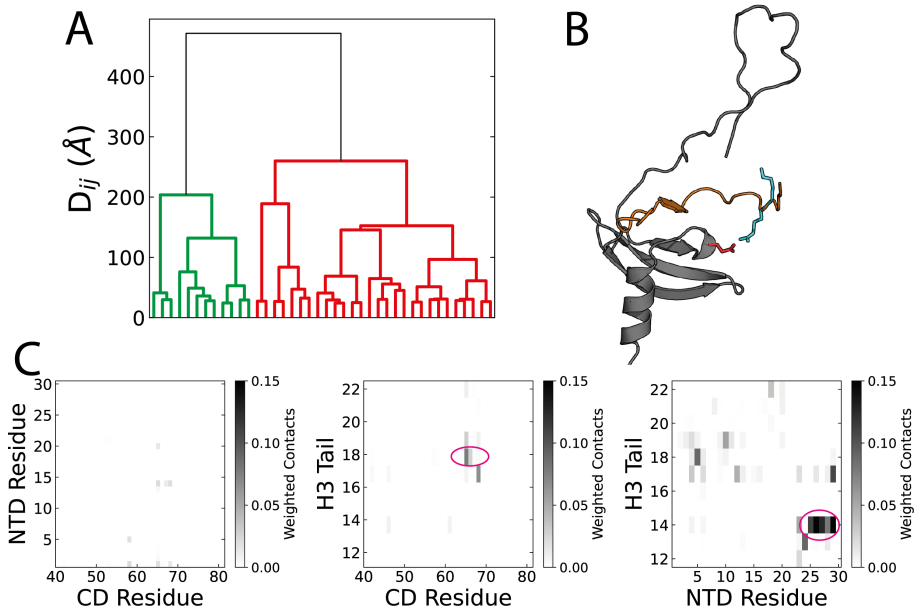


Figure 5.14 (A) Hierarchical clustering of the NTD/H3 IDRs visualized as a dendrogram with the two clusters identified coloured differently. **(B)** Representative conformations for the primary cluster identified in (A). The HP1γ is in grey and the H3 peptide is in orange. The E65 chromodomain residue is in red and the R17/K18 residues of the H3 tail are in cyan. **(C)** Normalized number of contacts between residues of the H3 tail, NTD and the chromodomain. A contact was assumed if a non-hydrogen atom of one residue was within 3.2 Å of a non-hydrogen atom of the other. Circles highlight salt-bridges between H3 tail basic residues and acidic residues of the HP1γ NTD/CD.

The accumulated MD sampling provides a hypothesis for the molecular mechanism of the differential H3K9$_{\text{met}}$ affinities of the HP1 subtypes (Figure 5.4B); where the NTD, the CD and PTMs together participate in enabling the basic residues within the H3 tail to form subtype-dependant differing patterns of salt-bridges. Thus, while the $_{7}$ARK$_{\text{met}}$S$_{10}$ segment of the H3 tail might serve as the CD recognition motif [144], the HP1 affinity is instead tuned by residues at the distal end of the tail away from the hydrophobic cage binding site [437]. Figure 5.15 validates this hypothesis where the salt-bridge frequencies of the H3 tail basic residues when within HP1α/$\alpha_{\text{phos.}}$/β and γ varies proportionally to their experimentally observed binding affinities (Figure 5.4B) [437, 441]. In fact, a repeat of the BE-Metad simulations of the H3-bound HP1β and γ subtypes with the D49H/E56K and D58H/E65K mutations respectively to match the HP1α CD, lead to a reduction in the H3 tail interactions to the levels of WT HP1α – a result consistent with the experimental observations of Hiragami-Hamada et al. [437] where this CD swap lead to a swap in H3 affinity.

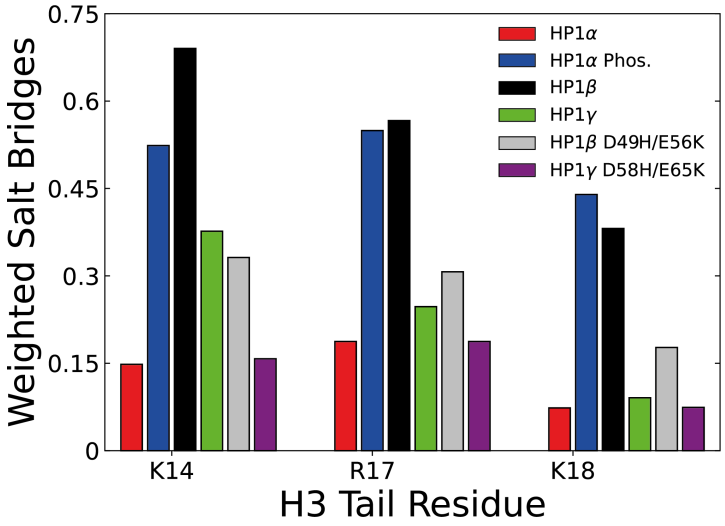


Figure 5.15 The normalized number of salt-bridges formed by the K14, R17 and K18 residues of the H3 tail from BE-Metad simulations with the various modelled HP1 systems. A salt-bridge was assumed if the basic residues' guanidinium/butylammonium group was within 3.4 Å of HP1 Asp/Glu residues' acetate group [467].

5.5 Conclusion and Further Work

In this chapter, a large set of Biased-Exchange Metadynamics simulations is used to sample the conformational landscape of the disordered N-terminal domains of the three HP1α, HP1β and HP1γ subtypes – both in-isolation and when bound to the H3K9$_{\text{met}}$ peptide. The WT HP1α NTD adopts an auto-inhibited conformation that precludes its acidic residues from interacting with

the basic H3 tail. This conformation can be modified by the phosphorylation of Serine residues within the IDR to permit the 'unconfining' of the acidic residues to facilitate H3 binding. In contrast, the HP1β and HP1γ subtypes promote H3 binding through the synergetic interactions of the CD and NTD, where specific inter-domain interactions in-turn orient negative residues to facilitate H3 binding. Together, the results offer a structural hypothesis of H3-HP1 binding where the tail's $_{7}\text{ARK}_{\text{met}}\text{S}_{10}$ segment serves as the hydrophobic cage recognition motif while the extended H3 tail tunes affinity. Importantly, this computationally predicted role for the distal end of the H3 tail correlates with the experimentally observed effect where a swap of these basic residues to Alanine abolishes HP1 binding [144, 437].

The results of the atomistic simulations within this chapter while offering an interesting testable hypothesis of HP1 functioning, should be contextualized as a subset of the large multi-domain interaction interface between HP1 and the nucleosome that includes its CSD to the PxVxL H3 motif [468], hinge to DNA [449] and even the CD to the nucleosomal surface [417, 447]. The work however serves as an important basis for investigating the functional role of other post-translational modifications in H3-HP1 interactions – the most relevant being the simultaneous dual PTM of K9_{met} and S10_{phos} within the H3 tail [469–471] that has been identified to function as a binary on-off switch [472, 473].

Chapter 6

Summary and Outlook

The use of Molecular Dynamics (MD) simulations to characterise protein conformations dates back to 1977 when McCammon, Gelin and Karplus modelled the globular 58 amino acid bovine pancreatic trypsin inhibitor for 8.8 ps [474]. The field of protein MD simulations has since come a long way with the development of software like the Folding@Home project [475] and hardware like Anton [476] now even permitting the breach of the millisecond mark [195]. However, despite Intrinsically Disordered Proteins (IDP) making up nearly 30% of the human proteome [477], they had hitherto lagged considerably behind their globular counterparts in their amenability to study through MD simulations [134].

The gap has however closed considerably over the past decade due to significant efforts along two directions – **(1)** the identification of the drawbacks of traditional protein force fields in describing IDP conformations [169, 478, 479] and the subsequent development of improved parameters [168, 361, 480, 481], and **(2)** the development of enhanced sampling methodology to allow improved modelling of the IDP phase space [198, 202, 211, 216] together with their efficient computational implementations [182, 298, 297]. In this thesis, I attempted to take advantage of these developments to understand the structural mechanisms of functioning of disordered regions with a set of proteins that form the core of chromatin structure – histones, linker histones H1 and the heterochromatin protein HP1.

- In Chapter 3, a combination of enhanced-sampling MD simulations was used to characterize the H1 linker histone protein when bound to the nucleosome. The disordered carboxy-terminal domain adopts a predominantly disordered yet compact configuration that allows its positioning between the two linker DNA strands and thereby enabling a screening of the entering/exiting arms. In contrast, the short amino-terminal domain underwent a disorder-to-order transition to form an α-helix binding one DNA strand. The asymmetric interactions of the disordered regions

in-turn induces an asymmetric nucleosome with unevenly curved DNA strands. In addition, the work demonstrates how the delicate balance of electrostatic interactions can be modified through the introduction of post-translational modifications like phosphorylations.

- In Chapter 4, the enhanced sampling simulation methods were applied to determine the molecular basis of the differential functioning of the three different linker histone subtypes H1.0, H1.1 and H1.2. When isolated in solution, the amino terminal domains of all three subtypes adopt primarily unstructured conformations. However, a sequence analysis of the NTDs suggests a built-in amphiphilicity where positive Lys residues are at every third/fourth site – an arrangement ideal to form a positive face when forming a helix. Correspondingly, the reduction of inter-Lysine sidechain repulsion leads the domain to undergo a disorder-to-order transition to helical conformations. Crucially, this helical propensity and thus the positive 'face' varies proportionally to the subtypes' experimental nucleosomal affinities.

- In Chapter 5, the MD simulations were applied to determine the molecular basis for the differential affinities of the Heterochromatin Protein 1 (HP1) α,β and γ subtypes for the methylated H3 histone. The disordered amino terminal domains of HP1(s) display a sequestration of charged residues towards opposite ends of the region – an arrangement whose interactions are varied in a subtype-dependant manner to differentially unshackle the negative residues to bind the positive H3 tail. Crucially, the simulations also rationalize the roles of two distinct sequence effects identified to promote H3 binding – **(1)** phosphorylation of Ser residues within the NTD of the HP1α subtype and **(2)** acidic patches within the globular chromodomains of the HP1β and γ subtypes.

The proteins investigated within this thesis however make up an extremely small subset of the disordered proteome spectrum that makes up the genome [107]. Transcription factors [90], activating domains [482], architectural proteins [408] and PTM enzymes [483] all play significant roles in reading/modifying chromatin structure and are made up of physiologically relevant disordered regions. In fact, even the proteins investigated in this thesis possess subtypes that are significantly divergent within their disordered segments. For example, the H1x (H1.10) somatic linker histone subtype has been predicted to possess a NTD with greater DNA affinity than even the H1.0 subtype modelled in Chapter 4 [256]. Similarly, the CENP-A variant of the core histone H3 possesses a disordered tail lacking the ARK_{met} motif [484] necessary for recognizing and

binding HP1 [144]. Work in this thesis provides the basis extendable in a piecewise manner to understand the overlapping roles of the multitude of disordered regions.

Work in this thesis also investigated the phosphorylation post-translational modification that modifies the conformational preference of H1 and HP1 proteins. However, IDPs are extensively decorated with a network of post-translational modifications [154] and improvements in Mass-Spectrometry techniques [485] are allowing the discovery of novel modifications [76] and modification sites [263] – some with hitherto unidentified physiological roles [486, 487]. Provided the appropriate parameterization of the novel residues [488, 489], the enhanced sampling methodology of this work is extendable to understand the roles/functioning of these PTMs.

Finally, the atomistic modelling work in this thesis is computationally inexpedient to understand chromatin structure beyond the nucleosomal scale. However, the IDP ensembles generated from the enhanced sampling simulations are mappable to generate bespoke coarse-grain parameters to allow their impact on nanoscale chromatin to be more accurately modelled [490].

References

[1] D. Nelson, A. L. Lehninger, and M. M. Cox, *Lehninger principles of biochemistry*. Macmillan, 2008.

[2] J. Watson and F. Crick, "Molecular Structure Of Nucleic Acids: A structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[3] A. L. Olins and D. E. Olins, "Spheroid chromatin units (v bodies)," *Science*, vol. 183, no. 4122, pp. 330–332, 1974.

[4] P. Oudet, M. Gross-Bellard, and P. Chambon, "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit," *Cell*, vol. 4, no. 4, pp. 281–300, 1975.

[5] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 A resolution.," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[6] M. Bhasin, E. L. Reinherz, and P. A. Reche, "Recognition and Classification of Histones Using Support Vector Machine," *Journal of Computational Biology*, vol. 13, no. 1, pp. 102–112, 2006.

[7] L. Mariño-Ramírez, M. G. Kann, B. A. Shoemaker, and D. Landsman, "Histone structure and nucleosome stability," *Expert Review of Proteomics*, vol. 2, no. 5, pp. 719–729, 2005.

[8] G. Arents and E. N. Moudrianakis, "The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization.," *Proceedings of the National Academy of Sciences*, vol. 92, no. 24, pp. 11170–11174, 1995.

[9] B. R. Zhou, J. Jiang, H. Feng, R. Ghirlando, T. S. Xiao, and Y. Bai, "Structural Mechanisms of Nucleosome Recognition by Linker Histones," *Molecular Cell*, vol. 59, no. 4, pp. 628–638, 2015.

[10] K. Luger and T. J. Richmond, "The histone tails of the nucleosome," *Current Opinion in Genetics and Development*, vol. 8, no. 2, pp. 140–146, 1998.

[11] J. C. Hansen, C. Tse, and A. P. Wolffe, "Structure and function of the core histone N-termini: More than meets the eye," *Biochemistry*, vol. 37, no. 51, pp. 17637–17641, 1998.

[12] F. Gordon, K. Luger, and J. C. Hansen, "The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays," *Journal of Biological Chemistry*, vol. 280, no. 40, pp. 33701–33706, 2005.

[13] C. Tse and J. C. Hansen, "Hybrid trypsinized nucleosomal arrays: Identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction," *Biochemistry*, vol. 36, no. 38, pp. 11381–11388, 1997.

[14] G. Arya and T. Schlick, "A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments," *Journal of Physical Chemistry A*, vol. 113, no. 16, pp. 4045–4059, 2009.

[15] R. Collepardo-Guevara, G. Portella, M. Vendruscolo, D. Frenkel, T. Schlick, and M. Orozco, "Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study," *Journal of the American Chemical Society*, vol. 137, no. 32, pp. 10205–10215, 2015.

[16] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 A resolution," *Journal of Molecular Biology*, vol. 319, no. 5, pp. 1097–1113, 2002.

[17] V. Ramakrishnan, "Histone H1 and chromatin higher-order structure," *Critical Reviews in Eukaryotic Gene Expression*, vol. 7, no. 3, pp. 215–230, 1997.

[18] R. T. Simpson, "Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones," *Biochemistry*, vol. 17, no. 25, pp. 5524–5531, 1978.

[19] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[20] D. J. Tremethick, "Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber," *Cell*, vol. 128, no. 4, pp. 651–654, 2007.

[21] G. Ozer, A. Luque, and T. Schlick, "The chromatin fiber: Multiscale problems and approaches," *Current Opinion in Structural Biology*, vol. 31, no. 1, pp. 124–139, 2015.

[22] E. Fussner, R. W. Ching, and D. P. Bazett-Jones, "Living without 30nm chromatin fibers," *Trends in Biochemical Sciences*, vol. 36, no. 1, pp. 1–6, 2011.

[23] Y. Joti, T. Hikima, Y. Nishino, F. Kamda, S. Hihara, H. Takata, T. Ishikawa, and K. Maeshima, "Chromosomes without a 30-nm chromatin fiber," *Nucleus*, vol. 3, no. 5, pp. 1–7, 2012.

[24] P. J. J. Robinson, L. Fairall, V. A. T. Huynh, and D. Rhodes, "EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure.," *Proceedings of the National Academy of Sciences*, vol. 103, no. 17, pp. 6506–6511, 2006.

[25] K. van Holde and J. Zlatanova, "Chromatin Higher Order Structure: Chasing a Mirage?," *Journal of Biological Chemistry*, vol. 270, no. 15, pp. 8373–8376, 1995.

[26] D. Z. Staynov, "Possible nucleosome arrangements in the higher-order structure of chromatin," *International Journal of Biological Macromolecules*, vol. 5, no. 1, pp. 3–9, 1983.

[27] J. T. Finch and A. Klug, "Solenoidal model for superstructure in chromatin.," *Proceedings of the National Academy of Sciences*, vol. 73, no. 6, pp. 1897–1901, 1976.

[28] G. Li and D. Reinberg, "Chromatin higher-order structures and gene regulation," *Current Opinion in Genetics and Development*, vol. 21, no. 2, pp. 175–186, 2011.

[29] S. Williams, B. Athey, L. Muglia, R. Schappe, A. Gough, and J. Langmore, "Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length," *Biophysical Journal*, vol. 49, no. 1, pp. 233–248, 1986.

[30] B. Dorigo, T. Schalh, A. Kulangara, S. Duda, R. Schroeder, and T. Richmond, "Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber," *Science*, vol. 306, no. 5701, pp. 1571–1573, 2004.

[31] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond, "X-ray structure of a tetranucleosome and its implications for the chromatin fibre.," *Nature*, vol. 436, no. 7047, pp. 138–141, 2005.

[32] J. R. Daban, "Physical constraints in the condensation of eukaryotic chromosomes. Local concentration of DNA versus linear packing ratio in higher order chromatin structures," *Biochemistry*, vol. 39, no. 14, pp. 3861–3866, 2000.

[33] P. J. Robinson and D. Rhodes, "Structure of the '30 nm' chromatin fibre: A key role for the linker histone," *Current Opinion in Structural Biology*, vol. 16, no. 3, pp. 336–343, 2006.

[34] J. Widom and A. Klug, "Structure of the 300A chromatin filament: X-ray diffraction from oriented samples," *Cell*, vol. 43, no. 1, pp. 207–213, 1985.

[35] O. Perišić, R. Collepardo-Guevara, and T. Schlick, "Modeling studies of chromatin fiber structure as a function of DNA linker length," *Journal of Molecular Biology*, vol. 403, no. 5, pp. 777–802, 2010.

[36] S. Baldi, P. Korber, and P. B. Becker, "Beads on a string—nucleosome array arrangements and folding of the chromatin fiber," *Nature Structural and Molecular Biology*, vol. 27, no. 2, pp. 109–118, 2020.

[37] K. Maeshima, S. Ide, and M. Babokhov, "Dynamic chromatin organization without the 30-nm fiber," *Current Opinion in Cell Biology*, vol. 58, pp. 95–104, 2019.

[38] A. Shakya, S. Park, N. Rana, and J. T. King, "Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization," *Biophysical Journal*, vol. 118, no. 3, pp. 753–764, 2020.

[39] O. L. Kantidze and S. V. Razin, "Weak interactions in higher-order chromatin organization," *Nucleic acids research*, vol. 48, no. 9, pp. 4614–4626, 2020.

[40] M. Mir, W. Bickmore, E. E. Furlong, and G. Narlikar, "Chromatin topology, condensates and gene regulation: Shifting paradigms or just a phase?," *Development*, vol. 146, no. 19, pp. 1–6, 2019.

[41] A. G. Larson and G. J. Narlikar, "The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation," *Biochemistry*, vol. 57, no. 17, pp. 2540–2548, 2018.

[42] J. Zhou, J. Y. Fan, D. Rangasamy, and D. J. Tremethick, "The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression," *Nature Structural and Molecular Biology*, vol. 14, no. 11, pp. 1070–1076, 2007.

[43] G. J. Narlikar, H. Y. Fan, and R. E. Kingston, "Cooperation between complexes that regulate chromatin structure and transcription," *Cell*, vol. 108, no. 4, pp. 475–487, 2002.

[44] T. Phillips and K. Shaw, "Chromatin Remodelin in Eukaryotes," *Nature Education*, vol. 1, no. 1, pp. 209–210, 2008.

[45] S. T. Kosak and M. Groudine, "Gene order and dynamic domains," *Science*, vol. 306, no. 5696, pp. 644–647, 2004.

[46] A. M. Oudelaar, J. O. Davies, L. L. Hanssen, J. M. Telenius, R. Schwessinger, Y. Liu, J. M. Brown, D. J. Downes, A. M. Chiariello, S. Bianco, M. Nicodemi, V. J. Buckle, J. Dekker, D. R. Higgs, and J. R. Hughes, "Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains," *Nature Genetics*, vol. 50, no. 12, pp. 1744–1751, 2018.

[47] B. Li, M. Carey, and J. L. Workman, "The Role of Chromatin during Transcription," *Cell*, vol. 128, no. 4, pp. 707–719, 2007.

[48] J. Boule, J. Mozziconacci, and C. Lavelle, "The polymorphisms of the chromatin fiber," *Journal of Physics: Condensed Matter* , vol. 27, pp. 1–14, 2015.

[49] J. I. Wu, J. Lessard, and G. R. Crabtree, "Understanding the Words of Chromatin Regulation," *Cell*, vol. 136, no. 2, pp. 200–206, 2009.

[50] S. L. Berger, "The complex language of chromatin regulation during transcription," *Nature*, vol. 447, no. 7143, pp. 407–412, 2007.

[51] C. Jiang and B. F. Pugh, "Nucleosome positioning and gene regulation: Advances through genomics," *Nature Reviews Genetics*, vol. 10, no. 3, pp. 161–172, 2009.

[52] C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer, "Dynamic genome architecture in the nuclear space: Regulation of gene expression in three dimensions," *Nature Reviews Genetics*, vol. 8, no. 2, pp. 104–115, 2007.

[53] B. Bonev and G. Cavalli, "Organization and function of the 3D genome," *Nature Reviews Genetics*, vol. 17, no. 11, pp. 661–678, 2016.

[54] T. I. Lee and R. A. Young, "Transcriptional regulation and its misregulation in disease," *Cell*, vol. 152, no. 6, pp. 1237–1251, 2013.

[55] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, no. 5532, pp. 1074–1080, 2001.

[56] M. Radman-Livaja and O. J. Rando, "Nucleosome positioning: How is it established, and why does it matter?," *Developmental Biology*, vol. 339, no. 2, pp. 258–266, 2010.

[57] V. B. Teif and K. Rippe, "Predicting nucleosome positions on the DNA: Combining intrinsic sequence preferences and remodeler activities," *Nucleic Acids Research*, vol. 37, no. 17, pp. 5641–5655, 2009.

[58] L. Ho and G. R. Crabtree, "Chromatin remodelling during development," *Nature*, vol. 463, no. 7280, pp. 474–484, 2010.

[59] G. J. Narlikar, R. Sundaramoorthy, and T. Owen-Hughes, "Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes," *Cell*, vol. 154, no. 3, pp. 490–503, 2013.

[60] D. C. Hargreaves and G. R. Crabtree, "ATP-dependent chromatin remodeling: Genetics, genomics and mechanisms," *Cell Research*, vol. 21, no. 3, pp. 396–420, 2011.

[61] V. K. Gangaraju and B. Bartholomew, "Mechanisms of ATP dependent chromatin remodeling," *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 618, no. 1-2, pp. 3–17, 2007.

[62] C. R. Clapier, J. Iwasa, B. R. Cairns, and C. L. Peterson, "Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 7, pp. 407–422, 2017.

[63] M. Vignali, A. H. Hassan, K. E. Neely, and J. L. Workman, "ATP-Dependent Chromatin-Remodeling Complexes," *Molecular and Cellular Biology*, vol. 20, no. 6, pp. 1899–1910, 2000.

[64] C. B. Gerhold and S. M. Gasser, "INO80 and SWR complexes: Relating structure to function in chromatin remodeling," *Trends in Cell Biology*, vol. 24, no. 11, pp. 619–631, 2014.

[65] A. Saha, J. Wittmeyer, and B. R. Cairns, "Chromatin remodelling: The industrial revolution of DNA around histones," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 6, pp. 437–447, 2006.

[66] W. Chen, Q. Zhu, Y. Liu, and Q. Zhang, "Chromatin Remodeling and Plant Immunity," *Advances in Protein Chemistry and Structural Biology*, vol. 106, pp. 243–260, jan 2017.

[67] C. B. Gerhold, M. H. Hauer, and S. M. Gasser, "INO80-C and SWR-C: Guardians of the genome," *Journal of Molecular Biology*, vol. 427, no. 3, pp. 637–651, 2015.

[68] A. Portela and M. Esteller, "Epigenetic modifications and human disease," *Nature Biotechnology*, vol. 28, no. 10, pp. 1057–1068, 2010.

[69] B. D. Strahl and C. D. Allis, "The language of covalent histone modifications," *Nature*, vol. 403, no. 6765, pp. 41–45, 2000.

[70] T. Kouzarides, "Chromatin Modifications and Their Function," *Cell*, vol. 128, no. 4, pp. 693–705, 2007.

[71] E. S. Witze, W. M. Old, K. A. Resing, and N. G. Ahn, "Mapping protein post-translational modifications with mass spectrometry," *Nature methods*, vol. 4, no. 10, pp. 798–806, 2007.

[72] M. Tan, H. Luo, S. Lee, F. Jin, J. S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocka, Y. Ye, S. Khochbin, B. Ren, and Y. Zhao, "Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification," *Cell*, vol. 146, no. 6, pp. 1016–1028, 2011.

[73] T. Jiang, X. Zhou, K. Taghizadeh, M. Dong, and P. C. Dedon, "N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage.," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 60–65, 2007.

[74] E. L. Mersfelder and M. R. Parthun, "The tale beyond the tail: Histone core domain modifications and the regulation of chromatin structure," *Nucleic Acids Research*, vol. 34, no. 9, pp. 2653–2662, 2006.

[75] L. Zhang, E. E. Eugeni, M. R. Parthun, and M. A. Freitas, "Identification of novel histone post-translational modifications by peptide mass fingerprinting," *Chromosoma*, vol. 112, no. 2, pp. 77–86, 2003.

[76] M. A. Christophorou, G. Castelo-Branco, R. P. Halley-Stott, C. S. Oliveira, R. Loos, A. Radzisheuskaya, K. A. Mowen, P. Bertone, J. C. R. Silva, M. Zernicka-Goetz, M. L. Nielsen, J. B. Gurdon, and T. Kouzarides, "Citrullination regulates pluripotency and histone H1 binding to chromatin.," *Nature*, vol. 507, no. 7490, pp. 104–108, 2014.

[77] Y. Wang, M. Li, S. Stadler, S. Correll, P. Li, D. Wang, R. Hayama, L. Leonelli, H. Han, S. A. Grigoryev, C. D. Allis, and S. A. Coonrod, "Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation," *Journal of Cell Biology*, vol. 184, no. 2, pp. 205–213, 2009.

[78] ChemAxon, "MarvinSketch v6.1.0."

[79] G. D. Bowman and M. G. Poirier, "Post-translational modifications of histones that influence nucleosome dynamics," *Chemical Reviews*, vol. 115, no. 6, pp. 2274–2295, 2015.

[80] P. Tessarz and T. Kouzarides, "Histone core modifications regulating nucleosome structure and dynamics," *Nature Reviews Molecular Cell Biology*, vol. 15, no. 11, pp. 703–708, 2014.

[81] J. A. North, S. Javaid, M. B. Ferdinand, N. Chatterjee, J. W. Picking, M. Shoffner, R. J. Nakkula, B. Bartholomew, J. J. Ottesen, R. Fishel, and M. G. Poirier, "Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling," *Nucleic Acids Research*, vol. 39, no. 15, pp. 6465–6474, 2011.

[82] J. A. North, M. Šimon, M. B. Ferdinand, M. A. Shoffner, J. W. Picking, C. J. Howard, A. M. Mooney, J. Van Noort, M. G. Poirier, and J. J. Ottesen, "Histone H3 phosphorylation near the nucleosome dyad alters chromatin structure," *Nucleic Acids Research*, vol. 42, no. 8, pp. 4922–4933, 2014.

[83] G. Euskirchen, R. K. Auerbach, and M. Snyder, "SWI/SNF chromatin-remodeling factors: Multiscale analyses and diverse functions," *Journal of Biological Chemistry*, vol. 287, no. 37, pp. 30897–30905, 2012.

[84] A. J. Ruthenburg, H. Li, T. A. Milne, S. Dewell, R. K. McGinty, M. Yuen, B. Ueberheide, Y. Dou, T. W. Muir, D. J. Patel, and C. D. Allis, "Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions," *Cell*, vol. 145, no. 5, pp. 692–706, 2011.

[85] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life," *Journal of Molecular Biology*, vol. 337, no. 3, pp. 635–645, 2004.

[86] S. Longhi, "Structural disorder within paramyxoviral nucleoproteins," *FEBS Letters*, vol. 589, no. 19, pp. 2649–2659, 2015.

[87] S. Longhi, L. M. Bloyet, S. Gianni, and D. Gerlier, "How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication," *Cellular and Molecular Life Sciences*, vol. 74, no. 17, pp. 3091–3118, 2017.

[88] A. Kirilyuk, M. Shimoji, J. Catania, G. Sahu, N. Pattabiraman, A. Giordano, C. Albanese, I. Mocchetti, J. A. Toretsky, V. N. Uversky, and M. L. Avantaggiati, "An Intrinsically Disordered Region of the Acetyltransferase p300 with Similarity to Prion-Like Domains Plays a Role in Aggregation," *PLoS ONE*, vol. 7, no. 11, 2012.

[89] B. Shaffer, S. McGraw, S. C. Xiao, D. Chan, J. Trasler, and J. R. Chaillet, "The Dnmt1 intrinsically disordered domain regulates genomic methylation during development," *Genetics*, vol. 199, no. 2, pp. 533–541, 2015.

[90] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder in transcription factors," *Biochemistry*, vol. 45, no. 22, pp. 6873–6888, 2006.

[91] L. Staby, C. O'Shea, M. Willemoës, F. Theisen, B. B. Kragelund, and K. Skriver, "Eukaryotic transcription factors: Paradigms of protein intrinsic disorder," *Biochemical Journal*, vol. 474, no. 15, pp. 2509–2532, 2017.

[92] A. K. Dunker, C. J. Oldfield, J. Meng, P. Romero, J. Y. Yang, J. W. Chen, V. Vacic, Z. Obradovic, and V. N. Uversky, "The unfoldomics decade: An update on intrinsically disordered proteins," *BMC Genomics*, vol. 9, no. 2, pp. 1–26, 2008.

[93] V. Vacic, V. N. Uversky, A. K. Dunker, and S. Lonardi, "Composition Profiler: A tool for discovery and visualization of amino acid composition differences," *BMC Bioinformatics*, vol. 8, no. 211, pp. 1–24, 2007.

[94] A. Campen, R. Williams, C. Brown, J. Meng, V. Uversky, and A. Dunker, "TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder," *Protein & Peptide Letters*, vol. 15, no. 9, pp. 956–963, 2008.

[95] C. M. Baker and R. B. Best, "Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 3, pp. 182–198, 2014.

[96] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence complexity of disordered protein," *Proteins: Structure, Function and Genetics*, vol. 42, no. 1, pp. 38–48, 2001.

[97] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi, "Molecular interactions underlying liquidliquid phase separation of the FUS low-complexity domain," *Nature Structural and Molecular Biology*, vol. 26, no. 7, pp. 637–648, 2019.

[98] S. Das, U. Pal, S. Das, K. Bagga, A. Roy, A. Mrigwani, and N. C. Maiti, "Sequence complexity of amyloidogenic regions in intrinsically disordered human proteins," *PLoS ONE*, vol. 9, no. 3, 2014.

[99] J. Jorda, B. Xue, V. N. Uversky, and A. V. Kajava, "Protein tandem repeats - The more perfect, the less structured," *FEBS Journal*, vol. 277, no. 12, pp. 2673–2682, 2010.

[100] V. N. Uversky, "Intrinsically disordered proteins and their "Mysterious" (meta)physics," *Frontiers in Physics*, vol. 7, no. 10, pp. 1–18, 2019.

[101] V. N. Uversky, "Unusual biophysics of intrinsically disordered proteins," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1834, no. 5, pp. 932–951, 2013.

[102] V. N. Uversky, "Paradoxes and wonders of intrinsic disorder: Stability of instability," *Intrinsically Disordered Proteins*, vol. 5, no. 1, p. e1327757, 2017.

[103] M. Arai, K. Sugase, H. J. Dyson, and P. E. Wright, "Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding," *Proceedings of the National Academy of Sciences*, vol. 112, no. 31, pp. 9614–9619, 2015.

[104] M. Kjaergaard, L. Andersen, L. D. Nielsen, and K. Teilum, "A folded excited state of ligand-free nuclear coactivator binding domain (NCBD) underlies plasticity in ligand recognition," *Biochemistry*, vol. 52, no. 10, pp. 1686–1693, 2013.

[105] I. Radhakrishnan, G. C. Pérez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright, "Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions," *Cell*, vol. 91, no. 6, pp. 741–752, 1997.

[106] P. S. Brzovic, C. C. Heikaus, L. Kisselev, R. Vernon, E. Herbig, D. Pacheco, L. Warfield, P. Littlefield, D. Baker, R. E. Klevit, and S. Hahn, "The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex," *Molecular Cell*, vol. 44, no. 6, pp. 942–953, 2011.

[107] S. L. Shammas, "Mechanistic roles of protein disorder within transcription," *Current Opinion in Structural Biology*, vol. 42, pp. 155–161, 2017.

[108] R. V. Mannige, "Dynamic new world: Refining our view of protein structure, function and evolution," *Proteomes*, vol. 2, no. 1, pp. 128–153, 2014.

[109] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Manipulating Proteins, DNA, and RNA," in *Molecular Biology of the Cell*, New York: Garland Science, 4 ed., 2002.

[110] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 8868–8873, 2000.

[111] Y. Huang and Z. Liu, "Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism," *Journal of Molecular Biology*, vol. 393, no. 5, pp. 1143–1159, 2009.

[112] K. Gunasekaran, C. J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov, "Extended disordered proteins: Targeting function with less scaffold," *Trends in Biochemical Sciences*, vol. 28, no. 2, pp. 81–85, 2003.

[113] D. Vuzman and Y. Levy, "DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail," *Proceedings of the National Academy of Sciences*, vol. 107, no. 49, pp. 21004–21009, 2010.

[114] D. Vuzman, A. Azia, and Y. Levy, "Searching DNA via a "Monkey Bar" Mechanism: The Significance of Disordered Tails," *Journal of Molecular Biology*, vol. 396, no. 3, pp. 674–684, 2010.

[115] J. H. Laity, H. J. Dyson, and P. E. Wright, "DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers.," *Journal of molecular biology*, vol. 295, pp. 719–727, jan 2000.

[116] J. Chen, "Towards the physical basis of how intrinsic disorder mediates protein function," *Archives of Biochemistry and Biophysics*, vol. 524, no. 2, pp. 123–131, 2012.

[117] H. J. Dyson, "Roles of intrinsic disorder in protein-nucleic acid interactions," *Molecular BioSystems*, vol. 8, no. 1, pp. 97–104, 2012.

[118] P. Zhang, K. Torres, X. Liu, C.-g. Liu, and R. E. Pollock, "An Overview of Chromatin-Regulating Proteins in Cells," *Current Protein & Peptide Science*, vol. 17, no. 5, pp. 401–410, 2016.

[119] J. Gsponer and M. Madan Babu, "The rules of disorder or why disorder rules," *Progress in Biophysics and Molecular Biology*, vol. 99, no. 2-3, pp. 94–103, 2009.

[120] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky, "Flexible nets: The roles of intrinsic disorder in protein interaction networks," *FEBS Journal*, vol. 272, no. 20, pp. 5129–5148, 2005.

[121] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Qu, V. N. Uversky, and A. K. Dunker, "Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners," *BMC Genomics*, vol. 9, no. 1, pp. 1–10, 2008.

[122] A. A. Kalashnikova, R. A. Rogge, and J. C. Hansen, "Linker histone H1 and protein-protein interactions," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1859, no. 3, pp. 455–461, 2016.

[123] A. L. Turner, M. Watson, O. G. Wilkins, L. Cato, A. Travers, J. O. Thomas, and K. Stott, "Highly disordered histone H1DNA model complexes and their condensates," *Proceedings of the National Academy of Sciences*, vol. 115, no. 47, pp. 11964–11969, 2018.

[124] M. J. Hendzel, M. A. Lever, E. Crawford, and J. P. H. Th'Ng, "The C-terminal Domain Is the Primary Determinant of Histone H1 Binding to Chromatin in Vivo," *Journal of Biological Chemistry*, vol. 279, no. 19, pp. 20028–20034, 2004.

[125] A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, and B. Schuler, "Extreme disorder in an ultrahigh-affinity protein complex," *Nature*, vol. 555, no. 7694, pp. 61–66, 2018.

[126] P. Heidarsson, D. Mercadante, A. Sottini, D. Nettels, M. Borgia, A. Borgia, S. Kilic, B. Fierz, R. Best, and B. Schuler, "Disordered Proteins Enable Histone Chaperoning on the Nucleosome," *bioRxiv*, 2020.

[127] S. Daujat, U. Zeissler, T. Waldmann, N. Happel, and R. Schneider, "HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding," *Journal of Biological Chemistry*, vol. 280, no. 45, pp. 38090–38095, 2005.

[128] Z. Liu and Y. Huang, "Advantages of proteins being disordered," *Protein Science*, vol. 23, no. 5, pp. 539–550, 2014.

[129] T. Vreven, H. Hwang, B. G. Pierce, and Z. Weng, "Prediction of protein-protein binding free energies," *Protein Science*, vol. 21, no. 3, pp. 396–404, 2012.

[130] I. Drobnak, N. De Jonge, S. Haesaerts, G. Vesnaver, R. Loris, and J. Lah, "Energetic basis of uncoupling folding from binding for an intrinsically disordered protein," *Journal of the American Chemical Society*, vol. 135, no. 4, pp. 1288–1294, 2013.

[131] L. Dahal, S. L. Shammas, and J. Clarke, "Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex," *Biophysical Journal*, vol. 113, no. 12, pp. 2706–2712, 2017.

[132] S. L. Shammas, M. D. Crabtree, L. Dahal, B. I. Wicky, and J. Clarke, "Insights into coupled folding and binding mechanisms from kinetic studies," *Journal of Biological Chemistry*, vol. 291, no. 13, pp. 6689–6695, 2016.

[133] D. H. Kim and K. H. Han, "Transient secondary structures as general target-binding motifs in intrinsically disordered proteins," *International Journal of Molecular Sciences*, vol. 19, no. 11, 2018.

[134] S. Bhattacharya and X. Lin, "Recent advances in computational protocols addressing intrinsically disordered proteins," *Biomolecules*, vol. 9, no. 4, pp. 1–22, 2019.

[135] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 1, pp. 18–29, 2015.

[136] P. Tompa, N. E. Davey, T. J. Gibson, and M. M. Babu, "A Million peptide motifs for the molecular biologist," *Molecular Cell*, vol. 55, no. 2, pp. 161–169, 2014.

[137] M. Fuxreiter, "Fold or not to fold upon binding — does it really matter?," *Current Opinion in Structural Biology*, vol. 54, pp. 19–25, 2019.

[138] H. J. Dyson and P. E. Wright, "Coupling of folding and binding for unstructured proteins," *Current Opinion in Structural Biology*, vol. 12, no. 1, pp. 54–60, 2002.

[139] P. E. Wright and H. J. Dyson, "Linking folding and binding," *Current Opinion in Structural Biology*, vol. 19, no. 1, pp. 31–38, 2009.

[140] S. J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H. Jane Dyson, R. M. Evans, and P. E. Wright, "Mutual synergistic folding in recruitment of cbp/p300 by p160 nuclear receptor coactivators," *Nature*, vol. 415, no. 6871, pp. 549–553, 2002.

[141] S. Brüschweiler, R. Konrat, and M. Tollinger, "Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core," *ACS Chemical Biology*, vol. 8, no. 7, pp. 1600–1610, 2013.

[142] K. L. Seldeen, C. B. McDonald, B. J. Deegan, and A. Farooq, "Evidence that the bZIP domains of the Jun transcription factor bind to DNA as monomers prior to folding and homodimerization," *Archives of Biochemistry and Biophysics*, vol. 480, no. 2, pp. 75–84, 2008.

[143] A. Srivastava, S. Ahmad, and M. M. Gromiha, "Deciphering RNA-recognition patterns of intrinsically disordered proteins," *International Journal of Molecular Sciences*, vol. 19, no. 6, 2018.

[144] S. A. Jacobs and S. Khorasanizadeh, "Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail," *Science*, vol. 295, no. 5562, pp. 2080–2083, 2002.

[145] L. M. Espinoza-Fonseca, "Reconciling binding mechanisms of intrinsically disordered proteins," *Biochemical and Biophysical Research Communications*, vol. 382, no. 3, pp. 479–482, 2009.

[146] J. Yang, M. Gao, J. Xiong, Z. Su, and Y. Huang, "Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding," *Protein Science*, vol. 28, no. 11, pp. 1952–1965, 2019.

[147] S. Chakrabortee, F. Meersman, G. S. Kaminski Schierle, C. W. Bertoncini, B. McGee, C. F. Kaminski, and A. Tunnacliffe, "Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance," *Proceedings of the National Academy of Sciences*, vol. 107, no. 37, pp. 16084–16089, 2010.

[148] T. Mittag, S. Orlicky, W. Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers, and J. D. Forman-Kay, "Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor," *Proceedings of the National Academy of Sciences*, vol. 105, no. 46, pp. 17772–17777, 2008.

[149] A. B. Sigalov, A. V. Zhuravleva, and V. Y. Orekhov, "Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form," *Biochimie*, vol. 89, no. 3, pp. 419–421, 2007.

[150] B. Sharma and S. Paul, "Understanding the Role of Temperature Change and the Presence of NaCl Salts on Caffeine Aggregation in Aqueous Solution: From Structural and Thermodynamics Point of View.," *Journal of Physical Chemistry B*, vol. 119, no. 21, pp. 6421–32, 2015.

[151] P. Tompa and M. Fuxreiter, "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions," *Trends in Biochemical Sciences*, vol. 33, no. 1, pp. 2–8, 2008.

[152] M. Fuxreiter, "Fuzziness in Protein Interactions—A Historical Perspective," *Journal of Molecular Biology*, vol. 430, no. 16, pp. 2278–2287, 2018.

[153] M. Fuxreiter and P. Tompa, "Fuzzy complexes: A more stochastic view of protein function," *Advances in Experimental Medicine and Biology*, vol. 725, pp. 1–14, 2012.

[154] V. Pejaver, W. L. Hsu, F. Xin, A. K. Dunker, V. N. Uversky, and P. Radivojac, "The structural and functional signatures of proteins that undergo multiple events of post-translational modification," *Protein Science*, vol. 23, no. 8, pp. 1077–1093, 2014.

[155] A. Bah and J. D. Forman-Kay, "Modulation of intrinsically disordered protein function by post-translational modifications," *Journal of Biological Chemistry*, vol. 291, no. 13, pp. 6696–6705, 2016.

[156] D. Winogradoff, I. Echeverria, D. A. Potoyan, and G. A. Papoian, "The acetylation landscape of the H4 histone tail: Disentangling the interplay between the specific and cumulative effects," *Journal of the American Chemical Society*, vol. 137, no. 19, pp. 6245–6253, 2015.

[157] L. Ou, A. M. Ferreira, S. Otieno, L. Xiao, D. Bashford, and R. W. Kriwacki, "Incomplete folding upon binding mediates Cdk4/cyclin D complex activation by tyrosine phosphorylation of inhibitor p27 protein," *Journal of Biological Chemistry*, vol. 286, no. 34, pp. 30142–30151, 2011.

[158] M. K. Yoon, D. M. Mitrea, L. Ou, and R. W. Kriwacki, "Cell cycle regulation by the intrinsically disordered proteins p21 and p27," *Biochemical Society Transactions*, vol. 40, no. 5, pp. 981–988, 2012.

[159] J. Ruan, H. Ouyang, M. F. Amaya, M. Ravichandran, P. Loppnau, J. Min, and J. Zang, "Structural basis of the chromodomain of Cbx3 bound to methylated peptides from histone H1 and G9a," *PLoS ONE*, vol. 7, no. 4, 2012.

[160] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*. San Diego: Academic Press, 2 ed., 2001.

[161] J. W. Ponder and D. A. Case, "Force fields for protein simulation," in *Advances in Protein Chemistry* (V. Daggett, ed.), pp. 27–85, Seattle: Academic Press, 2003.

[162] W. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.

[163] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. Van Gunsteren, "The GROMOS software for biomolecular simulation: GROMOS05," 2005.

[164] A. D. J. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins.," *Journal of Physical Chemistry B*, vol. 102, no. 18, pp. 3586–616, 1998.

[165] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, J. K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman., "A second generation force field for the simulation of proteins, nucleic acids and organic molecules.," *J. Am. Chem. Soc.*, vol. 117, no. 6, pp. 5179–5197, 1995.

[166] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins: Structure, Function and Bioinformatics*, vol. 78, no. 8, pp. 1950–1958, 2010.

[167] R. B. Best, W. Zheng, and J. Mittal, "Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association," *Journal of Chemical Theory and Computation*, vol. 10, no. 11, pp. 5113–5124, 2014.

[168] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, no. 1, pp. 71–73, 2016.

[169] S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. De Groot, and H. Grubmüller, "Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment," *Journal of Chemical Theory and Computation*, vol. 11, no. 11, pp. 5513–5524, 2015.

[170] P. Lagant, D. Nolde, and G. Vergoten, "Inclusion of a modified Urey-Bradley type force field in the Charmm program," in *Spectroscopy of Biological Molecules: New Directions*, pp. 409–410, 1999.

[171] A. R. Leach, *Molecular modeling: principles and applications*. Pearson Education, 2001.

[172] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[173] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The Missing Term in Effective Pair Potentials," *Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[174] W. L. Jorgensen and J. D. Madura, "Temperature and size dependence for monte carlo simulations of TIP4P water," *Molecular Physics*, vol. 56, no. 6, pp. 1381–1392, 1985.

[175] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building water models: A different approach," *Journal of Physical Chemistry Letters*, vol. 5, no. 21, pp. 3863–3871, 2014.

[176] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *Journal of Chemical Physics*, vol. 112, no. 20, pp. 8910–8922, 2000.

[177] J. L. Abascal and C. Vega, "A general purpose model for the condensed phases of water: TIP4P/2005," *Journal of Chemical Physics*, vol. 123, no. 23, 2005.

[178] R. W. Hockney, S. P. Goel, and J. W. Eastwood, "Quiet high-resolution computer models of a plasma," *Journal of Computational Physics*, vol. 14, no. 2, pp. 148–158, 1974.

[179] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: A Linear Constraint Solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[180] B. Hess, "P-LINCS: A parallel linear constraint solver for molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 1, pp. 116–122, 2008.

[181] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952–962, 1992.

[182] M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, and E. Lindah, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, no. 2, pp. 19–25, 2015.

[183] C. L. Wennberg, T. Murtola, S. Páll, M. J. Abraham, B. Hess, and E. Lindahl, "Direct-Space Corrections Enable Fast and Accurate Lorentz-Berthelot Combination Rule Lennard-Jones Lattice Summation," *Journal of Chemical Theory and Computation*, vol. 11, no. 12, pp. 5737–5746, 2015.

[184] J. W. Eastwood, R. W. Hockney, and D. N. Lawrence, "P3M3DP-The three-dimensional periodic particle-particle/ particle-mesh program," *Computer Physics Communications*, vol. 19, no. 2, pp. 215–261, 1980.

[185] M. Bergdorf, C. Peter, and P. H. Hünenberger, "Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study," *Journal of Chemical Physics*, vol. 119, no. 17, pp. 9129–9144, 2003.

[186] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, no. 12, p. 10089, 1993.

[187] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[188] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.

[189] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *Journal of Chemical Physics*, vol. 81, no. 8, p. 3684, 1984.

[190] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *Journal of Chemical Physics*, vol. 126, no. 1, 2007.

[191] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, and S. H. Scheres, "New tools for automated high-resolution cryo-EM structure determination in RELION-3," *eLife*, vol. 7, 2018.

[192] U. Shmueli, *Theories and techniques of crystal structure determination*. Oxford University Press, 2007.

[193] J. M. Würz, S. Kazemi, E. Schmidt, A. Bagaria, and P. Güntert, "NMR-based automated protein structure determination," *Archives of Biochemistry and Biophysics*, vol. 628, pp. 24–32, 2017.

[194] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.

[195] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. MacKenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, "Millisecond-scale molecular dynamics simulations on Anton," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pp. 1–11, 2009.

[196] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. 2017.

[197] R. H. Swendsen and J. S. Wang, "Replica Monte Carlo simulation of spin-glasses," *Physical Review Letters*, vol. 57, no. 21, pp. 2607–2609, 1986.

[198] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, "Replica exchange with solute tempering: A method for sampling biological systems in explicit water," *Proceedings of the National Academy of Sciences*, vol. 102, no. 39, pp. 13749–13754, 2005.

[199] B. Roux, "The calculation of the potential of mean force using computer simulations," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 275–282, 1995.

[200] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Physical Review Letters*, vol. 100, no. 2, 2008.

[201] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," 2005.

[202] L. Wang, R. A. Friesner, and B. J. Berne, "Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2)," *Journal of Physical Chemistry B*, vol. 115, no. 30, pp. 9431–9438, 2011.

[203] G. Bussi, "Hamiltonian replica exchange in GROMACS: A flexible implementation," *Molecular Physics*, vol. 112, no. 3-4, pp. 379–384, 2014.

[204] J. Kästner, "Umbrella sampling," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 6, pp. 932–942, 2011.

[205] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 1011–1021, 1992.

[206] M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations," *Computer Physics Communications*, vol. 135, no. 1, pp. 40–57, 2001.

[207] J. Hub, B. de Groot, and D. Van Der Spoel, "g_wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates," *Journal of Chemical Theory and Computation*, vol. 6, no. 12, pp. 3713–3720, 2010.

[208] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12562–12566, 2002.

[209] B. Ensing, M. DeVivo, Z. Liu, P. Moore, and M. Klein, "Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions," *Accounts of Chemical Research*, vol. 39, no. 2, pp. 73–81, 2006.

[210] D. Branduardi, F. L. Gervasio, A. Cavalli, M. Recanatini, and M. Parrinello, "The role of the peripheral anionic site and cation-π interactions in the ligand penetration of the human AChE gorge," *Journal of the American Chemical Society*, vol. 127, no. 25, pp. 9147–9155, 2005.

[211] S. Piana and A. Laio, "A Bias-Exchange Approach to Protein Folding," *The Journal of Physical Chemistry B*, vol. 111, no. 17, pp. 4553–4559, 2007.

[212] M. Bonomi, A. Barducci, and M. Parrinello, "Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics," *Journal of Computational Chemistry*, vol. 30, no. 11, pp. 1615–1621, 2009.

[213] G. Bussi and A. Laio, "Using metadynamics to explore complex free-energy landscapes," *Nature Reviews Physics*, vol. 2, no. 4, pp. 200–212, 2020.

[214] L. Delemotte, M. A. Kasimova, M. L. Klein, M. Tarek, and V. Carnevale, "Free-energy landscape of ion-channel voltage-sensor–domain activation," *Proceedings of the National Academy of Sciences*, vol. 112, no. 1, pp. 124–129, 2015.

[215] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, "Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics," *Journal of the American Chemical Society*, vol. 128, no. 41, pp. 13435–13441, 2006.

[216] M. Bonomi and M. Parrinello, "Enhanced sampling in the well-tempered ensemble," *Physical Review Letters*, vol. 104, no. 19, 2010.

[217] Y. C. Chen, "Beware of docking!," *Trends in Pharmacological Sciences*, vol. 36, no. 2, pp. 78–95, 2015.

[218] N. Brooijmans and I. D. Kuntz, "Molecular recognition and docking algorithms," *Annual Review of Biophysics and Biomolecular Structure*, vol. 32, pp. 335–373, 2003.

[219] G. Jones and P. Willett, "Docking small-molecule ligands into active sites," *Current Opinion in Biotechnology*, vol. 6, no. 6, pp. 652–656, 1995.

[220] B. D. Bursulaya, M. Totrov, R. Abagyan, and C. L. Brooks, "Comparative study of several algorithms for flexible ligand docking," *Journal of Computer-Aided Molecular Design*, vol. 17, no. 11, pp. 755–763, 2003.

[221] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2010.

[222] J. Meiler and D. Baker, "ROSETTALIGAND: Protein–Small Molecule Docking with Full Side-Chain Flexibility," *Proteins: Structure, Function and Bioinformatics*, vol. 65, no. 3, pp. 538–548, 2006.

[223] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking.," *Journal of Molecular Biology*, vol. 267, no. 3, pp. 727–748, 1997.

[224] K. A. Porter, I. Desta, D. Kozakov, and S. Vajda, "What method to use for protein–protein docking?," *Current Opinion in Structural Biology*, vol. 55, pp. 1–7, 2019.

[225] A. Singh, T. Dauzhenka, P. Kundrotas, M. J. Sternberg, and I. Vakser, "Application of Docking to Protein Models," *Biophysical Journal*, vol. 118, no. 3, p. 360a, 2020.

[226] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, "Protein-protein docking dealing with the unknown," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 317–342, 2010.

[227] T. Kirys, A. M. Ruvinsky, D. Singla, A. V. Tuzikov, P. J. Kundrotas, and I. A. Vakser, "Simulated unbound structures for benchmarking of protein docking in the Dockground resource," *BMC Bioinformatics*, vol. 16, no. 1, 2015.

[228] J. Cherfils and J. Janin, "Protein docking algorithms: simulating molecular recognition," *Current Opinion in Structural Biology*, vol. 3, no. 2, pp. 265–269, 1993.

[229] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin, "HADDOCK: A protein-protein docking approach based on biochemical or biophysical information," *Journal of the American Chemical Society*, vol. 125, no. 7, pp. 1731–1737, 2003.

[230] G. C. P. Van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. Van Dijk, S. J. De Vries, and A. M. J. J. Bonvin, "The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes," *Journal of Molecular Biology*, vol. 428, no. 4, pp. 720–725, 2016.

[231] J. Bednar, I. Garcia-Saez, R. Boopathi, A. R. Cutter, G. Papai, A. Reymer, S. H. Syed, I. N. Lone, O. Tonchev, C. Crucifix, H. Menoni, C. Papin, D. A. Skoufias, H. Kurumizaka, R. Lavery, A. Hamiche, J. J. Hayes, P. Schultz, D. Angelov, C. Petosa, and S. Dimitrov, "Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1," *Molecular Cell*, vol. 66, no. 3, pp. 384–397, 2017.

[232] S. H. Syed, D. Goutte-Gattat, N. Becker, S. Meyer, M. S. Shukla, J. J. Hayes, R. Everaers, D. Angelov, J. Bednar, and S. Dimitrov, "Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome," *Proceedings of the National Academy of Sciences*, vol. 107, no. 21, pp. 9620–9625, 2010.

[233] T. L. Caterino, H. Fang, and J. J. Hayes, "Nucleosome Linker DNA Contacts and Induces Specific Folding of the Intrinsically Disordered H1 Carboxyl-Terminal Domain," *Molecular and Cellular Biology*, vol. 31, no. 11, pp. 2341–2348, 2011.

[234] D. V. Fyodorov, B. R. Zhou, A. I. Skoultchi, and Y. Bai, "Emerging roles of linker histones in regulating chromatin structure and function," *Nature Reviews Molecular Cell Biology*, vol. 19, no. 3, pp. 192–206, 2017.

[235] Y. Zhang, M. Cooke, S. Panjwani, K. Cao, B. Krauth, P. Y. Ho, M. Medrzycki, D. T. Berhe, C. Pan, T. C. McDevitt, and Y. Fan, "Histone H1 depletion impairs embryonic stem cell differentiation," *PLoS Genetics*, vol. 8, no. 5, 2012.

[236] F. Thoma, Koller T., and A. Klug, "Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin," *Journal of Cell Biology*, vol. 83, no. 2 I, pp. 403–427, 1979.

[237] X. Shen, L. Yu, J. W. Weir, and M. A. Gorovsky, "Linker histories are not essential and affect chromatin condensation in vivo," *Cell*, vol. 82, no. 1, pp. 47–56, 1995.

[238] Y. Fan, T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock, and A. I. Skoultchi, "Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation," *Cell*, vol. 123, no. 7, pp. 1199–1212, 2005.

[239] T. J. Maresca, B. S. Freedman, and R. Heald, "Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts," *Journal of Cell Biology*, vol. 169, no. 6, pp. 859–869, 2005.

[240] X. Lu, S. N. Wontakal, A. V. Emelyanov, P. Morcillo, A. Y. Konev, D. V. Fyodorov, and A. I. Skoultchi, "Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure," *Genes and Development*, vol. 23, no. 4, pp. 452–465, 2009.

[241] T. Sera and A. P. Wolffe, "Role of histone H1 as an architectural determinant of chromatin structure and as a specific repressor of transcription on Xenopus oocyte 5S rRNA genes.," *Molecular and cellular biology*, vol. 18, no. 7, pp. 3668–3680, 1998.

[242] X. Shen and M. A. Gorovsky, "Linker histone H1 regulates specific gene expression but not global transcription in vivo," *Cell*, vol. 86, no. 3, pp. 475–483, 1996.

[243] C. M. Di Liegro, G. Schiera, and I. Di Liegro, "H1.0 Linker Histone as an Epigenetic Regulator of Cell Proliferation and Differentiation," *Genes*, vol. 9, no. 6, pp. 310–329, 2018.

[244] N. Happel and D. Doenecke, "Histone H1 and its isoforms: Contribution to chromatin structure and function," *Gene*, vol. 431, no. 1-2, pp. 1–12, 2009.

[245] J. Allan, P. G. Hartman, C. Crane-Robinson, and F. X. Aviles, "The structure of histone H1 and its location in chromatin," *Nature*, vol. 288, no. 5792, pp. 675–679, 1980.

[246] P. G. Hartman, G. E. Chapman, T. Moss, and E. M. Bradbury, "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," *European Journal of Biochemistry*, vol. 77, no. 1, pp. 45–51, 1977.

[247] K. L. Clark, E. D. Halay, E. Lai, and S. K. Burley, "Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5," *Nature*, vol. 364, no. 6436, pp. 412–420, 1993.

[248] K. S. Gajiwala and S. K. Burley, "Winged helix proteins," *Current Opinion in Structural Biology*, vol. 10, no. 1, pp. 110–116, 2000.

[249] J. Hutchinson, M. Cheema, J. Wang, K. Missiaen, R. Finn, R. Romero, J. Th'ng, M. Hendzel, and J. Ausio, "Interaction of chromatin with a histone H1 containing swapped N- and C-terminal domains," *Bioscience Reports*, vol. 35, no. 3, 2015.

[250] T. Misteli, A. Gunjan, R. Hock, M. Bustin, and D. T. Brown, "Dynamic binding of histone H1 to chromatin in living cells," *Nature*, vol. 408, no. 6814, pp. 877–881, 2000.

[251] T. J. Stasevich, F. Mueller, D. T. Brown, and J. G. McNally, "Dissecting the binding mechanism of the linker histone in live cells: An integrated FRAP analysis," *EMBO Journal*, vol. 29, no. 7, pp. 1225–1234, 2010.

[252] J. A. Subirana, "Analysis of the charge distribution in the C-terminal region of histone H1 as related to its interaction with DNA," *Biopolymers*, vol. 29, no. 10-11, pp. 1351–1357, 1990.

[253] T. W. Flanagan, J. K. Files, K. R. Casano, E. M. George, and D. T. Brown, "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*, vol. 5, no. 3, pp. 372–380, 2016.

[254] J. Allan, T. Mitchell, N. Harborne, L. Bohm, and C. Crane-Robinson, "Roles of H1 domains in determining higher order chromatin structure and H1 location," *Journal of Molecular Biology*, vol. 187, no. 4, pp. 591–601, 1986.

[255] P. Vyas and D. T. Brown, "N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H10 and H1c," *Journal of Biological Chemistry*, vol. 287, no. 15, pp. 11778–11787, 2012.

[256] M. Okuwaki, M. Abe, M. Hisaoka, and K. Nagata, "Regulation of cellular dynamics and chromosomal binding site preference of linker histones H1.0 and H1.X," *Molecular and Cellular Biology*, vol. 36, no. 21, pp. 2681–2696, 2016.

[257] C. Oberg and S. Belikov, "The N-terminal domain determines the affinity and specificity of H1 binding to chromatin," *Biochemical and Biophysical Research Communications*, vol. 420, no. 2, pp. 321–324, 2012.

[258] A. Kowalski and J. Pałyga, "Modulation of chromatin function through linker histone H1 variants," *Biology of the Cell*, vol. 108, no. 12, pp. 339–356, 2016.

[259] A. Izzo and R. Schneider, "The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1859, no. 3, pp. 486–495, 2016.

[260] T. W. Flanagan and D. Brown, "Molecular dynamics of histone H1," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1859, no. 3, pp. 468–475, 2016.

[261] M. A. Lever, J. P. Th'ng, X. Sun, and M. J. Hendzel, "Rapid exchange of histone H1.1 on chromatin in living human cells," *Nature*, vol. 408, no. 6814, pp. 873–876, 2000.

[262] T. Starkova, A. Polyanichko, T. Artamonova, M. Khodorkovskii, E. Kostyleva, E. Chikhirzhina, and A. Tomilin, "Post-translational modifications of linker histone H1 variants in mammals," *Physical Biology*, vol. 14, no. 16005-16017, 2017.

[263] J. R. Wiśniewski, A. Zougman, S. Krüger, and M. Mann, "Mass Spectrometric Mapping of Linker Histone H1 Variants Reveals Multiple Acetylations, Methylations, and Phosphorylation as Well as Differences between Cell Culture and Tissue," *Molecular & Cellular Proteomics*, vol. 6, no. 1, pp. 72–87, 2007.

[264] M. Kotliński, K. Rutowicz, Ł. Kniżewski, A. Palusiński, J. Olêdzki, A. Fogtman, T. Rubel, M. Koblowska, M. Dadlez, K. Ginalski, and A. Jerzmanowski, "Histone H1 variants in Arabidopsis are subject to numerous post-translational modifications, both conserved and previously unknown in histones, suggesting complex functions of H1 in plants," *PLoS ONE*, vol. 11, no. 1, 2016.

[265] A. Kuzmichev, T. Jenuwein, P. Tempst, and D. Reinberg, "Different Ezh2-containing complexes target methylation of histone H1 or nucleosomal histone H3," *Molecular Cell*, vol. 14, no. 2, pp. 183–193, 2004.

[266] P. Trojer, J. Zhang, M. Yonezawa, A. Schmidt, H. Zheng, T. Jenuwein, and D. Reinberg, "Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the jumonji domain-containing JMJD2/KDM4 proteins," *Journal of Biological Chemistry*, vol. 284, no. 13, pp. 8395–8405, 2009.

[267] S. L. Rulten, A. E. Fisher, I. Robert, M. C. Zuma, M. Rouleau, L. Ju, G. Poirier, B. Reina-San-Martin, and K. W. Caldecott, "PARP-3 and APLF function together to accelerate nonhomologous end-joining," *Molecular Cell*, vol. 41, no. 1, pp. 33–45, 2011.

[268] T. Thorslund, A. Ripplinger, S. Hoffmann, T. Wild, M. Uckelmann, B. Villumsen, T. Narita, T. K. Sixma, C. Choudhary, S. Bekker-Jensen, and N. Mailand, "Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage," *Nature*, vol. 527, no. 7578, pp. 389–393, 2015.

[269] T. Weiss, S. Hergeth, U. Zeissler, A. Izzo, P. Tropberger, B. M. Zee, M. Dundr, B. A. Garcia, S. Daujat, and R. Schneider, "Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D," *Epigenetics and Chromatin*, vol. 3, no. 7, pp. 1–13, 2010.

[270] N. Ogata, K. Ueda, H. Kagamiyama, and O. Hayaishi, "ADP-ribosylation of histone H1. Identification of glutamic acid residues 2, 14, and the COOH-terminal lysine residue as modification sites.," *Journal of Biological Chemistry*, vol. 255, no. 16, pp. 7616–7620, 1980.

[271] J. R. Wiśniewski, A. Zougman, and M. Mann, "Nϵ-Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function," *Nucleic Acids Research*, vol. 36, no. 2, pp. 570–577, 2008.

[272] A. Roque, I. Ponte, and P. Suau, "Post-translational modifications of the intrinsically disordered terminal domains of histone H1: effects on secondary structure and chromatin dynamics," *Chromosoma*, vol. 126, no. 1, pp. 83–91, 2017.

[273] R. A. Swank, J. P. Th'ng, X. W. Guo, J. Valdez, E. M. Bradbury, and L. R. Gurley, "Four distinct cyclin-dependent kinases phosphorylate histone H1 at all of its growth-related phosphorylation sites," *Biochemistry*, vol. 36, no. 45, pp. 13761–13768, 1997.

[274] B. Sarg, W. Helliger, H. Talasz, B. Förg, and H. H. Lindner, "Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: Identification of a novel phosphorylation site on histone H1," *Journal of Biological Chemistry*, vol. 281, no. 10, pp. 6573–6580, 2006.

[275] H. Talasz, W. Helliger, B. Puschendorf, and H. Lindner, "In vivo phosphorylation of histone H1 variants during the cell cycle," *Biochemistry*, vol. 35, no. 6, pp. 1761–1767, 1996.

[276] R. Liao and C. A. Mizzen, "Interphase H1 phosphorylation: Regulation and functions in chromatin," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1859, no. 3, pp. 476–485, 2016.

[277] L. R. Gurley, J. G. Valdez, and S. J. Buchanan, "Characterization of the Mitotic Specific Phosphorylation Site of Histone H1," *Journal of Biological Chemistry*, vol. 270, no. 46, pp. 27653–27660, 1995.

[278] J. P. Th'ng, X. W. Guo, R. A. Swank, H. A. Crissman, and E. M. Bradbury, "Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation," *Journal of Biological Chemistry*, vol. 269, no. 13, pp. 9568–9573, 1994.

[279] A. Roque, I. Ponte, J. L. R. Arrondo, and P. Suau, "Phosphorylation of the carboxy-terminal domain of histone H1: Effects on secondary structure and DNA condensation," *Nucleic Acids Research*, vol. 36, no. 14, pp. 4719–4726, 2008.

[280] R. Lopez, B. Sarg, H. Lindner, S. Bartolomé, I. Ponte, P. Suau, and A. Roque, "Linker histone partial phosphorylation: Effects on secondary structure and chromatin condensation," *Nucleic Acids Research*, vol. 43, no. 9, pp. 4463–4476, 2015.

[281] N. Raghuram, H. Strickfaden, D. McDonald, K. Williams, H. Fang, C. Mizzen, J. J. Hayes, J. Th'ng, and M. J. Hendzel, "Pin1 promotes histone H1 dephosphorylation and stabilizes its binding to chromatin," *Journal of Cell Biology*, vol. 203, no. 1, pp. 57–71, 2013.

[282] G. D. Bascom and T. Schlick, "Chromatin Fiber Folding Directed by Cooperative Histone Tail Acetylation and Linker Histone Binding," *Biophysical Journal*, vol. 114, no. 10, pp. 2376–2385, 2018.

[283] M. M. Duggan and J. O. Thomas, "Two DNA-binding sites on the globular domain of histone H5 are required for binding to both bulk and 5 S reconstituted nucleosomes," *Journal of Molecular Biology*, vol. 304, no. 1, pp. 21–33, 2000.

[284] O. Perišić, S. Portillo-Ledesma, and T. Schlick, "Sensitive effect of linker histone binding mode and subtype on chromatin condensation," *Nucleic acids research*, vol. 47, no. 10, pp. 4948–4957, 2019.

[285] M. Hagop Parseghian, "What is the role of histone H1 heterogeneity? A functional model emerges from a 50 year mystery," *AIMS Biophysics*, vol. 2, no. 4, pp. 724–772, 2015.

[286] D. Z. Staynov and C. Crane-Robinson, "Footprinting of linker histones H5 and H 1 on the nucleosome," *The EMBO Journal*, vol. 7, no. 12, pp. 3685–3691, 1988.

[287] D. T. Brown, T. Izard, and T. Misteli, "Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo.," *Nature Structural & Molecular Biology*, vol. 13, no. 3, pp. 250–255, 2006.

[288] B.-R. Zhou, H. Feng, H. Kato, L. Dai, Y. Yang, Y. Zhou, and Y. Bai, "Structural insights into the histone H1-nucleosome complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 48, pp. 19390–19395, 2013.

[289] D. C. Woods and J. Wereszczynski, "Elucidating the influence of linker histone variants on chromatosome dynamics and energetics," *Nucleic Acids Research*, vol. 48, pp. 3591–3604, mar 2020.

[290] Y. B. Zhou, S. E. Gerchman, V. Ramakrishnan, A. Travers, and S. Muyldermans, "Position and orientation of the globular domain of linker histone H5 on the nucleosome," *Nature*, vol. 395, no. 6700, pp. 402–405, 1998.

[291] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R. M. Xu, P. Zhu, and G. Li, "Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units," *Science*, vol. 344, no. 6182, pp. 376–380, 2014.

[292] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. L. Yeh, "UniProt: The Universal Protein knowledge base," *Nucleic Acids Research*, vol. 32, pp. 115–119, 2004.

[293] Schrödinger LLC, "The PyMOL Molecular Graphics System," 2002.

[294] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the Generalized Born Model Suitable for Macromolecules," *Journal of Physical Chemistry B*, vol. 104, no. 15, pp. 3712–3720, 2000.

[295] A. Onufriev, D. Bashford, and D. A. Case, "Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model," *Proteins: Structure, Function and Genetics*, vol. 55, no. 2, pp. 383–394, 2004.

[296] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro : an advanced semantic chemical editor , visualization , and analysis platform," *Journal of Cheminformatics*, vol. 4, no. 17, pp. 1758–2946, 2012.

[297] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," *Computer Physics Communications*, vol. 185, no. 2, pp. 604–613, 2014.

[298] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," *Computer Physics Communications*, vol. 180, no. 10, pp. 1961–1972, 2009.

[299] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[300] N. Homeyer, A. H. C. Horn, H. Lanig, and H. Sticht, "AMBER force-field parameters for phosphorylated amino acids in different protonation states: Phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine," *Journal of Molecular Modelling*, vol. 12, no. 3, pp. 281–289, 2006.

[301] K. K. Lee, E. Kim, C. Joo, J. Song, H. Han, and M. Cho, "Site-selective intramolecular hydrogen-bonding interactions in phosphorylated serine and threonine dipeptides," *Journal of Physical Chemistry B*, vol. 112, no. 51, pp. 16782–16787, 2008.

[302] F. Pietrucci and A. Laio, "A collective variable for the efficient exploration of protein beta-sheet structures: Application to SH3 and GB1," *Journal of Chemical Theory and Computation*, vol. 5, no. 9, pp. 2197–2201, 2009.

[303] A. Patriksson and D. van der Spoel, "A temperature predictor for parallel tempering simulations," *Physical Chemistry Chemical Physics*, vol. 10, no. 15, pp. 2073–2077, 2008.

[304] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*. John Wiley & Sons, 2011.

[305] H. Fang, S. Wei, T. H. Lee, and J. J. Hayes, "Chromatin structure-dependent conformations of the H1 CTD," *Nucleic Acids Research*, vol. 44, no. 19, pp. 9131–9141, 2016.

[306] A. E. White, A. R. Hieb, and K. Luger, "A quantitative investigation of linker histone interactions with nucleosomes and chromatin," *Scientific Reports*, vol. 6, 2016.

[307] X. Lu and J. C. Hansen, "Identification of Specific Functional Subdomains within the Linker Histone H10 C-terminal Domain," *Journal of Biological Chemistry*, vol. 279, no. 10, pp. 8701–8707, 2004.

[308] X. Lu, B. Hamkalo, M. H. Parseghian, and J. C. Hansen, "Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder," *Biochemistry*, vol. 48, no. 1, pp. 164–172, 2009.

[309] M. M. Bharath, N. R. Chandra, and M. R. Rao, "Prediction of an HMG-box fold in the C-terminal domain of histone H1: Insights into its role in DNA condensation," *Proteins: Structure, Function and Genetics*, vol. 49, no. 1, pp. 71–81, 2002.

[310] H. Fang, D. J. Clark, and J. J. Hayes, "DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain," *Nucleic Acids Research*, vol. 40, no. 4, pp. 1475–1484, 2012.

[311] R. C. Edgar, "MUSCLE: A multiple sequence alignment method with reduced time and space complexity," *BMC Bioinformatics*, vol. 5, 2004.

[312] R. C. Edgar, "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[313] M. Clamp, J. Cuff, S. M. Searle, and G. J. Barton, "The Jalview Java alignment editor," *Bioinformatics*, vol. 20, no. 3, pp. 426–427, 2004.

[314] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton, "Jalview Version 2-A multiple sequence alignment editor and analysis workbench," *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.

[315] D. J. Clark, C. S. Hill, S. R. Martin, and J. O. Thomas, "Alpha-helix in the carboxy-terminal domains of histones H1 and H5," *EMBO J*, vol. 7, no. 1, pp. 69–75, 1988.

[316] C. S. Hill, S. R. Martin, and J. O. Thomas, "A stable alpha-helical element in the carboxy-terminal domain of free and chromatin-bound histone H1 from sea urchin sperm," *EMBO J.*, vol. 8, no. 9, pp. 2591–2599, 1989.

[317] A. Roque, I. Iloro, I. Ponte, J. L. R. Arrondo, and P. Suau, "DNA-induced secondary structure of the carboxyl-terminal domain of histone H1," *Journal of Biological Chemistry*, vol. 280, no. 37, pp. 32141–32147, 2005.

[318] R. Vila, I. Ponte, M. A. Jiménez, M. Rico, and P. Suau, "An inducible helix-Gly-Gly-helix motif in the N-terminal domain of histone H1e: a CD and NMR study.," *Protein Science*, vol. 11, no. 2, pp. 214–220, 2002.

[319] R. Vila, I. Ponte, M. Collado, J. L. R. Arrondo, M. A. Jiménez, M. Rico, and P. Suau, "DNA-induced α-Helical Structure in the NH2-terminal Domain of Histone H1," *Journal of Biological Chemistry*, vol. 276, no. 49, pp. 46429–46435, 2001.

[320] A. Roque, I. Ponte, and P. Suau, "Role of charge neutralization in the folding of the carboxy-terminal domain of histone H1," *Journal of Physical Chemistry B*, vol. 113, no. 35, pp. 12061–12066, 2009.

[321] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983.

[322] A. Luque, R. Collepardo-Guevara, S. Grigoryev, and T. Schlick, "Dynamic condensation of linker histone C-terminal domain regulates chromatin structure," *Nucleic Acids Research*, vol. 42, no. 12, pp. 7553–7560, 2014.

[323] A. Kolinski and J. Skolnick, "Determinants of secondary structure of polypeptide chains: Interplay between short range and burial interactions," *Journal of Chemical Physics*, vol. 107, no. 3, pp. 953–964, 1997.

[324] F. Ding, R. K. Jha, and N. V. Dokholyan, "Scaling behavior and structure of denatured proteins," *Structure*, vol. 13, no. 7, pp. 1047–1054, 2005.

[325] M. Pasi, K. Zakrzewska, J. H. Maddocks, and R. Lavery, "Analyzing DNA curvature and its impact on the ionic environment: Application to molecular dynamics simulations of minicircles," *Nucleic Acids Research*, vol. 45, no. 7, pp. 4269–4277, 2017.

[326] C. Blanchet, M. Pasi, K. Zakrzewska, and R. Lavery, "CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures," *Nucleic Acids Research*, vol. 39, no. 2, 2011.

[327] A. Luque, G. Ozer, and T. Schlick, "Correlation among DNA Linker Length, Linker Histone Concentration, and Histone Tails in Chromatin," *Biophysical Journal*, vol. 110, no. 11, pp. 2309–2319, 2016.

[328] J. Jung, W. Nishima, M. Daniels, G. Bascom, C. Kobayashi, A. Adedoyin, M. Wall, A. Lappala, D. Phillips, W. Fischer, C. S. Tung, T. Schlick, Y. Sugita, and K. Y. Sanbonmatsu, "Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations," *Journal of Computational Chemistry*, vol. 40, no. 21, pp. 1919–1930, 2019.

[329] R. Collepardo-Guevara and T. Schlick, "Crucial role of dynamic linker histone binding and divalent ions for DNA accessibility and gene regulation revealed by mesoscale modeling of oligonucleosomes," *Nucleic Acids Research*, vol. 40, no. 18, pp. 8803–8817, 2012.

[330] O. Perišić and T. Schlick, "Dependence of the Linker Histone and Chromatin Condensation on the Nucleosome Environment," *Journal of Physical Chemistry B*, vol. 121, no. 33, pp. 7823–7832, 2017.

[331] H. E. Kasinsky, J. D. Lewis, J. B. Dacks, and J. Ausio, "Origin of H1 linker histones.," *FASEB Journal*, vol. 15, no. 1, pp. 34–42, 2001.

[332] J. D. Lewis, N. Saperas, Y. Song, M. J. Zamora, M. Chiva, and J. Ausio, "Histone H1 and the origin of protamines," *Proceedings of the National Academy of Sciences*, vol. 101, no. 12, pp. 4148–4152, 2004.

[333] I. Ponte, D. Romero, D. Yero, P. Suau, and A. Roque, "Complex evolutionary history of the mammalian histone H1.1-H1.5 gene family," *Molecular Biology and Evolution*, vol. 34, no. 3, pp. 545–558, 2017.

[334] J. R. Pehrson, "Core histone variants," *New Comprehensive Biochemistry*, vol. 39, pp. 188–204, 2004.

[335] M. H. Parseghian, A. H. Henschen, K. G. Krieglstein, and B. A. Hamkalo, "A proposal for a coherent mammalian histone H1 nomenclature correlated with amino acid sequences," *Protein Science*, vol. 3, no. 4, pp. 575–587, 1994.

[336] M. H. Parseghian and B. A. Hamkalo, "A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics," *Biochemistry and Cell Biology*, vol. 79, no. 3, pp. 289–304, 2001.

[337] P. B. Talbert, K. Ahmad, G. Almouzni, J. Ausiá, F. Berger, P. L. Bhalla, W. M. Bonner, W. Z. Cande, B. P. Chadwick, S. W. Chan, G. A. Cross, L. Cui, S. I. Dimitrov, D. Doenecke, J. M. Eirin-Lápez, M. A. Gorovsky, S. B. Hake, B. A. Hamkalo, S. Holec, S. E. Jacobsen, K. Kamieniarz, S. Khochbin, A. G. Ladurner, D. Landsman, J. A. Latham, B. Loppin, H. S. Malik, W. F. Marzluff, J. R. Pehrson, J. Postberg, R. Schneider, M. B. Singh, M. M. Smith, E. Thompson, M. E. Torres-Padilla, D. J. Tremethick, B. M. Turner, J. H. Waterborg, H. Wollmann, R. Yelagandula, B. Zhu, and S. Henikoff, "A unified phylogeny-based nomenclature for histone variants," *Epigenetics and Chromatin*, vol. 5, no. 1, pp. 1–19, 2012.

[338] Y. Fan, A. Sirotkin, R. G. Russell, J. Ayala, and A. I. Skoultchi, "Individual Somatic H1 Subtypes Are Dispensable for Mouse Development Even in Mice Lacking the H10 Replacement Subtype," *Molecular and Cellular Biology*, vol. 21, no. 23, pp. 7933–7943, 2001.

[339] E. M. George, T. Izard, S. D. Anderson, and D. T. Brown, "Nucleosome interaction surface of linker histone H1c is distinct from that of H10," *Journal of Biological Chemistry*, vol. 285, no. 27, pp. 20891–20896, 2010.

[340] M. Sancho, E. Diani, M. Beato, and A. Jordan, "Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth," *PLoS Genetics*, vol. 4, no. 10, 2008.

[341] H. Lee, R. Habas, and C. Abate-Shen, "Msx1 cooperates with histone H1b for inhibition of transcription and myogenesis," *Science*, vol. 304, no. 5677, pp. 1675–1678, 2004.

[342] A. Vaquero, M. Scher, D. Lee, H. Erdjument-Bromage, P. Tempst, and D. Reinberg, "Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin," *Molecular Cell*, vol. 16, no. 1, pp. 93–105, 2004.

[343] J. Zlatanova and D. Doenecke, "Histone H10: a major player in cell differentiation?," *FASEB Journal*, vol. 8, no. 15, pp. 1260–1268, 1994.

[344] B. Piña, P. Martínez, L. Simón, and P. Suau, "Differential kinetics of histone H1(0) accumulation in neuronal and glial cells from rat cerebral cortex during postnatal development," *Biochemical and Biophysical Research Communications*, vol. 123, no. 2, pp. 697–702, 1984.

[345] L. M. Garcia-Segura, S. Luquin, P. Martinez, M. T. Casas, and P. Suau, "Differential expression and gonadal hormone regulation of histone H1(0) in the developing and adult rat brain," *Brain Res Dev Brain Res*, vol. 73, no. 1, pp. 63–70, 1993.

[346] C. Pan and Y. Fan, "Role of H1 linker histones in mammalian development and stem cell differentiation," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1859, no. 3, pp. 496–509, 2016.

[347] S. Bhan, W. May, S. L. Warren, and D. B. Sittman, "Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H10, in gene expression regulation," *Gene*, vol. 414, no. 1, pp. 10–18, 2008.

[348] A. M. Sirotkin, W. Edelmann, G. Cheng, A. Klein-Szanto, R. Kucherlapati, and A. I. Skoultchi, "Mice develop normally without the H1(0) linker histone.," *Proceedings of the National Academy of Sciences*, vol. 92, no. 14, pp. 6434–8, 1995.

[349] N. Happel, E. Schulze, and D. Doenecke, "Characterisation of human histone H1x," *Biological Chemistry*, vol. 386, no. 6, pp. 541–551, 2005.

[350] S. Stoldt, D. Wenzel, E. Schulze, D. Doenecke, and N. Happel, "G1 phase-dependent nucleolar accumulation of human histone H1x," *Biology of the Cell*, vol. 99, no. 10, pp. 541–552, 2007.

[351] H. Takata, S. Matsunaga, A. Morimoto, R. Ono-Maniwa, S. Uchiyama, and K. Fukui, "H1.X with different properties from other linker histones is required for mitotic progression," *FEBS Letters*, vol. 581, no. 20, pp. 3783–3788, 2007.

[352] J. Warneboldt, F. Haller, O. Horstmann, B. C. Danner, L. Füzesi, D. Doenecke, and N. Happel, "Histone H1x is highly expressed in human neuroendocrine cells and tumours," *BMC Cancer*, vol. 8, no. 388, pp. 1–9, 2008.

[353] R. Mayor, A. Izquierdo-Bouldstridge, L. Millán-Ariño, A. Bustillos, C. Sampaio, N. Luque, and A. Jordan, "Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions," *Journal of Biological Chemistry*, vol. 290, no. 12, pp. 7474–7491, 2015.

[354] J. Clausell, N. Happel, T. K. Hale, D. Doenecke, and M. Beato, "Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF," *PLoS ONE*, vol. 4, no. 10, 2009.

[355] H. Talasz, N. Sapojnikova, W. Helliger, H. Lindner, and B. Puschendorf, "In vitro binding of H1 histone subtypes to nucleosomal organized mouse mammary tumor virus long terminal repeat promotor," *Journal of Biological Chemistry*, vol. 273, no. 48, pp. 32236–32243, 1999.

[356] J. P. H. Th'ng, R. Sung, M. Ye, and M. J. Hendzel, "H1 family histones in the nucleus: Control of binding and localization by the C-terminal domain," *Journal of Biological Chemistry*, vol. 280, no. 30, pp. 27809–27814, 2005.

[357] M. Orrego, I. Ponte, A. Roque, N. Buschati, X. Mora, and P. Suau, "Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin," *BMC Biology*, vol. 5, no. 22, pp. 1–11, 2007.

[358] M. A. Öztürk, G. V. Pachov, R. C. Wade, and V. Cojocaru, "Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome," *Nucleic Acids Research*, vol. 44, no. 14, pp. 6599–6613, 2016.

[359] L. Böhm and T. C. Mitchell, "Sequence conservation in the N-terminal domain of histone H1," *FEBS Letters*, vol. 193, no. 1, pp. 1–4, 1985.

[360] D. T. Brown, B. T. Alexander, and D. B. Sittman, "Differential effect of H1 variant overexpression on cell cycle progression and gene expression," *Nucleic Acids Research*, vol. 24, no. 3, pp. 486–493, 1996.

[361] R. B. Best and G. Hummer, "Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides," *Journal of Physical Chemistry B*, vol. 113, no. 26, pp. 9004–9015, 2009.

[362] R. B. Best, D. De Sancho, and J. Mittal, "Residue-specific α-helix propensities from molecular simulation," *Biophysical Journal*, vol. 102, no. 6, pp. 1462–1467, 2012.

[363] J. Åqvist, "Ion-water interaction potentials derived from free energy perturbation simulations," *Journal of Physical Chemistry*, vol. 94, no. 21, pp. 8021–8024, 1990.

[364] L. X. Dang, "Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study," *Journal of the American Chemical Society*, vol. 117, no. 26, pp. 6954–6960, 1995.

[365] D. Beglov and B. Roux, "Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations," *The Journal of Chemical Physics*, vol. 100, no. 12, pp. 9050–9063, 1994.

[366] Y. Luo and B. Roux, "Simulation of osmotic pressure in concentrated aqueous salt solutions," *Journal of Physical Chemistry Letters*, vol. 1, no. 1, pp. 183–189, 2010.

[367] R. M. Venable, Y. Luo, K. Gawrisch, B. Roux, and R. W. Pastor, "Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data," *Journal of Physical Chemistry B*, vol. 117, no. 35, pp. 10183–10192, 2013.

[368] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs.," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–88, 2005.

[369] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, "Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium," *Journal of Chemical Theory and Computation*, vol. 8, no. 1, pp. 348–362, 2012.

[370] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, "Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers," *Biophysical Journal*, vol. 92, no. 11, pp. 3817–3829, 2007.

[371] M. Wieczór and J. Czub, "How proteins bind to DNA: target discrimination and dynamic sequence search by the telomeric protein TRF1," *Nucleic Acids Research*, vol. 45, no. 13, pp. 7643–7654, 2017.

[372] M. Wieczór, A. Tobiszewski, P. Wityk, B. Tomiczek, and J. Czub, "Molecular recognition in complexes of TRF proteins with telomeric DNA," *PLoS ONE*, vol. 9, no. 2, 2014.

[373] G. Hedger, D. Shorthouse, H. Koldsø, and M. S. Sansom, "Free Energy Landscape of Lipid Interactions with Regulatory Binding Sites on the Transmembrane Domain of the EGF Receptor," *Journal of Physical Chemistry B*, vol. 120, no. 33, pp. 8154–8163, 2016.

[374] M. Bernetti, M. Masetti, F. Pietrucci, M. Blackledge, M. R. Jensen, M. Recanatini, L. Mollica, and A. Cavalli, "Structural and Kinetic Characterization of the Intrinsically Disordered Protein SeV NTAILthrough Enhanced Sampling Simulations," *Journal of Physical Chemistry B*, vol. 121, no. 41, pp. 9572–9582, 2017.

[375] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L. P. Wang, T. J. Lane, and V. S. Pande, "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories," *Biophysical Journal*, vol. 109, no. 8, pp. 1528–1532, 2015.

[376] A. P. Collins and P. C. Anderson, "Complete Coupled Binding-Folding Pathway of the Intrinsically Disordered Transcription Factor Protein Brinker Revealed by Molecular Dynamics Simulations and Markov State Modeling," *Biochemistry*, vol. 57, no. 30, pp. 4404–4420, 2018.

[377] B. R. Groveman, A. Kraus, L. D. Raymond, M. A. Dolan, K. J. Anson, D. W. Dorward, and B. Caughey, "Charge neutralization of the central lysine cluster in prion protein (PrP) promotes PrPSc-Like folding of recombinant PrP amyloids," *Journal of Biological Chemistry*, vol. 290, no. 2, pp. 1119–1128, 2015.

[378] E. Socher and H. Sticht, "Mimicking titration experiments with MD simulations: A protocol for the investigation of pH-dependent effects on proteins," *Scientific Reports*, 2016.

[379] F. Agou, J. P. Waller, Y. Yang, E. Guittet, and J. C. Gesquière, "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*, vol. 34, no. 2, pp. 569–576, 1995.

[380] N. P. Johnson, J. Lindstrom, W. A. Baase, and P. H. von Hippel, "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.," *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 4840–4844, 1994.

[381] M. K. Kim and Y. K. Kang, "Positional preference of proline in α-helices," *Protein Science*, vol. 8, no. 7, pp. 1492–1499, 1999.

[382] B. F. Fisher, S. H. Hong, and S. H. Gellman, "Helix Propensities of Amino Acid Residues via Thioester Exchange," *Journal of the American Chemical Society*, vol. 139, no. 38, pp. 13292–13295, 2017.

[383] S. Rauscher, V. Gapsys, B. de Groot, and H. Grubmüller, "Structural Ensembles of Intrinsically Disordered Proteins using Molecular Dynamics Simulation," *Biophysical Journal*, vol. 108, no. 2, p. 14a, 2015.

[384] I. S. Joung and T. E. Cheatham, "Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations," *Journal of Physical Chemistry B*, vol. 112, no. 30, pp. 9020–9041, 2008.

[385] C. O. Pabo and L. Nekludova, "Geometric analysis and comparison of protein-DNA interfaces: Why is there no simple code for recognition?," *Journal of Molecular Biology*, vol. 301, no. 3, pp. 597–624, 2000.

[386] M. Cassandri, A. Smirnov, F. Novelli, C. Pitolli, M. Agostini, M. Malewicz, G. Melino, and G. Raschellà, "Zinc-finger proteins in health and disease," *Cell Death Discovery*, vol. 3, no. 1, 2017.

[387] P. R. Manna, M. T. Dyson, and D. M. Stocco, "Role of basic leucine zipper proteins in transcriptional regulation of the steroidogenic acute regulatory protein gene," *Molecular and Cellular Endocrinology*, vol. 302, no. 1, pp. 1–11, 2009.

[388] J. Yoo and A. Aksimentiev, "Improved Parameterization of Amine-Carboxylate and Amine-Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields," *Journal of Chemical Theory and Computation*, vol. 12, no. 1, pp. 430–443, 2016.

[389] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[390] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, and M. Orozco, "Parmbsc1: a refined force field for DNA simulations.," *Nature Methods*, vol. 13, no. 1, pp. 55–58, 2015.

[391] J. Yoo and A. Aksimentiev, "New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions," *Physical Chemistry Chemical Physics*, vol. 20, no. 13, pp. 8432–8449, 2018.

[392] A. B. Adomas and P. Wade, "Chromatin Structure and Gene Expression: Function Follows Form," in *Environmental Epigenomics in Health and Disease*, pp. 189–205, 2013.

[393] Y. Chen, J. M. Tokuda, T. Topping, S. P. Meisburger, S. A. Pabit, L. M. Gloss, and L. Pollack, "Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core," *Proceedings of the National Academy of Sciences*, vol. 114, no. 2, pp. 334–339, 2017.

[394] S. Rudnizky, O. Malik, A. Bavly, L. Pnueli, P. Melamed, and A. Kaplan, "Nucleosome mobility and the regulation of gene expression: Insights from single-molecule studies," *Protein Science*, vol. 26, no. 7, pp. 1266–1277, 2017.

[395] R. J. Palstra, B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. De Laat, "The β-globin nuclear compartment in development and erythroid differentiation," 2003.

[396] M. A. Ricci, C. Manzo, M. F. García-Parajo, M. Lakadamyali, and M. P. Cosma, "Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo," *Cell*, vol. 160, no. 6, pp. 1145–1158, 2015.

[397] R. Drissen, R. J. Palstra, N. Gillemans, E. Splinter, F. Grosveld, S. Philipsen, and W. De Laat, "The active spatial organization of the β-globin locus requires the transcription factor EKLF," *Genes and Development*, vol. 18, no. 20, pp. 2485–2490, 2004.

[398] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, "Three-dimensional folding and functional organization principles of the Drosophila genome," *Cell*, vol. 148, no. 3, pp. 458–472, 2012.

[399] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. Van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard, "Spatial partitioning of the regulatory landscape of the X-inactivation centre," *Nature*, vol. 485, no. 7398, pp. 381–385, 2012.

[400] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, no. 7398, pp. 376–380, 2012.

[401] H. Tamaru, "Confining euchromatin/heterochromatin territory: Jumonji crosses the line," 2010.

[402] S. I. Grewal and S. Jia, "Heterochromatin revisited," *Nature Reviews Genetics*, vol. 8, no. 1, pp. 35–46, 2007.

[403] K. L. Huisinga, B. Brower-Toland, and S. C. Elgin, "The contradictory definitions of heterochromatin: Transcription and silencing," *Chromosoma*, vol. 115, no. 2, pp. 110–122, 2006.

[404] A. E. Fazary, Y. H. Ju, and H. S. Abd-Rabboh, "How does chromatin package DNA within nucleus and regulate gene expression?," *International Journal of Biological Macromolecules*, vol. 101, pp. 862–881, 2017.

[405] K. L. Jost, B. Bertulat, and M. C. Cardoso, "Heterochromatin and gene positioning: Inside, outside, any side?," *Chromosoma*, vol. 121, no. 6, pp. 555–563, 2012.

[406] D. A. Potoyan and G. A. Papoian, "Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 44, pp. 17857–17862, 2012.

[407] V. Krishnan, M. Z. Y. Chow, Z. Wang, L. Zhang, B. Liu, X. Liu, and Z. Zhou, "Histone H4 lysine 16 hypoacetylation is associated with defective DNA repair and premature senescence in Zmpste24-deficient mice," *Proceedings of the National Academy of Sciences*, vol. 108, no. 30, pp. 12325–12330, 2011.

[408] D. Canzio, A. Larson, and G. J. Narlikar, "Mechanisms of functional promiscuity by HP1 proteins," *Trends in Cell Biology*, vol. 24, no. 6, pp. 377–386, 2014.

[409] D. Vermaak and H. S. Malik, "Multiple Roles for Heterochromatin Protein 1 Genes in Drosophila," *Annual Review of Genetics*, vol. 43, no. 1, pp. 467–492, 2009.

[410] G. Lomberk, L. L. Wallrath, and R. Urrutia, "The Heterochromatin Protein 1 family," 2006.

[411] J. C. Eissenberg and S. C. Elgin, "The HP1 protein family: Getting a grip on chromatin," *Current Opinion in Genetics and Development*, vol. 10, no. 2, pp. 204–210, 2000.

[412] N. P. Cowieson, J. F. Partridge, R. C. Allshire, and P. J. McLaughlin, "Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis," *Current Biology*, vol. 10, no. 9, pp. 517–525, 2000.

[413] L. Kaustov, H. Ouyang, M. Amaya, A. Lemak, N. Nady, S. Duan, G. A. Wasney, Z. Li, M. Vedadi, M. Schapira, J. Min, and C. H. Arrowsmith, "Recognition and specificity determinants of the human Cbx chromodomains," *Journal of Biological Chemistry*, vol. 286, no. 1, pp. 521–529, 2011.

[414] A. Thiru, D. Nietlispach, H. R. Mott, M. Okuwaki, D. Lyon, P. R. Nielsen, M. Hirshberg, A. Verreault, N. V. Murzina, and E. D. Laue, "Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin," *EMBO Journal*, vol. 23, no. 3, pp. 489–499, 2004.

[415] V. Arcus, "OB-fold domains: A snapshot of the evolution of sequence, structure and function," *Current Opinion in Structural Biology*, vol. 12, no. 6, pp. 794–801, 2002.

[416] J. C. Eissenberg, "Structural biology of the chromodomain: Form and function," *Gene*, vol. 496, no. 2, pp. 69–78, 2012.

[417] S. Machida, Y. Takizawa, M. Ishimaru, Y. Sugita, S. Sekine, J. ichi Nakayama, M. Wolf, and H. Kurumizaka, "Structural Basis of Heterochromatin Formation by Human HP1," *Molecular Cell*, vol. 69, no. 3, pp. 385–397.e8, 2018.

[418] A. Kumar and H. Kono, "Heterochromatin protein 1 (HP1): interactions with itself and chromatin components," *Biophysical Reviews*, 2020.

[419] R. Assland and F. Stewart, "The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1," *Nucleic Acids Research*, vol. 23, no. 16, pp. 3168–3173, 1995.

[420] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept," *Annual Review of Biophysics*, vol. 37, no. 1, pp. 215–246, 2008.

[421] G. Nishibuchi and J. I. Nakayama, "Biochemical and structural properties of heterochromatin protein 1: Understanding its role in chromatin assembly," *Journal of Biochemistry*, vol. 156, no. 1, pp. 11–20, 2014.

[422] R. R. Meehan, C. F. Kao, and S. Pennings, "HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain," *EMBO Journal*, vol. 22, no. 12, pp. 3164–3174, 2003.

[423] C. Keller, R. Adaixo, R. Stunnenberg, K. J. Woolcock, S. Hiller, and M. Bühler, "HP1 Swi6 Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts," *Molecular Cell*, vol. 47, no. 2, pp. 215–227, 2012.

[424] R. K. Badugu, Y. Yoo, P. B. Singh, and R. Kellum, "Mutations in the heterochromatin protein 1 (HP1) hinge domain affect HP1 protein interactions and chromosomal distribution," *Chromosoma*, vol. 113, no. 7, pp. 370–384, 2005.

[425] C. L. Zhang, T. A. McKinsey, and E. N. Olson, "Association of Class II Histone Deacetylases with Heterochromatin Protein 1: Potential Role for Histone Methylation in Control of Muscle Differentiation," *Molecular and Cellular Biology*, vol. 22, no. 20, pp. 7302–7312, 2002.

[426] A. M. Ainsztein, S. E. Kandels-Lewis, A. M. Mackay, and W. C. Earnshaw, "INCENP centromere and spindle targeting: Identification of essential conserved motifs and involvement of heterochromatin protein HP1," *Journal of Cell Biology*, vol. 143, no. 7, pp. 1763–1774, 1998.

[427] C. Dinant and M. S. Luijsterburg, "The Emerging Role of HP1 in the DNA Damage Response," *Molecular and Cellular Biology*, vol. 29, no. 24, pp. 6335–6340, 2009.

[428] M. S. Luijsterburg, C. Dinant, H. Lans, J. Stap, E. Wiernasz, S. Lagerwerf, D. O. Warmerdam, M. Lindh, M. C. Brink, J. W. Dobrucki, J. A. Aten, M. I. Fousteri, G. Jansen, N. P. Dantuma, W. Vermeulen, L. H. Mullenders, A. B. Houtsmuller, P. J. Verschure, and R. Van Driel, "Heterochromatin protein 1 is recruited to various types of DNA damage," *Journal of Cell Biology*, vol. 185, no. 4, pp. 577–586, 2009.

[429] R. F. Luco, M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli, "Epigenetics in alternative pre-mRNA splicing," 2011.

[430] A. Yearim, S. Gelfman, R. Shayevitch, S. Melcer, O. Glaich, J. P. Mallm, M. Nissim-Rafinia, A. H. S. Cohen, K. Rippe, E. Meshorer, and G. Ast, "HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing," *Cell Reports*, vol. 10, no. 7, pp. 1122–1134, 2015.

[431] B. Perrini, L. Piacentini, L. Fanti, F. Altieri, S. Chichiarelli, M. Berloco, C. Turano, A. Ferraro, and S. Pimpinelli, "HP1 controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in Drosophila," *Molecular Cell*, vol. 15, no. 3, pp. 467–476, 2004.

[432] S. Canudas, B. R. Houghtaling, M. Bhanot, G. Sasa, S. A. Savage, A. A. Bertuch, and S. Smith, "A role for heterochromatin protein 1γ at human telomeres," *Genes and Development*, vol. 25, no. 17, pp. 1807–1819, 2011.

[433] C. R. Vakoc, S. A. Mandat, B. A. Olenchock, and G. A. Blobel, "Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin.," *Molecular cell*, vol. 19, pp. 381–391, aug 2005.

[434] S. H. Kwon and J. L. Workman, "The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression," *BioEssays*, vol. 33, no. 4, pp. 280–289, 2011.

[435] G. LeRoy, J. T. Weston, B. M. Zee, N. L. Young, M. D. Plazas-Mayorca, and B. A. Garcia, "Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications," *Molecular and Cellular Proteomics*, vol. 8, no. 11, pp. 2432–2442, 2009.

[436] Y. Mishima, M. Watanabe, T. Kawakami, C. D. Jayasinghe, J. Otani, Y. Kikugawa, M. Shirakawa, H. Kimura, O. Nishimura, S. Aimoto, S. Tajima, and I. Suetake, "Hinge and chromoshadow of HP1α participate in recognition of K9 methylated histone H3 in nucleosomes," *Journal of Molecular Biology*, vol. 425, no. 1, pp. 54–70, 2013.

[437] K. Hiragami-Hamada, K. Shinmyozu, D. Hamada, Y. Tatsu, K. Uegaki, S. Fujiwara, and J.-i. Nakayama, "N-Terminal Phosphorylation of HP1 Promotes Its Chromatin Binding," *Molecular and Cellular Biology*, vol. 31, no. 6, pp. 1186–1200, 2011.

[438] M. Billur, H. D. Bartunik, and P. B. Singh, "The essential function of HP1β: a case of the tail wagging the dog?," *Trends in Biochemical Sciences*, vol. 35, no. 2, pp. 115–123, 2010.

[439] A. Smallwood, G. C. Hon, F. Jin, R. E. Henry, J. M. Espinosa, and B. Ren, "CBX3 regulates efficient RNA processing genome-wide," *Genome Research*, vol. 22, no. 8, pp. 1426–1436, 2012.

[440] G. Nishibuchi, S. Machida, A. Osakabe, H. Murakoshi, K. Hiragami-Hamada, R. Nakagawa, W. Fischle, Y. Nishimura, H. Kurumizaka, H. Tagami, and J. I. Nakayama, "N-terminal phosphorylation of HP1α increases its nucleosome-binding specificity," *Nucleic Acids Research*, vol. 42, no. 20, pp. 12498–12511, 2014.

[441] L. C. Bryan, D. R. Weilandt, A. L. Bachmann, S. Kilic, C. C. Lechner, P. D. Odermatt, G. E. Fantner, S. Georgeon, O. Hantschel, V. Hatzimanikatis, and B. Fierz, "Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions," *Nucleic Acids Research*, vol. 45, no. 18, pp. 10504–10517, 2017.

[442] C. Maison and G. Almouzni, "HP1 and the dynamics of heterochromatin maintenance," *Nature Reviews Molecular Cell Biology*, vol. 5, no. 4, pp. 296–304, 2004.

[443] T. Zhao, T. Heyduk, and J. C. Eissenberg, "Phosphorylation Site Mutations in Heterochromatin Protein 1 (HP1) Reduce or Eliminate Silencing Activity," *Journal of Biological Chemistry*, vol. 276, no. 12, pp. 9512–9518, 2001.

[444] K. Kataoka, T. Noto, and K. Mochizuki, "Phosphorylation of an HP1-like protein is a prerequisite for heterochromatin body formation in Tetrahymena DNA elimination," *Proceedings of the National Academy of Sciences*, vol. 113, no. 32, pp. 9027–9032, 2016.

[445] S. Chen, C. Wang, L. Sun, D.-L. Wang, L. Chen, Z. Huang, Q. Yang, J. Gao, X.-B. Yang, J.-F. Chang, P. Chen, L. Lan, Z. Mao, and F.-L. Sun, "RAD6 Promotes Homologous Recombination Repair by Activating the Autophagy-Mediated Degradation of Heterochromatin Protein HP1," *Molecular and Cellular Biology*, vol. 35, no. 2, pp. 406–416, 2015.

[446] C. Maison, D. Bailly, D. Roche, R. M. De Oca, A. V. Probst, I. Vassias, F. Dingli, B. Lombard, D. Loew, J. P. Quivy, and G. Almouzni, "SUMOylation promotes de novo targeting of HP1a to pericentric heterochromatin," *Nature Genetics*, vol. 43, no. 3, pp. 220–227, 2011.

[447] M. Wiese, A. J. Bannister, S. Basu, W. Boucher, K. Wohlfahrt, M. A. Christophorou, M. L. Nielsen, D. Klenerman, E. D. Laue, and T. Kouzarides, "Citrullination of HP1γ chromodomain affects association with chromatin," *Epigenetics and Chromatin*, vol. 12, no. 1, 2019.

[448] F. Munari, M. J. Gajda, K. Hiragami-Hamada, W. Fischle, and M. Zweckstetter, "Characterization of the effects of phosphorylation by CK2 on the structure and binding properties of human HP1β," *FEBS Letters*, vol. 588, no. 7, pp. 1094–1099, 2014.

[449] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin," *Nature*, vol. 547, no. 7662, pp. 236–240, 2017.

[450] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, "Phase separation drives heterochromatin domain formation," *Nature*, vol. 547, no. 7662, pp. 241–245, 2017.

[451] H. Shimojo, A. Kawaguchi, T. Oda, N. Hashiguchi, S. Omori, K. Moritsugu, A. Kidera, K. Hiragami-Hamada, J. I. Nakayama, M. Sato, and Y. Nishimura, "Extended string-like binding of the phosphorylated HP1α N-terminal tail to the lysine 9-methylated histone H3 tail," *Scientific Reports*, vol. 6, pp. 1–15, 2016.

[452] S. Sanulli, M. J. Trnka, V. Dharmarajan, R. W. Tibble, B. D. Pascal, A. L. Burlingame, P. R. Griffin, J. D. Gross, and G. J. Narlikar, "HP1 reshapes nucleosome core to promote phase separation of heterochromatin," *Nature*, vol. 575, no. 7782, pp. 390–394, 2019.

[453] S. Arora, W. S. Horne, and K. Islam, "Engineering Methyllysine Writers and Readers for Allele-Specific Regulation of Protein-Protein Interactions," *Journal of the American Chemical Society*, vol. 141, no. 39, pp. 15466–15470, 2019.

[454] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank.," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[455] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the charmm force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models," *Journal of Chemical Theory and Computation*, vol. 6, no. 2, pp. 459–466, 2010.

[456] A. Warnecke, T. Sandalova, A. Achour, and R. A. Harris, "PyTMs: A useful PyMOL plugin for modeling common post-translational modifications," *BMC Bioinformatics*, vol. 15, no. 1, pp. 1–12, 2014.

[457] G. V. Papamokos, G. Tziatzos, D. G. Papageorgiou, S. D. Georgatos, A. S. Politou, and E. Kaxiras, "Structural role of RKS motifs in chromatin interactions: A molecular dynamics study of HP1 bound to a variably modified histone tail," *Biophysical Journal*, vol. 102, no. 8, pp. 1926–1933, 2012.

[458] T. Steinbrecher, J. Latzer, and D. A. Case, "Revised AMBER parameters for bioorganic phosphates," *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4405–4412, 2012.

[459] C. Grauffel, R. H. Stote, and A. Dejaegere, "Force field parameters for the simulation of modified histone tails," *Journal of Computational Chemistry*, vol. 31, no. 13, pp. 2434–2451, 2010.

[460] M. H. Feng, M. Philippopoulos, A. D. MacKerell, and C. Lim, "Structural characterization of the phosphotyrosine binding region of a high-affinity SH2 domain-phosphopeptide complex by molecular dynamics simulation and chemical shift calculations," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11265–11277, 1996.

[461] U. Baul, D. Chakraborty, M. L. Mugnai, J. E. Straub, and D. Thirumalai, "Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins," *Journal of Physical Chemistry B*, vol. 123, no. 16, pp. 3462–3474, 2019.

[462] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna,

S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020.

[463] A. E. Zambelli, "A data-driven approach to estimating the number of clusters in hierarchical clustering," *F1000Research*, vol. 5, 2016.

[464] A. Chakraborty, K. V. Prasanth, and S. G. Prasanth, "Dynamic phosphorylation of HP1α regulates mitotic progression in human cells," *Nature Communications*, vol. 5, no. 1, p. 3445, 2014.

[465] M. R. Machado, P. D. Dans, and S. Pantano, "Isoform-specific determinants in the HP1 binding to histone 3: Insights from molecular simulations," *Amino Acids*, vol. 38, no. 5, pp. 1571–1581, 2010.

[466] Y. K. Jiang, J. W. Zou, W. Yu-Qian, N. Zhang, Q. S. Yu, and Y. J. Jiang, "Molecular dynamics simulation on HP1 protein binding by histone H3 tail methylation and phosphoiylation," *International Journal of Quantum Chemistry*, vol. 109, no. 4, pp. 746–755, 2009.

[467] K. T. Debiec, A. M. Gronenborn, and L. T. Chong, "Evaluating the strength of salt bridges: A comparison of current biomolecular force fields," *Journal of Physical Chemistry B*, vol. 118, no. 24, pp. 6561–6569, 2014.

[468] Y. Liu, S. Qin, M. Lei, W. Tempel, Y. Zhang, P. Loppnau, Y. Li, and J. Min, "Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains revisited: Plasticity in the pseudosymmetric histone binding site of human HP1," *Journal of Biological Chemistry*, vol. 292, no. 14, pp. 5655–5664, 2017.

[469] W. Fischle, S. T. Boo, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, and C. D. Allis, "Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation," *Nature*, vol. 438, no. 7071, pp. 1116–1122, 2005.

[470] D. Demidov, S. Hesse, A. Tewes, T. Rutten, J. Fuchs, R. Karimi Ashtiyani, S. Lein, A. Fischer, G. Reuter, and A. Houben, "Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in Arabidopsis," *Plant Journal*, vol. 59, no. 2, pp. 221–230, 2009.

[471] J. Li, Q. Lin, H.-G. Yoon, Z.-Q. Huang, B. D. Strahl, C. D. Allis, and J. Wong, "Involvement of Histone Methylation and Phosphorylation in Regulation of Transcription by Thyroid Hormone Receptor," *Molecular and Cellular Biology*, vol. 22, no. 16, pp. 5688–5697, 2002.

[472] W. Fischle, Y. Wang, and C. D. Allis, "Binary switches and modification cassettes in histone biology and beyond," *Nature*, vol. 425, no. 6957, pp. 475–479, 2003.

[473] P. Sabbattini, M. Sjoberg, S. Nikic, A. Frangini, P. H. Holmqvist, N. Kunowska, T. Carroll, E. Brookes, S. J. Arthur, A. Pombo, and N. Dillon, "An H3K9/S10 methyl-phospho switch modulates Polycomb and Pol II binding at repressed genes during differentiation," *Molecular Biology of the Cell*, vol. 25, no. 6, pp. 904–915, 2014.

[474] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.

[475] M. Shirts and V. S. Pande, "Screen savers of the world unite," *Science*, vol. 290, no. 5498, pp. 1903–1904, 2000.

[476] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, "Anton, a special-purpose machine for molecular dynamics simulation," *Communications of the ACM*, vol. 51, no. 7, pp. 91–97, 2008.

[477] A. Deiana, S. Forcelloni, A. Porrello, and A. Giansanti, "Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell," *PLoS ONE*, vol. 14, no. 8, 2019.

[478] R. B. Best, N. V. Buchete, and G. Hummer, "Are current molecular dynamics force fields too helical?," *Biophysical Journal*, vol. 95, no. 1, pp. 7–9, 2008.

[479] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," *PLoS ONE*, vol. 7, no. 2, 2012.

[480] P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," *Proceedings of the National Academy of Sciences*, vol. 115, no. 21, pp. E4758–E4766, 2018.

[481] D. Song, R. Luo, and H. F. Chen, "The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins," *Journal of Chemical Information and Modeling*, vol. 57, no. 5, pp. 1166–1178, 2017.

[482] S. J. Triezenberg, "Structure and function of transcriptional activation domains," *Current Opinion in Genetics and Development*, vol. 5, no. 2, pp. 190–196, 1995.

[483] K. W. Barber and J. Rinehart, "The ABCs of PTMs," *Nature Chemical Biology*, vol. 14, no. 3, pp. 188–192, 2018.

[484] D. R. Foltz, L. E. Jansen, B. E. Black, A. O. Bailey, J. R. Yates, and D. W. Cleveland, "The human CENP-A centromeric nucleosome-associated complex," *Nature Cell Biology*, vol. 8, no. 5, pp. 458–469, 2006.

[485] Y. Chen, M. E. Hoover, X. Dang, A. A. Shomo, X. Guan, A. G. Marshall, M. A. Freitas, and N. L. Young, "Quantitative mass spectrometry reveals that intact histone H1 phosphorylations are variant specific and exhibit single molecule hierarchical dependence," *Molecular and Cellular Proteomics*, vol. 15, no. 3, pp. 818–833, 2016.

[486] F. Xin and P. Radivojac, "Post-translational modifications induce significant yet not extreme changes to protein structure," *Bioinformatics*, vol. 28, no. 22, pp. 2905–2913, 2012.

[487] G. Duan and D. Walther, "The Roles of Post-translational Modifications in the Context of Protein Interaction Networks," *PLoS Computational Biology*, vol. 11, no. 2, 2015.

[488] G. A. Khoury, J. P. Thompson, J. Smadbeck, C. A. Kieslich, and C. A. Floudas, "Forcefield-PTM: Ab initio charge and AMBER forcefield parameters for frequently occurring post-translational modifications," *Journal of Chemical Theory and Computation*, vol. 9, no. 12, pp. 5653–5674, 2013.

[489] D. Gfeller, O. Michielin, and V. Zoete, "SwissSidechain: A molecular and structural database of non-natural sidechains," *Nucleic Acids Research*, vol. 41, no. D1, pp. 327–332, 2013.

[490] A. Sridhar, S. E. Farr, G. Portella, T. Schlick, M. Orozco, and R. Collepardo-Guevara, "Emergence of chromatin hierarchical loops from protein disorder and nucleosome asymmetry," *Proceedings of the National Academy of Sciences*, vol. 117, no. 13, pp. 7216–7224, 2020.