Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Mehmet Ali Öztürk

born in Ankara, Turkey

Oral examination: 06.06.2018

A computational approach to decipher chromatosome structure determinants

Referees: Prof. Dr. Rebecca C. Wade

Prof. Dr. Rüdiger Hell

“...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.”

Richard Feynman

Acknowledgements

First, I want to thank Prof. Dr. Rebecca Wade for giving me the opportunity to work in an excellent research environment. Her guidance during projects encouraged and helped me to develop many skills.

I thank all current and past MCM group members for their support and friendship. Our discussions taught me many interesting aspects of different topics and were fun.

I am really grateful to Dr. Michael Martinez, Dr. Neil Bruce, Dr. Daria Kokh and Dr. Stefan Richter for answering my endless questions. Without their help, solving technical issues would have been harder. Further thanks to Dr. Neil Bruce for giving constructive comments to any draft that I have prepared.

I really appreciate HITS staff and administration for providing an exceptionally beautiful working environment. I was always proud to be a HITSter.

I thank our collaborators, Dr. Vlad Cojocaru, Dr. Xavier Salvatella and Dr. Biaoyang Lin for their hospitality and support during my visits to their labs. I learned a lot and enjoyed my time in Münster, Barcelona and Hangzhou.

Thanks to Patrick Friedrich and Ina Pöhner for their help in preparing my abstract in German.

I also learned from my TAC meetings with Prof. Dr. Jörg Langowski, Prof. Dr. Rüdiger Hell, Prof. Dr. Frauke Gräter and Prof. Dr. Sylvia Erhardt. Thanks a lot for their time and suggestions. Prof. Dr. Jörg Langowski inspired me with his enthusiasm. His unexpected passing is very unfortunate for the scientific community and I will always follow the critical thinking attitude with which he guided us.

I am grateful to the Klaus Tschira Foundation for financial support of my studies and its multifold contribution to science. Klaus Tschira and his foundation are role models for their contributions to society.

I thank the HBIGS program organizers for providing many interesting courses to develop our skills. It was a privilege to be part of the graduate school program.

I especially thank Dr. Musa Özboyaci and Ina Pöhner for being with me anytime I needed. We asked hard questions and tried to answer them together. It was an unforgettable journey.

Without the support of Mehmet Köse, Fatma Köse, Hasan Huseyin Köse, Ali Haydar Köse and Halil Ibrahim Köse, I would have really struggled. Thanks a lot for making me feel at home.

Finally, I want to thank my family for supporting me in every aspect that one can imagine. Even if they are thousands of kilometers away, I always had their support.

Abstract

In eukaryotic cells, DNA transcription, replication and repair events are controlled by the regulation of DNA compaction mechanisms that determine the open and closed chromatin states. Nucleosomes are the basic DNA packaging units of chromatin. The nucleosome core (NC) consists of a core histone protein octamer with approximately two tight superhelical turns of DNA wrapped around it. The NC is extended at its entry and exit points by linker DNA (L-DNA) and a linker histone (LH) protein binds between the two L-DNA arms to form a chromatosome. The dyad is the single DNA base pair between the nucleosome entry and exit points determining the symmetry axis and is used to define the position of LH binding to a nucleosome. For LH - nucleosome binding, previous studies indicate both on- and off-dyad binding modes, as well as different LH orientations. Thus, the molecular determinants of the structure of LH – nucleosome complex and the dynamics of LH – nucleosome binding are not fully understood.

The aim of the research described here was to obtain an atomic-detail level understanding of chromatosome formation. Analysis of the experimentally determined structures of LH – nucleosome complexes showed that instead of a single 3D structure, an ensemble of structures of LH – nucleosome complexes exists. To understand the distribution of these ensembles, normal mode analysis (NMA), standard and accelerated molecular dynamics (MD & AMD) and Brownian dynamics (BD) simulations were applied to LH, nucleosome and chromatosome systems. MD and AMD simulations showed that the globular domain of the LH (LH GD) prefers to be in its closed form in solution. Upon nucleosome binding, the LH GD structure transformed to an open structure due to hydrophobic interactions with the L-DNA of the nucleosome. Additionally, LH GD binding constrained the flexibility of the L-DNA and affected the directions of movement of the L-DNA arms. BD simulations indicated that various chromatosome configurations were possible depending on LH GD sequence and L-DNA opening angles. These findings suggest that LH – nucleosome binding is mediated by a combination of conformational selection and induced fit mechanisms.

Further BD simulations show that chromatosome configurations were affected by single point mutations in the LH GD and varied for different LH isoforms. My results indicate that by making specific single point mutation exchanges, it is possible to swap LH – nucleosome configurations among different LH GD isoforms. Similar shifts were observed in chromatosome configuration upon introduction of post translational modifications (PTMs) in the LH GD.

I applied BD simulations to compute dissociation rate constant (k_{off}) values and compare them with previously reported fluorescence recovery after photobleaching (FRAP) data on the binding of various LH mutants to chromatin. The results of the BD simulations correspond with the relative trends in measured FRAP recovery half-times (t_{50}) of LH – chromatin binding of various LH

mutants. The results thus enable the interpretation of the FRAP data in terms of a physical model of LH – nucleosome binding.

My thesis provides detailed insights into the structure, dynamics and kinetics of chromatosome formation in eukaryotes. The results presented in this work can guide further experiments on the sequence determinants of LH – nucleosome binding.

Zusammenfassung

In eukaryotischen Zellen werden DNA-Transkription, Replikation und Reparaturvorgänge durch die Regulierung der dichten Packung von DNA kontrolliert. Dies geschieht mittels Mechanismen, die die Dekondensation oder Kondensation von Chromatin beeinflussen. Die wesentlichen DNA-Verpackungseinheiten des Chromatins werden Nukleosomen genannt. Dabei besteht die Nukleosomen-Grundeinheit ('Nukleosomen Core', NC) aus einem Histonoktamer, um das etwa zwei flache superhelikale Windungen von DNA gewickelt sind. Erweitert wird der NC an DNA-Eintritts- sowie Austrittsstelle durch DNA-Linker (L-DNA). Mit dem zwischen den DNA-Linkern gebundenen Protein, dem Linker Histon (LH), bezeichnet man dies als Chromatosom. Das einzelne DNA-Basenpaar zwischen Eintritts- und Austrittsstelle der DNA am Nukleosom, welches die Symmetrieachse bestimmt und häufig verwendet wird, um die Position der Bindung von LH an ein Nukleosom zu beschreiben, wird 'Dyad' genannt. Für die Bindung von LH und Nukleosom zeigten vorherige Studien sowohl Bindung an der 'Dyad'-Position, als auch abseits des 'Dyad'. Ebenso wurden unterschiedliche Orientierungen des LH beobachtet, weshalb die Einflussfaktoren auf die Struktur von LH-Nukleosom Komplexen und die Dynamik der Bindung zwischen LH und Nukleosom auf molekularer Ebene nicht vollständig verstanden sind.

Das Ziel der hier beschriebenen Forschungsarbeit war es, ein Verständnis der Bildung von Chromatosomen auf atomarer Ebene zu erreichen. Die Analyse experimentell ermittelter Strukturen von LH-Nukleosom Komplexen zeigte, dass statt einer einzelnen 3D-Struktur ein Ensemble von Strukturen des LH-Nukleosom Komplexes existiert. Um die Verteilung dieser strukturellen Ensembles besser zu verstehen, wurden Normalschwingungsanalysen (NMA), Standard-Molekulardynamik-Simulationen (MD), beschleunigte Molekulardynamik-Simulationen mit erweitertem Sampling ('accelerated MD', AMD) und Simulationen Brown'scher Moleküldynamik (BD) mit LH-, Nukleosom- und Chromatosom-Systemen durchgeführt. MD und AMD Simulationen zeigten, dass die globuläre Domäne des LH (LH GD) in Lösung eine geschlossene Form bevorzugt. Bei Bindung an das Nukleosom wird die LH GD durch hydrophobe Wechselwirkungen mit der L-DNA des Nukleosoms in eine offene Struktur überführt. Zusätzlich schränkt die Bindung der LH GD die Flexibilität und Bewegungsrichtung der L-DNA ein. BD-Simulationen wiesen auf zahlreiche mögliche Chromatosomenkonfigurationen in Abhängigkeit von der Sequenz der LH GD und den L-DNA-Öffnungswinkeln hin. Diese Ergebnisse lassen darauf schließen, dass LH-Nukleosom-Bindung durch eine Kombination von konformationeller Selektion und 'induced fit'-Mechanismen vermittelt wird.

Weitere BD-Simulationen zeigten, dass Chromatosomen-Konfigurationen durch einzelne Punktmutationen in der LH GD beeinflusst wurden und sich zwischen unterschiedlichen Isoformen des LH unterschieden. Meine Ergebnisse deuten an, dass durch Austausch spezifischer einzelner

Punktmutationen zwischen verschiedenen LH GD Isoformen auch LH-Nukleosom-Konfigurationen unter diesen vertauscht werden können. Ähnliche Veränderungen der Chromatosomen-Konfiguration wurden als Ergebnis der Einführung posttranslationaler Modifikationen (PTMs) in der LH GD beobachtet.

Mittels BD-Simulationen berechnete ich weiterhin die Geschwindigkeitskonstante der Dissoziation (k_{off}) verschiedener LH-Mutanten von Chromatin und verglich diese mit zuvor beschriebenen Daten aus FRAP-Experimenten (fluorescence recovery after photobleaching). Die Ergebnisse der BD-Simulationen zeigen dabei für LH-Chromatin-Bindung verschiedener LH-Mutanten die gleichen relativen Tendenzen wie die in FRAP-Experimenten ermittelten Diffusionszeiten, die zur Wiederherstellung von 50% der Ausgangs-Fluoreszenzintensität nötig sind (t_{50}). Die Ergebnisse erlauben daher die Interpretation der FRAP-Daten im Sinne eines physikalischen Models von LH-Nukleosomen-Bindung.

Meine Arbeit liefert detaillierte Einblicke in Struktur, Dynamik und Kinetik der Chromatosomen-Bildung in Eukaryoten. Die hier gezeigten Ergebnisse können als Orientierungshilfe für weitere Experimente zur Aufklärung des Einflusses der Sequenz auf die LH-Nukleosomen-Bindung dienen.

Table of Contents

Abbreviations

AMD:	accelerated molecular dynamics
APBS:	Adaptive Poisson – Boltzmann Solver
BD:	Brownian dynamics
CENP-A:	H3-like centromeric protein
CH:	core histone
CL:	chemical cross-linking
CMD:	classical molecular dynamics
cryo-EM	cryo-electron microscopy
ECM:	effective charge methodology
ENM:	elastic network model
FRAP:	fluorescence recovery after photobleaching
FRET:	Förster resonance energy transfer
GD:	globular domain
GFP:	green fluorescent protein
gH1:	globular domain of linker histone (H1)
gH5:	globular domain of linker histone (H5)
gH5A:	open form of globular domain of LH (H5)
gH5B:	closed form of globular domain of LH (H5)
HADDOCK:	High Ambiguity Driven Protein-Protein Docking
ITC:	isothermal titration calorimetry
k_{off}:	dissociation rate constant
k_{on}:	association rate constant
L-DNA:	linker DNA
LH:	linker histone
MD:	molecular dynamics
MNase:	micrococcal nuclease
NC:	nucleosome core
N-DNA:	nucleosomal DNA
NMA:	normal mode analysis
NMR:	nuclear magnetic resonance
NUC:	nucleosome core particle
OH footprint:	hydroxyl radical foot-printing
PBC:	periodic boundary conditions
PCA:	principal component analysis
PDB:	Protein Data Bank

PRE:	paramagnetic relaxation enhancement
PTM:	post-translational modification
QM:	quantum mechanics
RMSD:	root-mean-square deviation
SAXS:	small angle X-ray scattering
SDA:	Simulation of Diffusional Association
t_{50}:	FRAP recovery half time
TIPnP:	transferable intermolecular potential n point
UHBD:	University of Houston Brownian Dynamics
WT:	wild type
X-ray:	X-ray crystallography

Introduction

In cells, biological processes like transcription, protein synthesis, replication, catalysis and signaling are regulated by selective intermolecular interactions. The strength and extent of these interactions are determined by the spatial arrangements of atoms in time. Understanding the time-dependent atomic-detail features of biomolecules is key to answering one of the fundamental questions in science: “What is the molecular basis of life?” (1). Since the establishment of molecular biology in the 20th century, significant methodological progress in biochemistry, genetics, microbiology and physics has allowed us to extend our knowledge about life (2). However, “How is ~2 m of DNA packed inside a human cell nucleus?” (3) is still a challenging question as the molecular details of DNA compaction are not fully understood.

In eukaryotic cells, DNA is wrapped around core histone proteins forming nucleosomes, which are the repeating units of the chromatin. The compaction of nucleosomes is governed by mediator proteins, called linker histones (LH), that bind to nucleosomes and form chromatosome complexes (4). Traditionally, the single DNA base pair defining the symmetry axis of the nucleosome core particle between linker DNA (L-DNA) arms is called the dyad point and used to refer to the position of LH binding to a nucleosome. Even though the crystal structures of the globular domain of a LH and of a nucleosome were determined in 1993 and 1997, respectively (5, 6), the first detailed structures of LH – nucleosome complexes have only been reported very recently (7–10). However, experimental studies on the structures of chromatosomes are not in full agreement, and two alternative LH – nucleosome binding modes (on- and off-dyad) have been reported. In the on-dyad binding mode, the globular domain of the LH interacts with both linker DNA arms and the nucleosomal DNA (N-DNA) on the dyad axis (7, 9). In the off-dyad binding mode, the globular domain of LH only interacts with N-DNA next to the dyad point and one of the L-DNAs (8, 10).

Various experimental approaches have been applied to understand the dynamics and structure of LH – nucleosome complexes from different perspectives. Among these methods, *micrococcal nuclease* (MNase) digestion (11, 12), chemical cross-linking (13, 14), hydroxyl radical footprinting (15) and fluorescence recovery after photobleaching (FRAP) (16) have provided information about LH – nucleosome binding in the cell. On the other hand, the structural information

obtained from these methods is at low resolution and not sufficient to understand the atomic details of chromatosome complex formation. X-ray crystallography (7, 9) and cryo-electron microscopy (cryo-EM) (10) were used to obtain atomic-level information on LH – nucleosome complexes. However, these methods are also limited, as they can only provide static structures and do not give further details on the dynamics of macromolecular interactions. Alternatively, both atomic resolution and dynamic features can be obtained by using NMR spectroscopy methods (8, 17). On the other hand, NMR spectroscopy signals are too complex to analyze for big biomolecular complexes and the method is generally only applicable to small biomolecules up to 80 kDa weight (18). Thus, current experimental methods are not able to provide a detailed understanding of the molecular determinants and dynamics of chromatosome formation and complementary methods are required.

The impact of computers on research has significantly increased since their principles were introduced by Alan Turing in the 1930s (19). Remarkably, on 11 May 1997, one of the most exciting events in history happened when a computer (Deep Blue) defeated the world chess champion Garry Kasparov (20). This event can be considered as a clear indication of the progress made in computing power, software and efficiency. Up to now, computing power has been increasing continuously following Moore's law (computer power doubles every two years) (21) and a new domain of science, computational biology, has evolved. Specifically, in the last four decades computational structural biology methods have significantly advanced (22) and the term "computational microscope" has been introduced (23).

Theoretically, all chemical and biological reactions can be analyzed at sub-atomistic level by using quantum mechanics (QM) simulations. Despite the availability of the theoretical framework of QM simulations, current computing power only allows us to analyze systems up to about 300 atoms at this level for very short time scales, making QM far from being applicable for most biomolecules (24). In order to overcome this issue, a simplified classical molecular simulation method that relies on solving Newton's equation of motion iteratively was developed, which is called molecular dynamics (MD) simulation. In MD, formation and breakage of chemical bonds is generally neglected. Such simplification enables all atom simulations of biomolecules to be performed on up to microsecond time scales (24). Thus, MD is a useful tool for obtaining information about the conformational sampling of biomolecules and has been applied to a wide variety of systems. Accelerated MD (AMD) simulations can help to solve the problem of limited conformational sampling in standard MD simulations by using boosted potential energies to overcome energy barriers (25). However, understanding the association events of biomolecules requires accessing even longer timescales. As a solution to this problem, Brownian dynamics (BD) simulations can be applied. In BD, the flexibility of biomolecules is usually not considered and rigid body structures are used in implicit solvent conditions which make it possible to study the formation of diffusional encounter

complexes of biomolecules (26). The combination of MD and BD methods is an effective approach to understand the association and dynamics of chromatosome formation at different scales.

Despite previous molecular simulation efforts on the nucleosome core particle (27), up to now, the molecular details of LH – nucleosome binding have not been investigated by a combined MD and BD approach. In this thesis, I aimed to complement experimental studies by using a combination of molecular modeling and simulation methods to understand chromatosome complex formation. The main research questions addressed are:

- What are the conformational states of the globular domain of the LH in solution? Which state is dominant?
- Do L-DNAs with different sequences have the same flexibility in nucleosomes?
- What are the mechanisms of LH – nucleosome binding? Is it induced-fit or conformational selection or both?
- How are the structures of the LH and nucleosome affected by chromatosome complex formation?
- Is there more than one chromatosome configuration? If so, what are the determinants of these structural ensembles?
- Do single point mutations and post-translational modifications (PTMs) of LHs affect chromatosome configuration?
- What are the kinetic parameters of LH – nucleosome association?

In order to answer these questions, molecular simulation methods including normal mode analysis (NMA), MD and BD were applied to LH, nucleosome and LH – nucleosome complex structures.

Thesis overview

This thesis consists of six Chapters:

In Chapter 1, the DNA packaging mechanism in eukaryotes is introduced by giving information on LH, nucleosome, chromatosome and chromatin structures from the literature.

In Chapter 2, the theoretical background to the molecular simulation tools used (NMA, MD and BD) is introduced.

In Chapter 3, MD and BD simulation results of LH, nucleosome and LH – nucleosome complex structures are provided. The determinants of LH – nucleosome binding are analyzed.

In Chapter 4, BD simulation results for the introduction of single point mutants and PTMs on LH - nucleosome binding are given. Isoform specific LH – nucleosome binding is also investigated.

In Chapter 5, BD simulation results for deriving mutant LH - nucleosome binding kinetic parameters are compared with experimental data.

In Chapter 6, a summary of the thesis is given. The limitations are discussed and complemented by future perspectives.

1

Packaging of DNA in eukaryotes

This chapter is in part adapted from the submitted perspective article “Towards an ensemble view of chromatosome structure: A paradigm shift from one to many” authored by Öztürk M. A., Cojocaru V. and Wade R. C.

In eukaryotes, DNA packing is achieved by an ordered protein – DNA complex formation mechanism and the structure of the compact DNA has an effect on whether the genetic material is used for transcription, replication or recombination (28). Thus, it is crucial to understand the DNA packaging mechanism. Initial compaction is maintained by formation of the complex of the 145 - 147 bp DNA and core histone proteins. This nucleosome complex is the basic repeating unit of the chromatin. With the binding of linker histone (LH) protein to nucleosome, the next level of compaction is achieved and chromatosome is assembled. Further packaging of thousands of chromatosomes forms chromatin. When chromatin is condensed, chromosome structure is formed (See Figure 1.1 for schematic representation). Additionally, chromatin is regulated in three different ways: ATP-dependent chromatin remodelers, post-translational modifications (PTMs) and exchange of histone variants (29). The last two mechanisms are investigated in this thesis. The structural units of DNA packaging with a focus on chromatosomes are explained in this chapter.

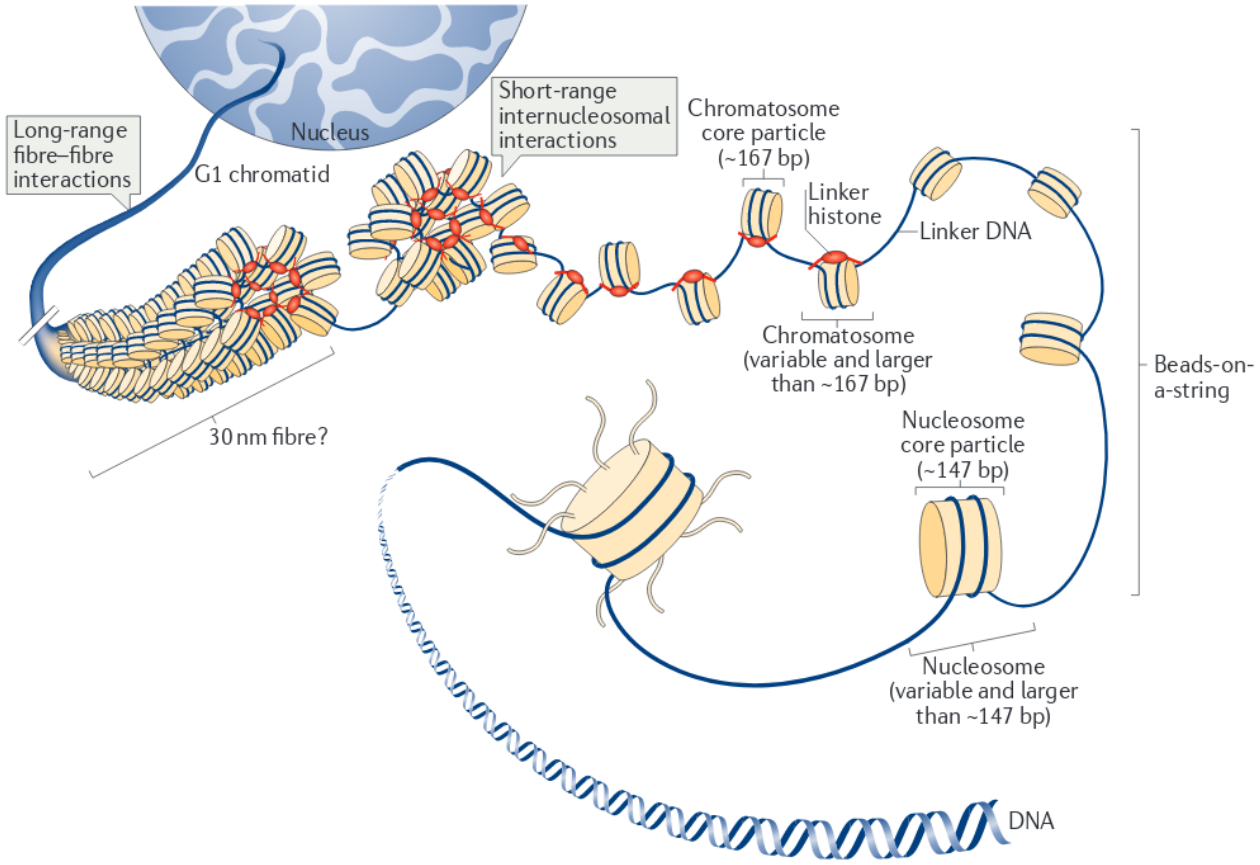


Figure 1.1: Schematic representation of DNA compaction in eukaryotes (Figure is obtained from Fyodorov D. V. et al. (4) and re-published with permission.)

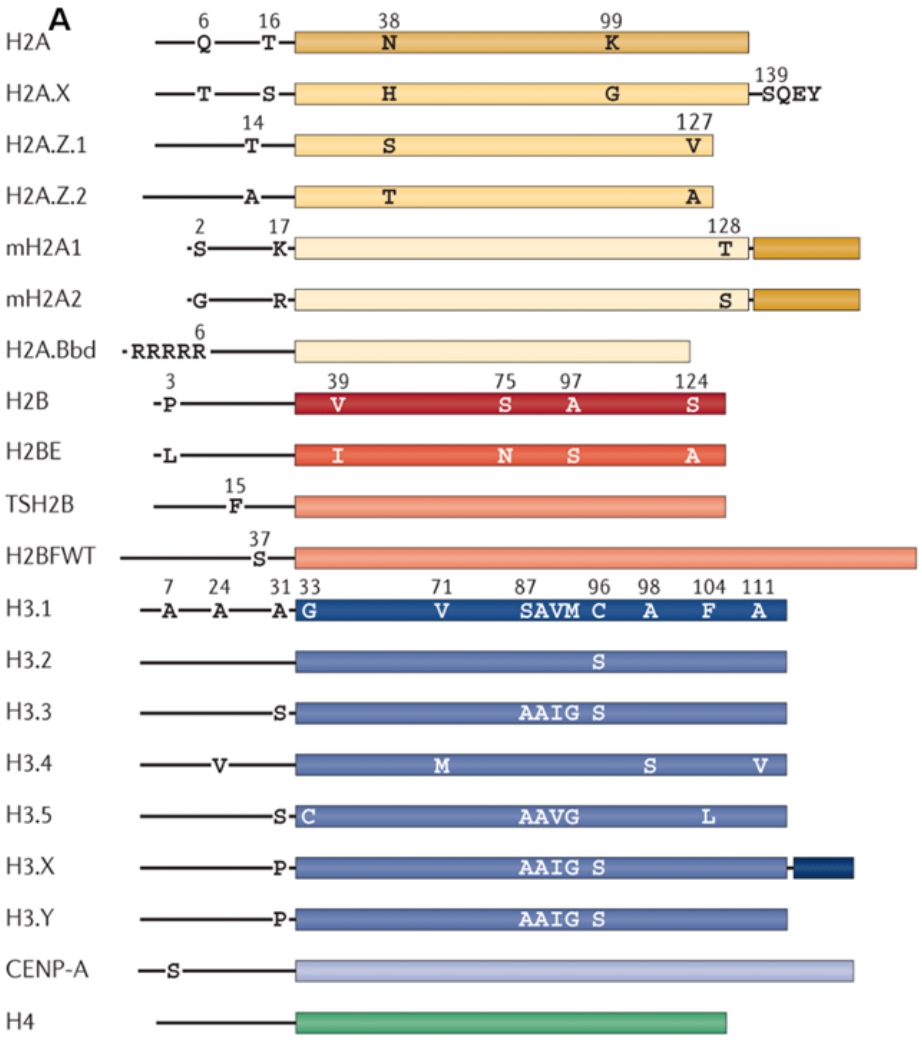
1.1 Chromatin

In eukaryotes, DNA is wrapped around core histone protein oligomers to form chromatin (30, 31) which further condenses into chromosomes during cell division. For cell function, it is crucial to dynamically compact about two meters of genetic material in such way that specific genes for transcription can be accessed when required (32). Despite more than 30 years of research, the mechanism of higher order chromatin compaction is not fully resolved (33, 34). In 1976, by using electron microscopy (EM), Finch and Klug showed that DNA forms 30-nm chromatin fibers and they named their chromatin model as “one-start helix” in which connected nucleosomes are folded as solenoids (35). Later in 1984, the “two-start helix” model for chromatin, in which nucleosomes interact with the following nucleosomes in a zigzag arrangement was suggested (36). Since then, the structure of chromatin has been investigated in various experimental studies by using X-ray crystallography, EM, cryo-EM and small angle X-ray scattering (SAXS) (37). Despite such efforts, determining a single 30-nm chromatin fiber structure is not possible as the variable linker DNA (L-DNA) length affects the solenoid or zigzag compaction and a 1 bp increase of the length of the L-DNA could cause 36° rotation of one nucleosome with respect to the neighbor (34, 38). Additionally,

in certain conditions both solenoid and zigzag structures can be present in the 30-nm chromatin fiber (33).

1.2 Nucleosome core particle

The nucleosome core particle in eukaryotes consists of 145 to 147 bp of nucleosomal DNA (N-DNA) wrapped around a histone octamer composed of two copies of each of the core histone proteins H2A, H2B, H3 and H4. In mammals there are various histone variants which introduces diversity in chromatin regulation. Recent knockdown and knockout studies and mutational analysis of histones indicate that histone variant regulation affects differentiation, proliferation, nuclear reprograming and meiosis functions of the cell (39). For example, H3-like centromeric protein (CENP-A) replaces H3 in the kinetochore and generates more flexible nucleosomes with reduced linker histone binding compared to H3 isoform nucleosomes (40). High expression levels of CENP-A are related with various cancer types (41).



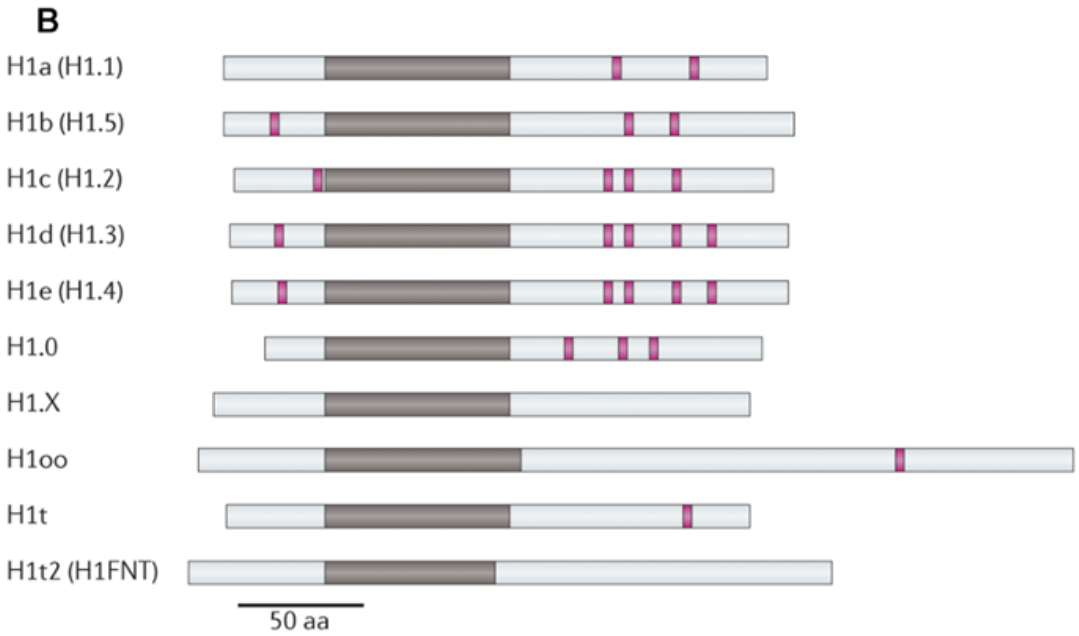


Figure 1.2: Human histone variants. **A-** Core histone variants of H2A (yellow), H2B (red), H3 (blue) and H4 (green) are shown. Color divergence is used to compare amino acid sequence of histones with the first entry of each histone family. Key amino acid differences and their positions are shown. Unstructured N- or C- termini are depicted in black lines. **B-** Human linker histone variants are shown. Globular domains are shown in dark gray and sites phosphorylated by cyclin-dependent kinases are shown in magenta. Unstructured N- or C- termini are depicted in gray (Figure is obtained from Maze I. et al. (39) and reprinted with permission.).

The first crystal structure of the nucleosome was reported in 1997 by Luger et al. (6). The structure of the nucleosome (Figure 1.2) is conserved from yeast to metazoans (28) with a disc-like structure of ~5.5 nm height and ~11 nm diameter.

The four core histones are around 11-15 kDa positively charged proteins and are highly conserved among eukaryotes. Core histones have very flexible N-terminal tails and short C-terminal tails. The H3/H4 dimer has lower affinity compared to H2A/H2B dimer in physiological conditions (7 vs 12 kcal/mol) (42). With the interaction of the H3:H3 interface, the H3/H4 dimer can self-associate to form a tetramer. Then, two H2A-H2B dimers join the complex to form the histone octamer (42).

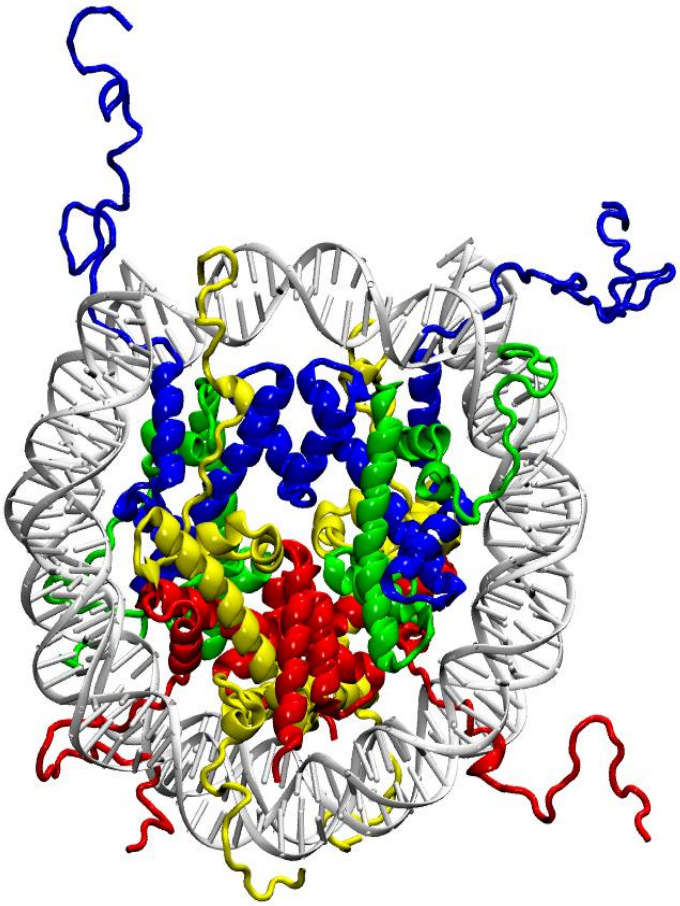


Figure 1.3: The X-ray crystal structure of the nucleosome core particle (PDB id: 1KX5 (43)) consisting of H2A (yellow), H2B (red), H3 (blue) and H4 (green) core histones. 147 bp DNA wrapped around core histones is shown in white.

1.3 Linker histone (LH)

In addition to the core histones, a linker histone (LH) protein, H1 or H5, can bind to the nucleosome between the two L-DNA arms to form a chromatosome complex (7, 13, 44, 45). LHs are composed of about 200 amino acid residues (Figure 1.3), and contain three distinct domains, a short (~40 residues) unstructured N-terminal tail, a relatively conserved globular domain (GD, ~80 residues) and a basic disordered C-terminal tail (~100 residues) (46). Previous studies have shown that even though the N- and C- terminal tails can affect the affinity and geometry of LH - nucleosome binding, they do not appear to affect the LH positioning (15). Furthermore, both the LH GD and the full length LH protect the same L-DNA from micrococcal nuclease digestion (47). Thus, the LH positioning on the chromatosome is mainly governed by the LH GD.

Recent research has shown that LH proteins have a range of functions, including roles in DNA replication, epigenetic regulation, genome stability and DNA repair (for a recent review see Fyodorov et al. (4)). Higher eukaryotes have a family of LH proteins consisting of a number of

variants, also referred to as subtypes, that have a relatively conserved GD and more variable N- and C- termini (48). It has been shown that LH variants can have different functions, tissue expression levels and DNA binding affinities (49–51). In mammals, there are seven standard H1 subtypes with varying sequence conservation, chromatin binding affinity and genomic distribution (48). H1 LH proteins have been shown to be essential for mouse development (52). For example, even though a single H1 isoform knock-out did not result in any significant phenotypic change, deletion of 3 isoforms was shown to be embryonically lethal (53–55). On the other hand, studies in unicellular eukaryotes, such as *Aspergillus nidulans* and *Tetrahymena thermophila* have indicated that knock-out of the sole H1 isoform is not lethal but can cause some genes to be up- or down-regulated (56, 57). Furthermore, it was previously reported that LHs behave as regulators of specific genes by affecting nucleosome spacing (55).

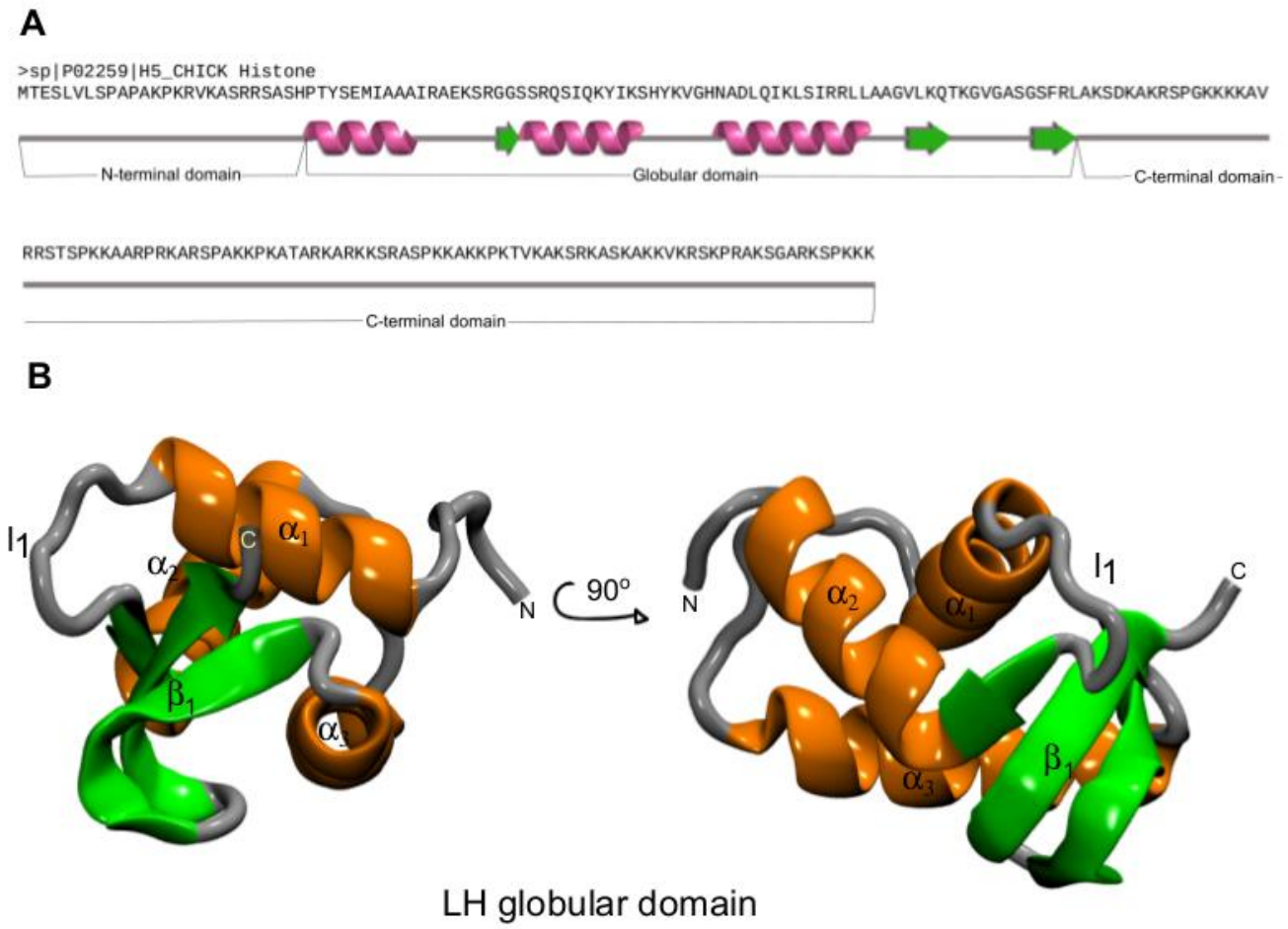


Figure 1.4: Sequence and structure of the *G. gallus* H5 **A-** Amino acid sequence of the *G. gallus* H5 is given with Uniprot accession number (P02259). The domains of the LH and secondary structure of the globular domain (GD) are shown below the protein sequence. **B-** Structure of the GD of the *G. gallus* H5 from PDB id: 1HST chain A (5). LH protein is shown in cartoon representation and colored according to secondary structure: α helices in orange, β sheets in green and unstructured regions in gray.

1.4 Chromatosome

Chromatosomes were first revealed by digestion of chromatin by a non-specific nuclease to consist of the nucleosome core, about 20 bp of L-DNA and one LH (58). The chromatosome can therefore be considered as a fundamental unit of the chromatin structure (59) and the determination of the three-dimensional structure of this subnucleosomal particle has been a longstanding goal.

In vitro reconstitution of nucleosomes requires certain conditions that are far from physiological conditions, such as 2M salt concentration, as well as suitable DNA and protein sequences (60). Obtaining chromatosomes in a form suitable for structure determination has been difficult. As can be seen in Table 1.1, the systems studied have a combination of DNA, core histone and LH sequences of different origins and DNA and protein constructs of different lengths. Moreover, the LH - nucleosome complexes were reconstituted and their structures determined under a range of environmental conditions, with different LH:nucleosome ratios, with different buffers and at different pH values and temperatures. Notably, the nucleosomes were reconstituted using salt dialysis against a gradually decreasing high salt buffer, the LH - nucleosome complexes were reconstituted by incubation at various ionic strength conditions, and the structural measurements were made at salt concentrations ranging from about 10 mM up to close to physiological ionic strength (Table 1.1). On the other hand, Schlick and colleagues showed that salt and LH concentration, L-DNA length, the presence of oligo-nucleosome systems, and synergistic folding of the LH C-terminal affect chromatin condensation and LH contacts with L-DNAs (61–63). Thus, the heterogeneity of the studied systems should be born in mind when considering the relevance of results with these *in vitro* systems for understanding chromatosome and chromatin structure in cell nuclei.

The nucleosome systems vary in the lengths of the L-DNA arms which each range from 10 to 30 bp. The first chromatosome structure solved (45) had a Widom 601 DNA sequence and core histone proteins from *Drosophila melanogaster*. A common component of the recent structural nucleosome studies is the synthetic 147 bp Widom 601 N-DNA sequence that wraps around the core histones and has a strong core histone octamer binding affinity (64). In structural studies, the choice of Widom 601 sequence, albeit unnatural, allowed researchers to obtain more stable nucleosomes (65). The first published report of the sequence of Widom 601 DNA was given in the study of Schalch et al. (66). A palindromic variant, Widom 601L, with higher core histone octamer affinity (L indicates that it was generated from the left half of the Widom 601 sequence), was also used (9, 67).

In the published studies of the structures of LH - nucleosome complexes, the core histones vary in origin as seen in Table 1.1. The core histones have flexible tails which are present in the sequences used in the experiments but often missing in the final structures determined. The extent

to which the flexible tails affect LH binding is unknown. On the other hand, Zhou et al. (45) reported that *D. melanogaster* H1 methyl groups are affected by paramagnetic relaxation enhancement (PRE) labeling of T119 in the H2A tail and that the disordered C-terminal tail of H2A folds upon LH binding. These results suggest that further research is required to understand the effects of the core histone tails on LH binding to the nucleosome. Experimentally, the LHs have been studied as full-length proteins and as globular domain constructs of varying lengths and, in some cases, with mutations to improve stability or switch key isoform residues. The N- and C- terminal domains are highly flexible and, therefore, their removal can be expected to facilitate crystallization.

1.4.1 Experiments to determine the structure of the LH - nucleosome complexes

The experimental methods for structure determination can vary in the level of detail and the amount of information that they provide, as well as the associated uncertainties, (for a recent review, see Mackay et al. (68).) For the first structure of a LH - nucleosome complex determined, Zhou et al. (45) mutated 4 residues of the *D. melanogaster* gH1 and obtained a more stable LH domain, similar to the *G. gallus* gH5. By using a gel shift assay and isothermal titration calorimetry (ITC), they showed that various mutant *D. melanogaster* H1 constructs (residues 37-132, 45-119, 37-211 and 37-256) have the same nucleosome binding affinities. The authors derived experimental constraints with NMR shifts and PRE for wild type (WT) and mutant *D. melanogaster* H1 binding to a nucleosome. The structure of the *D. melanogaster* gH1 was modeled by homology, based on the closed conformation of the *G. gallus* gH5 in the crystal structure (PDB id: 1HST, chain B) and a structural model of a LH GD-nucleosome complex was obtained by docking the LH GD and nucleosome with the HADDOCK program (69) using a small number of restraints derived from the combined experimental results. It should be noted that even though Zhou et al. (45) did their experiments with a Widom 601 N-DNA sequence, in their docking calculations they used the nucleosome structure with PDB id: 1ZBB whose DNA sequence is not Widom 601 but a palindromic sequence extracted from PDB id: 1KX5 (66). Later Zhou et al. (17) used a similar approach to study the binding of LHs from two different organisms.

A detailed model was only obtained when the first structure of a LH GD - nucleosome complex was determined by X-ray crystallography at 3.5 Å resolution (PDB id: 4QLC) (7). This model was supported by NMR data in the same publication. The first cryo-electron microscopy (cryo-EM) study (10) gave structures of chromatin fibers with 12*177 and 12*187 bp nucleosomes in the presence of full-length and WT *Homo sapiens* H1.4 with ~25 Å and 11 Å resolution, respectively. Both structures were in agreement with a zigzag two-start helix model for the 30 nm chromatin fiber. By averaging the densities of the central four nucleosomes in the 11 Å resolution map, Song et al. (10) deduced an off-dyad binding mode for H1. Although they proposed a specific orientation of H1 in the chromatosome, the low resolution of the electron density map means that other orientations

are also consistent with the data. This may reflect the variable and dynamic nature of chromatin due to molecular flexibility and variable DNA length, histone variants, and PTMs of the core histones and DNA.

Recently, Bednar et al. (9) have reported the first X-ray crystal structure for a complex containing a nucleosome with a full-length LH at 5.4 Å resolution (PDB id: 5NL0). Additionally, they applied cryo-EM, site-directed protein cross-linking and hydroxyl radical footprinting methods in the same study. For experiments, they used standard Widom 601 and palindromic Widom 601L DNA sequences, together with *H. sapiens* core histone and *Xenopus laevis* LH proteins. For deriving structural models, they used *X. laevis* core histone and LH protein sequences.

Table 1.1: Experimentally determined structures of LH – nucleosome complexes. The methods used and the sequences studied are given, followed by the details of the structural models derived from the experimental results.

	References	**Zhou et al., 2013**	**Song et al., 2014**	**Zhou et al., 2015**	**Zhou et al., 2016**	**Bednar et al., 2017**
Experimental details	**Structure Determination Methods**	NMR, PRE ITC, HADDOCK	Cryo-EM Ultracentrifugation	NMR, ITC, X-ray FRET, Ultracentrifugation	NMR, ITC HADDOCK Ultracentrifugation	Cryo-EM, X-ray OH footprint, CL
	147 bp N-DNA	synthetic DNA Widom 601	synthetic DNA Widom 601	synthetic DNA Widom 601	synthetic DNA Widom 601	synthetic DNA Widom 601 Widom 601L (*)
	# of L-DNA bp (**)	10+10 30+30	15+15 20+20	10+10	10+10 30+30	25+25
	Core Histones	*D. melanogaster*	*X. laevis*	*D. melanogaster*	*D. melanogaster*	*H. sapiens*
	Linker Histone (LH) (***)	*D. melanogaster* H1 (WT, 37-132, 45-119, 37-211, 37-256)	*H. sapiens* H1.4	*G. gallus* H5 (22-98, 24-98, 22-102 and 22-142), *D. melanogaster* (WT and 44-118), *X. laevis* H1	WT and mutant *G. gallus* H5 (24-98) *D. melanogaster* H1 (WT and 45-119) *X. laevis* H1.0 and *H. sapiens* H1.0	*X. laevis* H1.0b *H. sapiens* H1.5 (1-177 and 40-112)
	Environmental conditions for LH - nucleosome structural measurements (****)	Low IS pH 6.0 - 7.4	low IS, pH 8.0	NMR: low IS, X-ray, ITC, FRET: high IS pH 3.75 - 8.0	NMR: low IS ITC: high IS pH 7.4 - 8.0	Cryo-EM: Low IS X-ray: Medium IS pH 6.4
	Resolution (Å)	-	11 and 25	3.5	-	5.4
Structure details	**Basis for nucleosome structure**	Nucleosome from PDBs 1ZBB and 1KX5	Cryo-EM map fitted with nucleosome PDBs 1AOI and 1ZBB	Electron density fitted with nucleosome PDBs 4INM and 3MVD	DNA from PDB 4QLC	Electron density fitted with nucleosome PDB 3UT9

N-DNA	*H. sapiens* X chromosome α-satellite DNA Palindromic 147 bp	*H. sapiens* X chromosome α-satellite DNA Palindromic 146 bp	synthetic DNA Widom 601 147 bp	synthetic DNA Widom 601 147 bp	synthetic DNA Widom 601L (*) 145 bp
# of L-DNA bp (**)	10+10	15+15 20+20	10+10	0+0	26+26
Core Histones	*X. laevis*	*X. laevis*	*D. melanogaster*	None	*X. laevis*
CH Tails	Yes	No	No	No	No
Modeled LH sequence	*D. melanogaster* H1	*G. gallus* H5	*G. gallus* H5	*G. gallus* H5, *D. melanogaster* H1	*X. laevis* H1.0b
Modeled LH structure	From closed *G. gallus* LH PDB 1HST, chain B	From open *G. gallus* LH PDB 1HST, chain A	From closed *G. gallus* LH PDB 1HST, chain B	H5, from closed *G. gallus* LH PDB 1HST, chain B H1, from closed *G. gallus* LH PDB 1HST, chain B	From closed *G. gallus* LH PDB 1HST, chain B
LH Position	off-dyad	off-dyad	on-dyad	on-dyad off-dyad	on-dyad
PDB id of model			4QLC		5NL0

CH: Core histone, N-DNA: nucleosomal DNA, L-DNA: linker DNA, NMR: nuclear magnetic resonance, PRE: paramagnetic relaxation enhancement, ITC: isothermal titration calorimetry, HADDOCK: High Ambiguity Driven protein-protein DOCKing, Cryo-EM: cryo-electron microscopy, X-ray: X-ray crystallography, FRET: Förster resonance energy transfer, OH footprint: hydroxyl radical foot-printing, CL: chemical cross-linking. (*) The Widom 601L N-DNA sequence is the palindrome of the left half of the Widom 601 N-DNA sequence (**) The number of bp for each L-DNA arm is given, e.g. 10+10 denotes L-DNA1 with 10 bp and L-DNA2 with 10 bp. (***) Residue ranges are given in parentheses. (****) The ionic strength is classified as low: 10-20 mM, medium: ca. 50 mM and high: ca. 100-120 mM.

1.4.2 Position of the LH with respect to nucleosome

Biochemical experiments performed by *micrococcal nuclease (*MNase) digestion (11, 12), chemical cross-linking (13, 14), FRAP (16) and hydroxyl radical footprinting (15) have previously indicated either on- or off-dyad binding of LH proteins to nucleosomes. Additionally, molecular modeling and simulation studies resulted in various on- and off-dyad LH binding modes (16, 70–73). The recent structure determinations by NMR, X-ray crystallography and cryo-EM reported by Zhou et al. (7, 17, 45) Song et al. (10) and Bednar et al. (9), show both on- and off-dyad binding modes for the LH (See Figure 1.5 and Table 1.1).

In the off-dyad configuration, the LH GD interacts with only one of the L-DNAs and binds to the N-DNA adjacent to the dyad axis. Zhou et al. (45) showed that *D. melanogaster* H1, both as a full-length wild-type construct (residues 1-256) and in a truncated form (residues 37-211), binds off-dyad to a nucleosome with a 147 bp Widom 601 DNA sequence, two 10 bp L-DNAs and *D. melanogaster* core histones. Song et al. (10) showed that full-length wild-type *H. sapiens* H1.4 (residues 1-219) binds off-dyad to a chromatin composed of nucleosomes of 147 bp Widom 601 DNA sequence and two 15 or 20 bp L-DNAs wound around *X. laevis* core histones. However, it should be noted that the authors cross-linked LHs to the nucleosomes which may cause artifacts. Zhou et al. (17) reported that both full-length and truncated (residues 45-119) *D. melanogaster* H1 bind in an off-dyad position to a nucleosome of 147 bp Widom 601 DNA sequence and two 10 or 30 bp L-DNAs with *D. melanogaster* core histones.

In the on-dyad configuration, the LH interacts with both L-DNAs and the N-DNA on the dyad axis. Zhou et al. (7) found that *G. gallus* H5 (residues 22-98, 22-102, 22-142) binds on-dyad to a nucleosome of 147 bp Widom 601 DNA sequence and two 10 bp L-DNAs with *D. melanogaster* core histones. Most recently, Bednar et al. (9) reported that full length *X. laevis* H1, *H. sapiens* H1.5 (residues 1-77) and *H. sapiens* H1.5 (residues 40-112) bind on-dyad to nucleosomes with 147 bp Widom 601 DNA sequence and two 25 bp L-DNAs with *H. sapiens* core histones.

1.4.3 Orientation of the LH - nucleosome complexes

The experimentally derived structures not only show two positions of the LH on the nucleosome – on- and off-dyad – but also show different orientations with the l_3 loop of the LH interacting with either N-DNA or L-DNA, see Figure 1.5. In several cases, the experimental data can be fit with more than one orientation of the LH GD, i.e. its orientation cannot be unambiguously defined from the experimental results. Computational docking can help to identify the preferred orientation, for example, as applied by Zhou et al. (7, 17). However, in such efforts, the sequences of the modeled structures should ideally be exactly the same as the sequences used experimentally, which has not always been the case (Table 1.1). Computer simulations can also provide insights into

the mechanism of association and previous Brownian dynamics simulations of the association of gH5 to a nucleosome showed off-dyad LH-nucleosome binding (73).

Consideration of the positions occupied by the flexible LH N- and C-termini also serves to limit the possible orientations a LH can adopt on a nucleosome. For example, the LH configuration proposed by Zhou et al. (45) has a very close contact between the C-terminus of the H1 and L-DNA, which may not represent the full length LH system *in vivo*. Recently, the cryo-EM structure of Bednar et al. (9) revealed that the C-terminal domain of the LH localizes on one of the L-DNAs and introduces an asymmetry in the structure of the Widom 601 nucleosome.

Zhou et al. (7, 45) indicated that single residue mutations in the LH GD can significantly affect the LH – nucleosome binding affinity. Further experimental and computational analysis with mutant LHs is necessary to understand whether point mutations lead to positional or orientational shifts of the LH GD with respect to the nucleosome. This aspect is important as, generally, experiments on LH - nucleosome complexes are carried out with a mutant LH and care is therefore required in interpretation of the data with respect to wild type or post-translationally modified LHs.

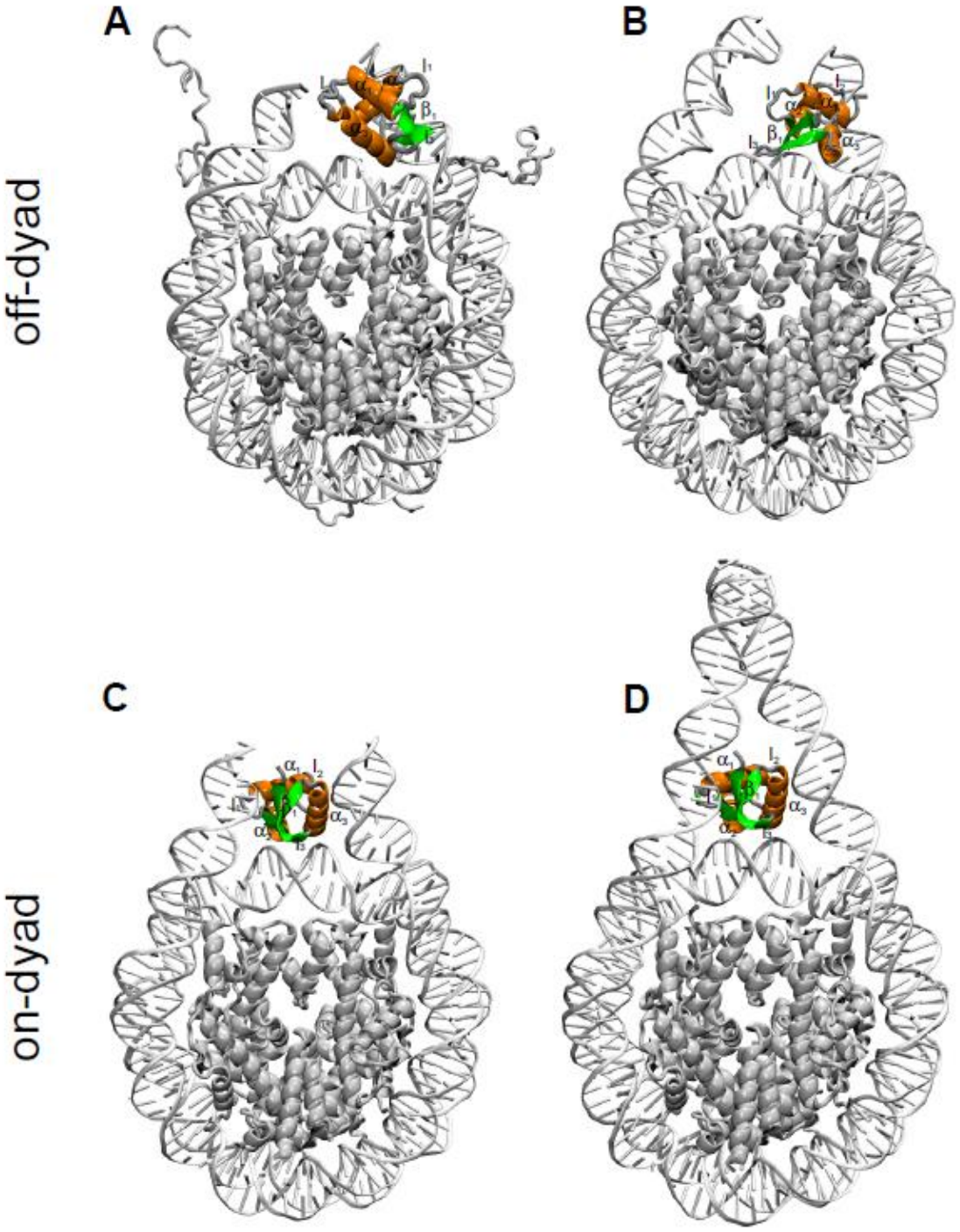


Figure 1.5: Four recently determined 3D structures of LH - nucleosome complexes. **A-** Off-dyad binding of *D. melanogaster* gH1 to a nucleosome, as reported by Zhou et al. (45) (PDB file provided by Yawen Bai.) **B-** Off-dyad binding of *H. sapiens* H1.4 GD to a nucleosome as reported by Song et al. (10) (Structure provided by Ping Zhu.) **C-** On-dyad binding of *G. gallus* gH5 to a nucleosome as reported by Zhou et al. (7), PDB id: 4QLC. **D-** On-dyad binding of *X. laevis* gH1 to a nucleosome as reported by Bednar et al. (9), PDB id: 5NL0. LH proteins are shown in cartoon representation and colored according to secondary structure: α helices in orange, β sheets in green and unstructured regions in gray. DNA is shown in light gray and core histones are shown in dark gray.

1.4.4 Evidence for LH - nucleosome structural ensembles

As summarized in Table 1.1, different experimental studies indicated on- or off-dyad binding of the LH to the nucleosome. A detailed analysis of the previous studies indicates that instead of a single chromatosome configuration, there are structural ensemble of chromatosomes. Below, experimental support to this hypothesis is introduced.

Zhou et al. (45) demonstrated the first systematic approach to combining various experimental methods for determination of the structure of the LH – nucleosome complex. Apart from using various lengths and mutants of *D. melanogaster* LH, they showed that the construction of nucleosomes with H2A.Z core histones resulted in an undetectable level of LH binding as measured by ITC. This suggests that, depending on the composition of the core histones of the nucleosome, there could be various LH – nucleosome binding affinities and different chromatosome ensembles. Song et al. (10) reported that the tetranucleosomal units of the 12mer structure have an interaction of the N-terminus of H4 and the acidic patch of the H2A-H2B dimer, which was suggested to be the reason for the twist between the tetranucleosomal units. Such a twist would allow a wide range of L-DNA angles to be present in higher order nucleosome structures and multi-nucleosome units in the higher order chromatin structure could allow various chromatosome ensembles, as also shown by mesoscale simulations of the chromatin (63). Zhou et al. (7) reported the first crystal structure of a chromatosome containing *G. gallus* gH5. The clear observation of the side chains of R47, R73, R74 and R94 implied that gH5 makes stable interactions with the dyad N-DNA and both L-DNA arms. This suggests that specific residues are responsible for the affinity to the nucleosome and the stability of the chromatosome structure. Similarly, the authors showed that gH5 undergoes conformational rearrangement upon nucleosome binding and they reported that the gH5 l_2 loop is more stable than the l_3 loop in its free form.

Zhou et al. (7) reported that, relative to the LH-bound state, the L-DNA arms of the free nucleosome are ~10 Å further apart, which would affect the higher order chromatin structure and dynamics. Furthermore, in the same publication, the authors applied sedimentation assays on 12*177 bp nucleosomes with *D. melanogaster* full length H1 and gH1, *X. laevis* H1 and *G. gallus* H5 and gH5. They found that the gH5 has a ~6S higher sedimentation coefficient compared to *D. melanogaster* gH1 in nucleosome arrays and this could be an indication of different nucleosome complexation mechanisms for the respective LHs. The authors also mention that NMR analysis showed that the H1x isoform has an α_1 helix two helical turns longer than that of gH1, and that this could lead to a specific nucleosome binding mechanism.

Lastly, Bednar et al. (9) showed that LH tails introduce an asymmetry into higher order chromatin structure as the C-terminal tail of the LH only interacts with one of the L-DNAs in the Widom 601 nucleosome. This feature could affect the accessibility of the nucleosomes for LH binding in the chromatin structure and could facilitate or block LH binding to certain conformations.

Interestingly, in addition to cryo-EM and crystal structure determination, the authors conducted hydroxyl-radical footprinting analysis and showed that both *X. laevis* full-length H1 and gH1 have similar symmetric footprints on the core of the DNA, indicating that the LH tails have a limited effect on the LH binding site of the chromatosome. Furthermore, hydroxyl-radical footprinting variations of LH – nucleosome binding also support the availability of chromatosome ensembles rather than a single chromatosome structure. In order to exclude the off-dyad LH binding in solution, the authors conducted hydroxyl-radical footprinting experiments on nucleosomes that lack both the L-DNAs and observed similar DNA protection patterns as observed for the full nucleosome particle. Finally, the authors pointed out that the PTMs of the LH could change the electrostatic potential of the protein and could result in regulation of chromatin structure. These recent studies show that different experimental approaches can be combined to understand the structure and dynamics of the chromatosome. Particularly, solution methods, like hydroxyl radical footprinting indicate that instead of a single chromatosome structure, the LH – nucleosome complex exists as structural ensemble of chromatosomes.

1.5 Post-translational modifications of chromatin

Chromatin is not static and its structure is regulated by covalent post-translational modifications of histone proteins and methylation of DNA, which impact the formation of euchromatin (open state) or heterochromatin (closed state). These mechanisms are called epigenetic regulation. DNA methylation is not the focus of this thesis and further information can be found in the reference (74).

In 1964, Allfrey et al. reported the first study of histone methylation and acetylation (75). Since then, further core histone PTMs of phosphorylation, ubiquitination, and sumoylation have been reported (76, 77). Most PTMs occur in positively charged termini of histones (see Figure 1.6). As the flexible histone tails are in contact with neighbor nucleosomes, their PTMs impact the nucleosome compaction (77). Core histone PTMs will not be investigated in this thesis and further information can be found in other references (78–80).

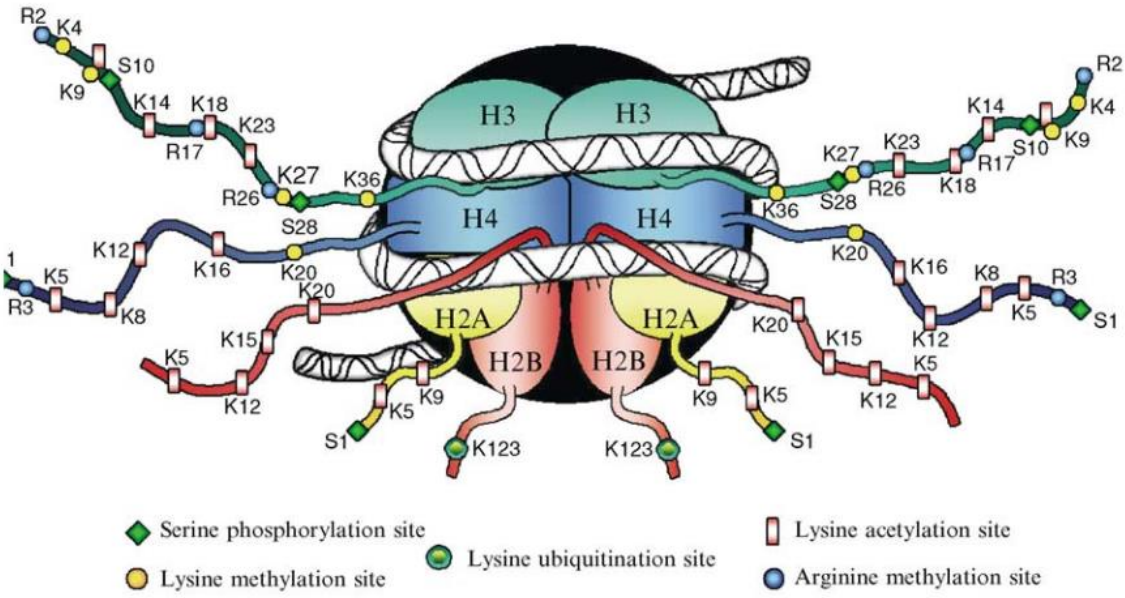


Figure 1.6: PTMs of core histone tails. (Figure obtained from Wood A. et al. (76) and reprinted with permission.)

The first PTM of a LH was reported in 1972 (81). Since then, many studies have shown that LHs can have methylation, acetylation, ADPribosylation, ubiquitination, formylation and PARylation PTMs (82–93). Izzo and Schneider recently extensively reviewed human and mouse H1 PTMs (94). They reported that H1 phosphorylation can have opposing effects on chromatin condensation. Horn et al. suggested that H1 phosphorylation may regulate ATP-dependent chromatin remodeling enzymes and thus impact chromatin compaction (95). Furthermore, high H1 phosphorylation levels are linked with DNA repair (96), apoptosis (97), cellular aging (98) and cancer events (99). H1 methylation is also associated with heterochromatin organization (100) and cell cycle regulated chromatin binding (87). However, although many sites of variant specific PTMs have been characterized, the phenotypic impact of individual LH PTMs is often unknown (101).

2

Methodological background

In cells, there are various molecular association events that contribute to processes such as transcription, protein synthesis, replication, catalysis and signaling to maintain the function. These events have been investigated by different experimental techniques of chemistry, physics and molecular biology. However, the complexity of these systems makes it hard to understand all of their molecular details. Thus, complementary methods are required. Recently, significant progress has been made in computing power and molecular simulation methods are becoming more popular to investigate the dynamics and kinetics of biomolecules (Figure 2.1). In normal mode analysis (NMA) harmonic oscillators are used to explore the dynamics of the biomolecules. In molecular dynamics (MD) method, Newton's equation of motion is numerically solved to obtain the dynamics of the system for all atoms. In Brownian dynamics (BD) methodology, the solvent is considered implicitly and longer time scale simulations are possible by simulating diffusion of rigid solute structures.

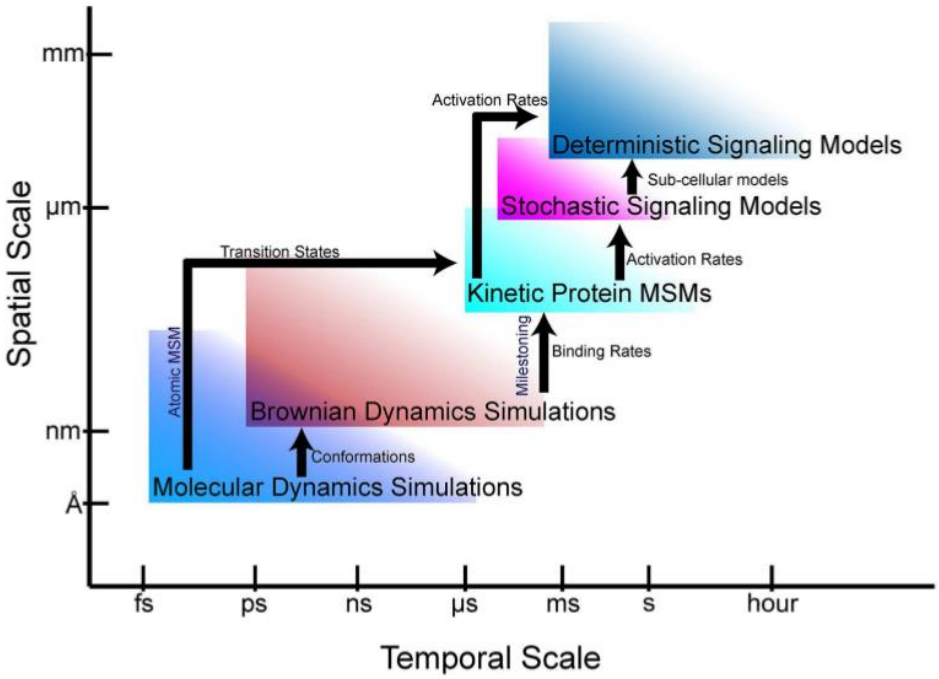


Figure 2.1: Spatial and temporal scales of various simulation methods. Information relation between techniques is shown by arrows. (Figure is obtained from Boras B.W. et al. (102) and re-published under Open Access CC BY license.)

In this thesis, NMA is used for determining the large amplitude motions of the nucleosome linker DNA arms. MD is used for conformational sampling of the nucleosome, linker histone and LH - nucleosome complex. BD is used to determine the diffusional encounter complex structures and kinetics of LH - nucleosome binding. The theoretical background of these methods is introduced below.

2.1 Normal mode analysis

In NMA, atoms are connected via springs and the eigenvectors of the Hessian matrix are calculated for the harmonic potential of a local minimum. Even though all atom representation can be used, usually only Cα atoms are considered for connecting the atoms via springs (see Figure 2.2). Low-frequency modes are not dependent on the force field used as they are only dependent on the connectivity (103). In NMA, if all atoms are treated a molecule with N atoms has 3 degrees of freedom for each atom. Subtraction of 3 rotational and 3 translational degrees of freedom of the whole molecule results in 3N-6 normal modes.

If Ri and Rj are the generalized coordinates around the minimum energy state $Rmin$, the Hessian matrix K of the potential energy E in Eq. (2.1) below gives the normal modes.

$$K_{ij} = \left[\frac{\partial^2 E}{\partial R_i \, \partial R_j}\right]_{R_{min}} \tag{2.1}$$

Modes represent the directionality of the movement but not the amplitude. High-frequency modes are related with local molecular movements, whereas low-frequency modes determine the motion of the domains, which are usually related with the function of proteins and/or nucleic acids. NMA is mainly used to investigate the large amplitude movements on different scales.

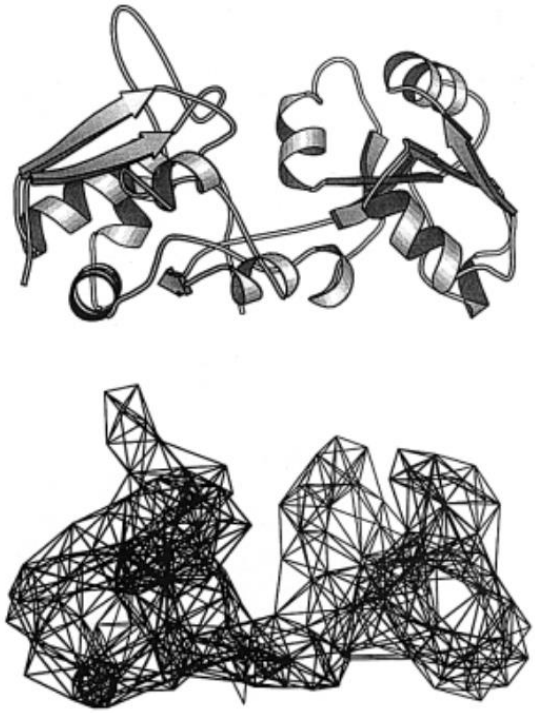


Figure 2.2: Elastic network model of lysine-arginine-ornithine binding protein. Model is generated by connecting Cα atoms closer than 8 Å. (Figure is obtained from Tama F. et al. (104) and re-printed with permission.)

In this thesis, the NOMAD – Ref web server is used for the NMA calculations of the nucleosome structure (105). In NOMAD – Ref, elastic network model (ENM) is used. In this model, for the coordinates a and b, a Hooken potential is applied in a certain cut-off to replace the empirical potential. The potential between atoms a and b is defined by:

$$E\,(r_a, r_b) = \frac{C}{2}\left(\lceil r_{a,b} \rceil - \lceil r^o_{a,b} \rceil\right)^2 \tag{2.2}$$

Here, C is the spring constant assumed to be the same for all interacting pairs. $r_{a,b}$ and $r^o_{a,b}$ are instantaneous and equilibrium distances, respectively.

2.2 Molecular dynamics (MD)

2.2.1 Molecular mechanics theory

The motion of atoms can be analyzed by Newtonian dynamics principles with the Born-Oppenheimer assumption, in which movement of the atomic nuclei can be separated from electrons. As such, in molecular mechanics, the dynamics of atoms can be defined by using empirical parameters, called force field, by using the following potential energy function E:

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{vdW} + E_{Coulomb} \quad (2.3)$$

Here, the potential energy is the sum of the bond energies, that are bond stretching (E_{bond}), bond angle bending (E_{angle}) and dihedral torsion ($E_{dihedral}$) as well as non-bonded energies, van der Waals potential (E_{vdW}) and Coulomb potential ($E_{Coulomb}$).

E_{bond} and E_{angle} terms are calculated by using springs and Hooke`s law (Eq. (2.4, 2.5)),

$$E_{bond}(l) = \frac{k_s}{2}(l - l_0)^2 \quad (2.4)$$

Here, k_s is the force constant for bond stretching, l_0 is the equilibrium bond length and l is the actual bond length.

$$E_{angle}(\theta) = \frac{k_b}{2}(\theta - \theta_0)^2 \quad (2.5)$$

Here, k_b is the force constant for bond angle bending, θ_0 is the equilibrium bond angle and θ is the actual bond angle.

Dihedral torsion energy ($E_{dihedral}$) is calculated with the following formula:

$$E_{dihedral}(\omega) = \frac{E_n}{2}[1 + \cos(n\omega - \gamma)] \quad (2.6)$$

Here, ω is the dihedral angle, E_n is the height of the energy barrier, n is the periodicity of the potential function and γ is the phase shift angle.

The van der Waals potential (E_{vdW}) in Eq. (2.3) is defined by the Lennard Jones potential as:

$$E_{vdW}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.7)$$

Here, ε_{ij} is depth of the potential well, r_{ij} is the distance between atoms i and j, σ_{ij} is the collision diameter.

Coulomb potential ($E_{Coulomb}$) in Eq. (2.3) is defined by:

$$E_{Coulomb} = \sum_{i=1}^{N_a} \sum_{j=i+1}^{N_b} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \quad (2.8)$$

Here, N_a and N_b are the number of partial charges, q_i and q_j are the partial charges on atoms, ϵ_0 is the dielectric permittivity of the vacuum, ϵ_r is the dielectric permittivity of the solvent and r_{ij} is the distance between atoms.

Parameters for force fields are derived from experiments of small biomolecules. Commonly used force fields in biomolecular simulations are AMBER (106), GROMOS (107) and CHARMM (108). In this thesis, all-atom AMBER force field (109) modified for DNA (ff99) (110), protein (ff99SB) (111) and DNA backbone (parmbsc0) (112) are used.

2.2.2 Energy minimization

The potential energy of a system is defined by the interaction of all atoms with each other. The potential energy is dependent on the coordinates of the atoms that generate the potential energy surface. For molecules, there can be many energy minima and transitions occur between them. Minimization is finding the coordinates of the system that has a minimum energy. For a system in a minimum state of energy E and Cartesian coordinates of the atoms x_i, the first derivative of E is zero and the second derivatives are positive:

$$\frac{\partial E}{\partial x_i} = 0; \frac{\partial^2 E}{\partial x_i^2} > 0 \quad (2.9)$$

In biomolecular simulations, energy minimization is usually applied before the MD simulations and the most commonly used minimization methods are steepest descent and conjugate gradient (113). In the steepest descent method, minimization is achieved by moving in the direction of the net force, similar to walking straight downhill. For this purpose, orthogonal gradients and direction of steps are used. However, this method can have an oscillatory behavior in narrow energy valleys. In the conjugate gradient method, each gradient is orthogonal to previous gradients and each direction is conjugate compared to previous directions, which solves the oscillation problem.

2.2.3 MD simulations

In MD simulations, the coordinates of the atoms are determined by using Newton's equation of motion for the forces calculated from the previous step. For the given initial position r and velocity v of each particle i, the force F can be calculated from the potential energy E:

$$F_i = \frac{\partial E}{\partial r_i} \quad (2.10)$$

Then acceleration of each particle, a is calculated by,

$$a_i = \frac{F_i}{m_i} \tag{2.11}$$

where m_i is the mass of each atom. Velocities and coordinates of each step can be determined by using the following algorithm:

1- Initial $(t = 0)$ position r_i and velocity v_i is given for $i = 0$ and a short time step Δt is chosen
2- Force and acceleration are calculated by using Eq. (2.10) and Eq. (2.11)
3- Atoms are moved $r_{i+1} = r_i + v_i\,\Delta t\, + \frac{1}{2} a \Delta t^{\,2}$
4- Boundary conditions, temperature and pressure control is applied
5- Time and iteration step increased $t = t\, +\, \Delta t$ and $i = i\, + 1$
6- Go to Step 2

The most common method to integrate Newton's equation of motion is the *Verlet* algorithm (114). In *Verlet* method, Taylor expansion approximated positions and dynamic properties are used. In order to calculate the new positions $r_{t+\,\Delta t}$ at time $t\, +\, \Delta t$, the positions and accelerations of time t and the position of the previous step $r_{t-\,\Delta t}$ are used:

$$r_{t+\,\Delta t} = \, r_t + \, \Delta t\, v_i + \frac{1}{2} a_t \Delta t^{\,2} + \cdots \tag{2.12}$$

$$r_{t-\,\Delta t} = \, r_t - \, \Delta t\, v_i + \frac{1}{2} a_t \Delta t^{\,2} - \cdots \tag{2.13}$$

The sum of these equations gives:

$$r_{t\,+\Delta t} = \, 2r_t - \, r_{t-\,\Delta t} + a_t \Delta t^{\,2} \tag{2.14}$$

And the velocity is calculated by,

$$v_{t\,+\Delta t} = [r_{t\,+\Delta t} - \, r_{t-\,\Delta t}]/2\Delta t \tag{2.15}$$

The time step of the integration is usually chosen as 1 fs (femtosecond), which is in the time scale of the fast vibrations of bonds to hydrogen atoms. Usually bonds with hydrogens are not in focus as their dynamics has a little effect on large scale movements. By using algorithms like SHAKE (115), it is possible to constrain bonds with hydrogens. This approach allows a bigger (2 fs) time step to be used and thus the computational efficiency is increased.

With the iterative MD process, a time-dependent simulation coordinate trajectory is obtained and additional parameters like pressure, temperature and energy. Commonly used biomolecule MD simulation software packages are AMBER (116), NAMD (117) and GROMACS (118). In this thesis, the AMBER and NAMD software packages are used.

2.2.3.1 Constant pressure and temperature in MD simulations

In mathematical physics, the probability distribution of a system's states is defined as statistical ensemble. In MD simulations, generally a classical algorithmic framework to sample the motions of a molecule is used in constant number of particles, constant volume and constant energy (NVE, microcanonical) ensemble. However, biological processes usually occur at constant number of particles, constant volume and constant temperature (NVT, canonical) ensemble and / or constant number of particles, constant pressure and constant temperature (NPT, isothermal-isobaric) ensemble. In order to compare the simulation results with experiments, MD simulations need to be maintained in these ensembles.

In NVT MD simulations, the most commonly used method to maintain temperature was introduced by Berendsen (119). In an unconstrained system, temperature is related with the time average of the kinetic energy as:

$$< E_k > = \frac{3}{2} N k_B T \tag{2.16}$$

Here, E_k is the kinetic energy, k_B is the Boltzmann constant and T is the temperature. Berendsen (119) used an external heat bath to keep the temperature constant. The atom velocities of the system are reassigned as follows:

$$\frac{dT(t)}{dt} = \frac{(T_{bath} - T(t))}{\tau} \tag{2.17}$$

Here, bath and the system are coupled with the coupling parameter τ.

In NPT MD simulations, the pressure is controlled by changing the volume of the simulation cell. Isothermal compressibility, κ is used to determine the amount of volume fluctuation.

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.18}$$

The most commonly used NPT methods are Nosé - Hoover Langevin piston method (120) and Berendsen barostat (119). In this thesis, both the NVT and the NPT ensembles are used in MD simulations.

2.2.3.2 Periodic boundary conditions (PBC)

In MD simulations, it is aimed to obtain macroscopic properties of the system which requires avoiding boundary effects. In PBC, an atom exiting from the simulation box re-enters from the opposite site. The box shapes used in MD that allow translational operations in all dimensions are: cube, hexagonal prism, truncated octahedron, rhombic dodecahedron and elongated dodecahedron (121). However, there are also some limitations of PBC. The fluctuations that have a bigger

wavelength than the cell size cannot be observed. Additionally, a sufficiently large box size should be used to avoid artificial long-range interactions with the image molecule.

2.2.3.3 Non-bonded cutoff

Normally, in MD simulations, calculating all interactions between N atoms requires $O(N^2)$ time (122). However, this is computationally very expensive. As the non-bonded interactions would be very limited beyond a certain atom pair distance, they are either neglected or approximated by using a reaction field, particle mesh Ewald or fast multipole methods (123). Such approaches can reduce the computational time to $O(N)$. Determination of the cutoff should be done carefully as there could be artificial long-range interactions due to periodic boundary conditions. In this thesis, the particle mesh Ewald method (124) with a 10 Å cutoff is used.

2.2.3.4 Water models

Despite methodological advances, MD simulations mostly can only reach up to microsecond timescales, which can be shorter than biologically significant molecular movements. One important aspect that can reduce the computational complexity is the way the solvent, usually water for biomolecules, is treated. In MD, water can be treated by implicit or explicit models. In implicit models, the number of particles is reduced by the continuum approximation of the discrete solvent. Further computational efficiency is gained by faster conformational space sampling. The most commonly used implicit solvent models are Poisson – Boltzmann and generalized Born (125).

Proper consideration of water is important in MD simulations as it determines the hydrophobicity of proteins and Coulomb screening between protein charges. Several water models are developed. The most commonly used water models are transferable intermolecular potential n point models (TIPnP) from Jorgensen lab and extended simple point charge model (SPC/E) from Berendsen lab (126). In this thesis, the TIP3P water model is used (127).

2.2.3.2 Accelerated molecular dynamics (AMD)

In classical MD simulations of biological systems, it is often not possible to observe all interesting properties of the system as it requires long time scales to overcome the local energy barriers of the potential energy, which a molecule can be trapped in. In order to overcome such energy barriers and increase the sampling of the molecule's conformation, a non-directional bias can be added to the energy function to increase the probability of sampling of different conformations like in AMD (128). This approach helps to access millisecond events in shorter simulation times (25). In this thesis, a modified $V^M_{dihedral}$ potential is applied as follows.

$$V^M_{dihedral} = \begin{Bmatrix} V_{dihedral} & if \;\; V_{dihedral} \geq E_{dihedral} \\ V_{dihedral} + \Delta V_{dihedral} & if \;\; V_{dihedral} < E_{dihedral} \end{Bmatrix} \quad (2.19)$$

Here, $V_{dihedral}$ is the intrinsic dihedral potential, $E_{dihedral}$ is the reference potential and $\Delta V_{dihedral}$ is the boost potential. The boost potential in Eq. (2.19) is calculated as follows with the α acceleration factor:

$$\Delta V_{dihedral} = \frac{(E_{dihedral} - V_{dihedral})^2}{\alpha + (E_{dihedral} - V_{dihedral})} \quad (2.20)$$

In this thesis, AMD is applied to obtain higher conformational sampling of linker histone, nucleosome and chromatosome particles.

2.3 Brownian dynamics (BD)

Brownian motion is named after Robert Brown who first identified the motions of small particles in fluids by using light microscopy. Brownian motion can be simulated with Brownian dynamics (BD) simulations, where particles move due to the forces generated by the stochastic collisions with solvent molecules (26).

In BD, for the given time step Δt, the diffusive 3D translational displacement Δr of a particle is determined by:

$$\langle \Delta r^2 \rangle = 6D\Delta t \quad (2.21)$$

Here, D is the translational coefficient and for spherical objects it can be calculated by using Boltzmann constant k_b, absolute temperature T, solvent viscosity η and radius of the particle a,

$$D = \frac{k_b T}{6\pi\eta a} \quad (2.22)$$

Ermak and McCammon developed an algorithm (129) to simulate the association of proteins by considering rotational and translational motions.

The following formula is used to calculate the translational displacement:

$$\Delta r = \frac{D\Delta t}{k_b T} F + R \quad (2.23)$$

Here F is the force acting on the molecule and R is the random displacement satisfying the conditions $\langle R \rangle = 0$ and $\langle \mathrm{R}^2 \rangle = 6D\Delta t$.

For the rotational displacement the following formula is used:

$$\Delta\phi = \frac{D_R \Delta t}{k_b T} T + \Theta \quad (2.24)$$

Here D_R is the rotational diffusion coefficient, T is the torque and Θ is a random rotational angle satisfying $\langle \Theta \rangle = 0$ and $\langle \Theta^2 \rangle = 6D_R\Delta t$.

It should be noted that typically in BD the molecules are treated as non-flexible rigid bodies. However, this is an approximation of biomolecules in solution and association events can occur with induced fit and/or conformational selection which requires flexibility of molecules. On the other hand, this simplification allows to investigate the biomolecular association at greater time scales, which is very difficult to access in molecular dynamics simulations. These principles can be found incorporated in a software to simulate the interaction and association of biomolecules.

2.3.1 Simulation of Diffusional Association (SDA) software

Depending on their interaction energies, biomolecules can form diffusional encounter complexes which are close to their bound state. By using rigid structures and an appropriate force field, BD simulations can be applied to understand the interaction energies and kinetics of biomolecular association. For this purpose, the SDA software (SDA7, http://mcm.h-its.org/sda7/, version 7) is developed (130). Theoretical background of the SDA7 is introduced in the following sections:

2.3.1.1 Interaction energies

In the SDA software, electrostatic, electrostatic desolvation and non-polar desolvation interaction energies are used for simulating association of two solutes (130). The following formula is used to calculate the total interaction energy between solutes:

$$\Delta G = \frac{1}{2}\sum_{i}\phi_{el_1}\,q_{i_2} + \frac{1}{2}\sum_{i}\phi_{el_2}\,q_{j_1} + \sum_{i}\phi_{edesolv_1}\,{q_{i_2}}^2 + \sum_{i}\phi_{edesolv_2}\,{q_{j_1}}^2 + \sum_{m}\phi_{npdesolv_1}\,SASA_{m_2} + \sum_{n}\phi_{npdesolv_2}\,SASA_{n_1} \quad (2.25)$$

Here, the first two terms are for the electrostatic interaction energy, the second and third terms are for the electrostatic desolvation energy, and the last two terms are for the non-polar desolvation energies. q_i is the net charge, $\phi_{edesolv_1}$is the electrostaric desolvation potential of the solutes, $\phi_{npdesolv_1}$is the non-polar desolvation potential of the solutes and $SASA$ is the solvent accessible surface area of the solutes. ϕ_{el_1} and $\phi_{el_2}$are the electrostatic potentials of two solutes which are calculated by solving the nonlinear Poisson - Boltzmann equation:

$$-\nabla\epsilon\,\nabla\phi = p + \sum_{i} q_i\,n_i \exp\left(-\frac{q_i\,\phi}{k_B T}\right) \quad (2.26)$$

Here, ϕ is the molecular electrostatic potential, ϵ is the position dependent dielectric permittivity, p is the molecular charge density, q_i is the net charge and n_i is the concentration of the ions of the solvent. The Adaptive Poisson-Boltzmann Solver (APBS) (131) and University of Houston Brownian Dynamics (UHBD) (132) are two commonly used software packages to compute electrostatic potentials of biomolecules. In this thesis, APBS software (131) is used for these calculations.

2.3.1.2 Effective charges

In SDA, the electrostatic intermolecular potential energies between two solutes are calculated in each step of the trajectory. However, due to high number of atoms and their charges, it is not computationally efficient to compute energies for all atoms in each step. In order to overcome this, effective charge methodology (ECM) (133) is used. In ECM, by using small number of charge assignments on Glu, Asp, Lys and Arg residues, C and N-termini of the proteins and P atoms of the nucleic acids, the electrostatic potential of a given molecule can be regenerated.

2.3.1.3 Electrostatic energy

In SDA, the following equation is used for the calculation of the electrostatic interaction energy between two solutes:

$$\Delta G = \frac{1}{2}\sum_{i} \phi_{el_2}(r) {q_{i_1}}^{eff}(r) + \frac{1}{2}\sum_{j} \phi_{el_1}(r)\, {q_{j_2}}^{eff}(r) \tag{2.27}$$

Here, ϕ_{el_1} and ϕ_{el_2} are electrostatic potentials which are fitted by the effective charges ${q_{i_1}}^{eff}(r)$ and ${q_{j_2}}^{eff}(r)$. In SDA, the electrostatic potential grids generated by APBS or UHBD of each molecule are used for calculating the electrostatic potential energy of two solutes.

2.3.1.4 Electrostatic desolvation energy

In SDA, the electrostatic desolvation energy is computed by using the following formula (134)

$$\Delta G_{edesolv}^{1-2} = \sum_{i_1} \phi_{edesolv_2}(r_{i_1})\,[{q_{i_1}}^{eff}(r)]^2 + \sum_{i_2} \phi_{edesolv_2}(r_{i_2})\,[{q_{i_2}}^{eff}(r)]^2 \tag{2.28}$$

in which $\phi_{edesolv}(r)$ is the electrostatic desolvation potential of one of the solutes and computed by,

$$\phi_{edesolv}(r) = \alpha \frac{\epsilon_s - \epsilon_p}{\epsilon_s\,(2\epsilon_s + \epsilon_p)} \sum_{j} {a_j}^3 \frac{(1+\kappa r_j)^2}{{r_j}^4}\, e^{-2\kappa r_j} \tag{2.29}$$

Here, ${q_{i_1}}^{eff}$ is the effective charge, α is an empirical scaling factor, ϵ_s is the solvent dielectric constant, ϵ_p solute interior dielectric constant, κ is the inverse of the Debye length. This sum is calculated for all solute atoms j with radius a_j.

2.3.1.5 Non-polar desolvation energy

In SDA, the non-polar desolvation energy mentioned in Eq. (2.25) is calculated by the formula (135):

$$\phi_{npdesolv} = \beta c \begin{cases} 1 & r_{min} \\ \dfrac{b - r_{min}}{b - a} & a < r_{min} < b \\ 0 & r_{min} > b \end{cases} \tag{2.30}$$

Here, β is the proportionality constant between buried area and non-polar desolvation energy, c is a factor to prevent double counting of the buried area, a and b are used to define the maximum distance of an interacting solute atom from the interacting surface.

2.3.2 SDA docking

The interaction energy and structure of the diffusional encounter complexes of biomolecules can be obtained by using SDA (130). In BD simulations with SDA, one of the solutes is centered and the second one is randomly placed on a surface of a sphere with radius b, rotational and translational diffusion is considered and in each step interaction forces are calculated. Simulations are conducted until the second solute reaches predefined surface of sphere radius $c \gg b$ (See Figure 2. 3) or the reaction criteria conditions are satisfied. Minimum energy structures are recorded and later used for clustering by using a hierarchical agglomerative average-linkage clustering algorithm to define the encounter complex structures.

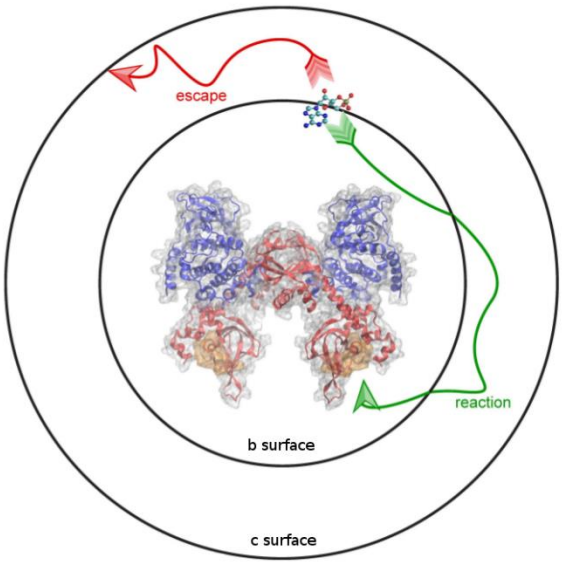


Figure 2.3: Schematic representation of the BD simulation. b and c surfaces are shown. (Figure is obtained from Boras B.W. et al. (102) and re-published under Open Access CC BY license.)

In this thesis, SDA docking is applied to determine the configurations of the various linker histone – nucleosome complexes.

2.3.3 SDA association rate calculation

It is important to understand the association of biomolecules as some of them occur very fast and others are relatively slow. In general, the bimolecular association rate constant of two molecules at the separation distance $r = b$ can be determined by using analytical Smoluchowski equation (136):

$$k_{on}(b) = \frac{4\pi D}{\int_b^\infty \frac{e^{U(r)/kT}}{r^2} dr} \tag{2.31}$$

Here $U(r)$ is a centrosymmetric interaction potential between the spheres and D is the diffusion constant. By generating thousands of trajectories, the probability of satisfying encounter complex criteria, β, can be determined and the previous equation can be written as,

$$k_{on}^{BD} = k_{on}(b) \frac{\beta}{1 - (1 - \beta)\frac{k_{on}(b)}{k_{on}(c)}} \tag{2.32}$$

For the efficient determination of the association rate constants, it is crucial to define a set of polar contacts of reaction criteria. This usually consists of minimum 2 independent contacts at 6 Å distance.

In this thesis, SDA association methodology is applied to investigate the association rate constants of wild type and mutant linker histone – nucleosome binding.

3

Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome

This chapter is adapted from the published research article “Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome” (137) authored by Öztürk M. A., Pachov G. V., Wade R. C. and Cojocaru V. MD and AMD simulations were run by Pachov G. V. and Cojocaru V. BD simulations and analysis of the MD simulation trajectories were conducted by Öztürk M.A.

3.1 Purpose of research

As introduced in Chapter 2, an important contributor to DNA compaction in chromatin structure is the LH. The 1:1 ratio of LH – nucleosome binding generates a chromatosome (138, 59). In addition to their structural significance, LHs have different isoforms and can regulate replication and transcription (139–141). Previously, one-start solenoidal helix (142) and two-start zig-zag helix (66) models of the 30 nm chromatin model were proposed, in which L-DNA can have different lengths (143) and conformations. Furthermore, previous computational and experimental studies proposed on- or off-dyad binding modes for LH - nucleosome binding (7, 9, 10, 14, 16, 45, 70, 71, 73, 144–147). Asymmetric (off-dyad) chromatosome models were obtained from NMR (45), site directed mutagenesis (17) and cryo-electron microscopy (cryo-EM) (10). However, X-ray (7, 9), NMR (148) and cryo-EM studies (9) proposed an on-dyad chromatosome structure. Interestingly, in these studies different LH and DNA sequences were used (See Table 1.1). Additionally, in some cases L-DNAs were not modeled or they were only used as static structures with different lengths. On the

other hand, the highly dynamic nature of the DNA in chromatin (149) can be the reason for the different binding modes reported in the literature.

The first X-ray structure of the globular domain of a LH (gH5) was obtained for *G. gallus* (PDB id: 1HST (5), 2.5 Å resolution). The structure is composed of a helix-turn-helix motif and a β-hairpin and consists of open (chain A – gH5A) and closed (chain B – gH5B) forms of the globular domain. However, the molecular significance of these states is not known. Previously, Pachov et al. (73) applied BD docking simulations of gH5B to a nucleosome structure, whose flexibility was obtained by NMA. Their simulations revealed an off-dyad LH - nucleosome binding (73). Additionally, Cui et al. reported that LH – nucleosome interactions are stabilized by sequence-specific hydrophobic interactions with AT-rich DNA (71). As BD simulations of Pachov et al. did not model either short-range hydrophobic interactions or conformational relaxation, the interactions reported by Cui et al. could not be confirmed. As a result, how the sequences and the conformational dynamics of the LH and the nucleosome affect the chromatosome configuration was not fully understood.

Here, classical molecular dynamics (CMD) and accelerated molecular dynamics (AMD) simulations are applied to understand the conformational plasticity of the nucleosome and of gH5 in free and off-dyad nucleosome-bound forms. Additionally, BD simulations are used to determine the dynamic pathways of the LH – nucleosome complex assembly.

3.2 Material and methods

Methods introduced in Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4 were conducted by Pachov G. V. and Cojocaru V.

3.2.1 Selection of starting structures

Open and closed *G. gallus* globular domain LH structures were obtained from PDB id: 1HST (5) chain A (gH5A) and chain B (gH5B), respectively. The crystal structure (PDB id 1KX5 (66), 1.9 Å resolution) was used for the nucleosome core particle (NUC). 10 base pairs of L-DNA were added to each end and histone tails were removed. By using the tleap module of the AMBER software (150) hydrogen atoms were added at pH 7. As the starting configuration for the molecular dynamics simulations of the LH – nucleosome complex, the structure of the NUC-gH5B encounter complex with the dominant binding mode of gH5B as described by Pachov et al. (73) was used. There were significant steric clashes between gH5A and the L-DNA for the initial superposition of the crystal structure of gH5A to the starting configuration of the NUC-gH5B complex. To resolve these, two snapshots from the initial 20 ns of the CMD of gH5A were superimposed by minimizing the root mean square deviation (RMSD) of the gH5 non-hydrogen atoms from the NUC-gH5B configuration. Then, these snapshots of gH5A were transferred into the NUC-gH5B structure after removal of

gH5B. Two starting models that have minimum steric clashes of the NUC-gH5A complex (NUC-gH5A and NUC-gH5A*) were obtained.

3.2.2 Setup of MD simulations

74 residue long unbound structures of gH5A and gH5B were neutralized with 11 Cl^- ions and solvated in explicit TIP3P water (127) of a truncated octahedral box. Water molecules with a minimal distance between any solute and solvent atoms of 12 Å at 300 K were used. There were 22134 and 24345 atoms respectively. Then, NUC-gH5B, NUC-gH5A, and NUC-gH5A* models were neutralized with 226 Na^+ ions and were solvated in a truncated octahedral box containing an additional ~50 mM NaCl. The total Na^+ ion concentration was ~200 mM and the minimal distance between the solute, including the neutralizing Na^+ ions, and solvent atoms was 4 Å. There were 198303 atoms for each LH – nucleosome complex system. Additionally, a system with the free nucleosome having 197127 atoms was setup by removing gH5B and 11 Cl^- ions from the solvated NUC-gH5B system.

The all-atom AMBER force field (109) modified for DNA (ff99) (110) and proteins (ff99SB) (111) with further corrections for the DNA backbone (parmbsc0) (112) were used. The Joung-Cheatham parameters optimized for TIP3P water (151) were used for the ions. Energy minimization was applied with the AMBER software (150) for all systems in 11 stages of 1000 steps each (100 steepest descent and 900 conjugate gradient) with decreasing restraints on non-hydrogen solute atoms (the force constants were 500, 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0 kcal/mol $Å^2$) with cut-offs exceeding the system size.

3.2.3 Classical MD simulations

NAMD software (117) was used to equilibrate the systems in three stages of 25, 100 and 250 ps. During the first two stages, the temperature increased from 100 to 300 K. All non-hydrogen solute atoms and ions were also harmonically restrained with force constants of 100 and 10 kcal/mol $Å^2$ in the NVT ensemble. 300 K no restraint conditions were applied and the density was equilibrated in the NPT ensemble during the third state. The temperature of 300 K was maintained using Langevin dynamics with a damping coefficient of 2 ps^{-1}. The pressure of 1.01325 bar was maintained using the Nose-Hoover-Langevin piston method with a period of 100 fs and decay of 50 fs. Water molecules were kept rigid by using the SHAKE algorithm (152). Throughout the equilibration, the integration time step was kept at 1 fs. Then, CMD simulations were performed (see Table 3.1) in which the temperature and pressure were maintained with the Berendsen weak coupling algorithm (with relaxation times of 1 and 5 ps, respectively, and compressibility of $4.57 \cdot 10^{-5}$ bar^{-1}) to avoid any potential influence of the Langevin equation on the dynamics of the systems. 2 fs integration time step was used after the equilibration and all bonds involving hydrogens were kept rigid by applying the SHAKE algorithm. The electrostatic interactions were computed by using the particle mesh

Ewald algorithm (124) with a grid spacing of 1 Å and a cut-off of 10 Å and all simulations were performed under periodic boundary conditions.

3.2.4 Accelerated molecular dynamics (AMD) simulations

As introduced in Chapter 2, by using AMD simulations (128), it is possible to accurately describe biomolecular dynamics on time scales significantly shorter than those required by CMD (25). The parameters used and the length of the simulations are shown in Table 3.1.

Table 3.1: Molecular dynamics simulations performed (Table was prepared by Cojocaru V. and re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

Simulation	Parameters	H5 gH5B	H5 gH5A	NUC	NUC-gH5B	NUC-gH5A	NUC-gH5A*
CMD-01	Time (ns)	600	600	100	324	324	324
	V_{DIHED} (kcal/mol)	749.1±10.9	750.7±10.9	16095.7±53.2	16828.5±52.7	16844.2±51.7	16847.0±53.2
CMD-02	Time (ns)	600	600	324	105	100	100
	V_{DIHED} (kcal/mol)	748.3±11.2	749.9±11.3	16075.7±50.5	16844.6±53.4	16847.7±55.4	16853.7±53.1
AMD-01	Time (ns)	200	208	-	108	100	100
	α	44.4	44.4	-	850.5	850.5	850.5
	E_{DIHED} (kcal/mol)	976.8	976.8	-	21097.0	21101.2	21101.2
	ΔV_{DIHED} (kcal/mol)	8.5±3.8	8.6±3.8	-	113.2±7.4	114.0±7.8	114.4±7.8
AMD-02	Time (ns)	200	212	-	114	112	112
	α	51.8	51.8	-	850.5	850.5	850.5
	E_{DIHED} (kcal/mol)	1005.7	1013.8	-	22798.0	22802.2	22802.2
	ΔV_{DIHED} (kcal/mol)	11.5±4.5	12.3±4.7	-	193.6±10.6	192.0±10.9	192.4±10.9
AMD-03	Time (ns)	200	208	-	108	112	112
	α	59.2	59.2	-	850.5	850.5	850.5
	E_{DIHED} (kcal/mol)	1050.8	1050.8	-	24499.0	24503.2	24503.2
	ΔV_{DIHED} (kcal/mol)	15.9±5.4	16.2±5.5	-	281.5±14.4	281.8±14.1	281.9±14.3
AMD-04	Time (ns)	200	200	-	-	-	-
	α	51.8	51.8	-	-	-	-
	E_{DIHED} (kcal/mol)	1109.3	1117.4	-	-	-	-
	ΔV_{DIHED} (kcal/mol)	18.4±6.2	19.5±6.4	-	-	-	-
AMD-05	Time (ns)	200	200	-	-	-	-
	α	51.8	51.8	-	-	-	-
	E_{DIHED} (kcal/mol)	1212.9	1221.0	-	-	-	-
	ΔV_{DIHED} (kcal/mol)	26.9±8.1	27.2±8.1	-	-	-	-

3.2.5 Analysis of structural dynamics

To describe the conformational plasticity of gH5, first two vectors v_H and v_B that thread through the two structural elements involved were defined, the helix α_3 and the sheet β_1, respectively (Figure 3.1): (i) v_H connects the geometric centers of the second and third turns of the gH5 helix α_3, defined by the backbone atoms (C, N, O, CA) of residues 67–71 and 71–75, respectively; (ii) v_B connects the geometric center of the backbone atoms of residues 82 and 94 with the geometric

center of the backbone atoms of residues 83 and 93. In addition, the center of the turn was defined between the two β strands as the geometric center of the backbone atoms of residues 87–91. The vectors v_{HT} and v_{BT} connect this point with the centers of vectors v_H and v_B, respectively. The latter two are connected by the vector v_{BH}. Finally, two angles Φ_1 and Φ_2 were defined as follows (Figure 3.1): Φ_1 is the angle between vectors v_H and v_{HT}, Φ_2 is the angle between the vectors v_{BH} and v_{BT}.

To characterize the structural dynamics within the LH – nucleosome complex, a reference xyz coordinate system was set-up based on two vectors v_N^1 and v_N^2 (Figure 3.5A). v_N^1 connects the geometric centers of nucleotides 45–48, 287–290, 123–126, 209–212 and nucleotides 83–86, 249–252 whereas v_N^2 connects the geometric centers of nucleotides 100–103, 232–235, 24–27, 311–314 and nucleotides 144–147, 188–191, 66–69, 266–269. v_N^1 was defined along the dyad axis and crosses v_N^2, approximately in the center of the nucleosome. The origin of the xyz coordinate system was defined at the point where v_N^1 crosses the nucleosomal DNA at the geometric center of nucleotides 83–86, 249–252. Then, the x-axis was defined to extend along v_N^1, the y-axis was defined along the cross product of x and v_N^2, and the z-axis was defined along the cross product of x and y. The orientation of gH5 with respect to the N-DNA was described by the angles θ_1 and θ_2 (Figure 3.5B), where θ_1 = the angle between the xy projection of v_H and the x axis and θ_2 = the angle between the yz projection of v_H and the z-axis. The motions of the L-DNAs were described using the angles γ_1 and γ_2 (Figure 3.11A and B), where γ_1 = the angle between the xz projection of the vector v_L^1 or v_L^2 and the z-axis, and γ_2 = the angle between the xy projection of v_L^1 or v_L^2 and the y axis. v_L^1 and v_L^2 were defined based on selected DNA residues along the helical axis of the two L-DNAs. v_L^1 connects the geometric centers of nucleotides 12–15, 320–323 and 2–5, 330–333, whereas v_L^2 connects the geometric centers of nucleotides 153–156, 179–182 and 163–166, 169–172 (Figure 3.11A). The numbering of the DNA nucleotides starts from 1 and 168 at the 5′ ends of L-DNA1 and L-DNA2, respectively, and runs to 167 and 334 at the 3′ ends of L-DNA2 and L-DNA1, respectively. All non-hydrogen atoms were used to define the nucleotides. All vector-based angle calculations were performed in VMD (153).

To analyze the slow motions of the L-DNAs, the essential dynamics of the nucleosome core particle was calculated from principal component analysis (PCA) of the CMD simulations with the CPPTRAJ program (154). For this, first gH5 was removed and all non-hydrogen atoms of the nucleosome core particle were superimposed. Secondly, the covariance matrix was calculated and diagonalized to extract the first 25 eigenvectors and their eigenvalues. Then, the trajectory was projected onto the 25 calculated modes, and the minimum and maximum projection values for each mode were extracted. Finally, these values were used to generate individual trajectories along each

mode and the motions of the L-DNAs in the trajectories of the first two modes that contributed most to the overall dynamics were analyzed.

3.2.6 Vector variance analysis

The variances of the defined LH vectors (V_{BH} V_{BT} V_{HT} V_H V_B) and nucleosome vectors (V_{N1} V_{N2} V_{L1} V_{L2}) due to intrinsic dynamics were determined with VMD as follows: (i) for each snapshot saved, the points defining each vector (Figures 3.1, 3.5 and 3.11) were superimposed onto the first frame of the trajectory; (ii) all angles between the vector at each time step were calculated; (iii) the values of these angles were plotted in Figures 3.2A and 3.6A in box plots with 25 % threshold and the maximum value as the upper limit. Additionally, the variance of the measured angles in each simulation is given on top of each box.

3.2.7 Analysis of residue-residue contacts

To analyze the hydrophobic interactions between V87, A89 and the helix α_3 in LH simulations the minimal distance between the methyl group carbons of V87 and A89 and all non-hydrogen atoms of helix α_3 was calculated (Figure 3.3). To analyze the hydrogen bonds between R74, Q83 and the backbone of turn β1, the minimal distance of the non-hydrogen atoms in the polar groups of the sidechains and the N and O atoms of the backbone of turn β_1 was calculated (Figure 3.3). The number of contacts between R73 and R74 of gH5 and DNA bases in the N-DNA were determined by imposing a 4.5 Å threshold contact distance during LH – nucleosome complex simulations (Figure 3.15).

3.2.8 BD simulations

Both gH5A and gH5B were docked using rigid body BD based docking simulations to eight nucleosome structures with different L-DNA1 conformations and L-DNA2 fixed in a specific, highly-populated conformation. These were selected from the CMD simulation without LH based on the γ_1 and γ_2 angles (Figure 3.13C and D). In addition, gH5B was docked in the nucleosome structure taken from the recent LH – nucleosome complex structure by Zhou et al. (PDB id: 4QLC, 3.5 Å resolution) (7) using the protocol of Pachov et al. (73). In short, NMA was applied using the NOMAD-Ref web-server (105) to generate nucleosome conformations with different degrees of L-DNA opening. The original structure (conformation 0), as well as two conformations with RMSD of 1 and 2 Å, respectively (all non-hydrogen atoms superimposed) along the first mode (‘conformation 1’ and ‘conformation 2’), were selected. The RMSD of the L-DNAs in these two structures from the original structure was 4.7 and 9.2 Å (the non-hydrogen atoms of the core histones superimposed), respectively.

First, polar hydrogen atoms were added to the structures by using PDB2PQR 1.8 (155) and partial atomic charges and atomic radii were assigned from the AMBER99 force field. The electrostatic potential was calculated for all structures by solving the non-linear Poisson–Boltzmann equation on a grid with a 1 Å spacing and dimension of 193^3 in APBS 1.4 (156) at temperature 298.15 K. The solvent and solute dielectric constants were 78.54 and 2, respectively and the ionic strength was 100 mM. Higher solute dielectric constants of 4, 6 and 8 were also tested for docking gH5 to the highly populated conformation of the nucleosome from snapshot 5 (Figures 3.13C and D). The results were insensitive to the varying solute dielectric constant in this range. To define dielectric boundary conditions, the van der Waals surface was used.

The BD simulations were performed with the SDA7 (Simulation of Diffusional Association) software (130) using electrostatic interaction forces. Short-range interactions were neglected, and a 0.5 Å excluded volume criterion to prevent overlap was applied. Effective charges were assigned to charged residues on the protein and to P atoms on the DNA using the ECM program (133). The trajectories were started randomly on a sphere at a center-to-center distance of b = 280 Å and stopped at a center-to-center distance of c = 500 Å. The time step was set to 1 ps for center-to-center distances up to 160 Å and increased linearly up to 100 ps at a distance of 260 Å. A total of 20 000 trajectories were generated for each pair of LH-nucleosome conformations simulated. The diffusional encounter complex was considered formed when the following two geometric conditions were satisfied: (i) the center-to-center distance of gH5 and the nucleosome <73 Å, and (ii) the nucleosome dyad point and gH5 separation <40 Å. The interaction energies and the coordinates of a complex were recorded if the RMSD to previously recorded complexes was >1 Å and the interaction energy was within the 5000 lowest (most favorable) energy complexes recorded. A complex with RMSD < 1 Å to a previously recorded complex but lower energy was recorded as a substitute of that complex. The 5000 recorded complexes were clustered into 10 groups according to the backbone RMSD values between them. Upon ranking the clusters by their population during the BD simulations, representative structures of the clusters were generated.

3.3 Results and discussion

3.3.1 Conformational plasticity of the LH

To describe the conformational plasticity of gH5 during the simulations (Table 3.1), the Φ_1 and Φ_2 angles were defined (Figure 3.1A and B, see Materials and Methods for details) using vectors with a small angular variance due to intrinsic internal motions (see Material and Methods and Figure 3.2A). In the crystal structure of gH5 the angles in the gH5B conformation are Φ_1 =

103.62° and Φ_2 = 53.06° whereas for the gH5A conformation, they are Φ_1 = 95.00° and Φ_2 = 91.73° (Figure 3.1).

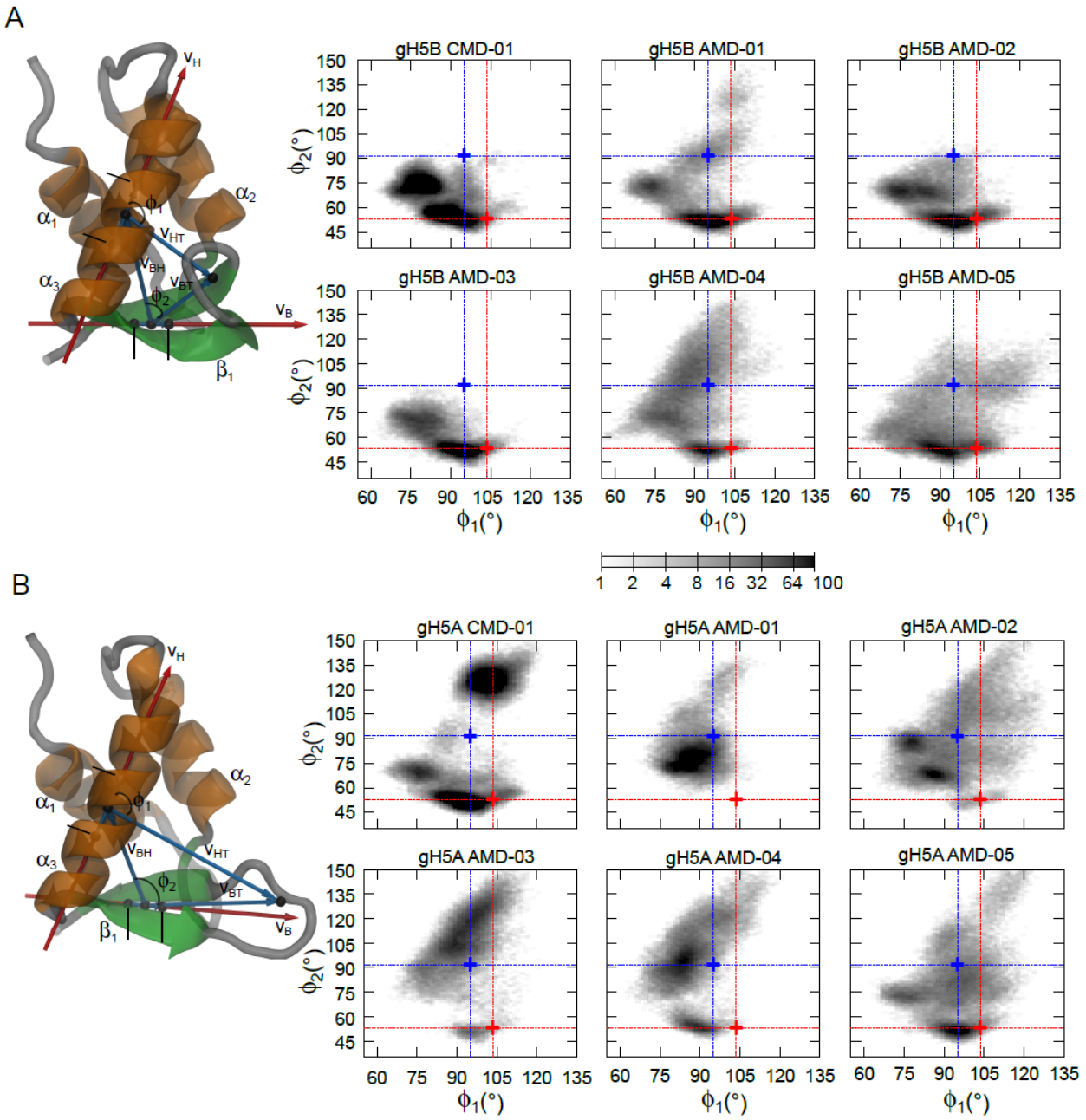
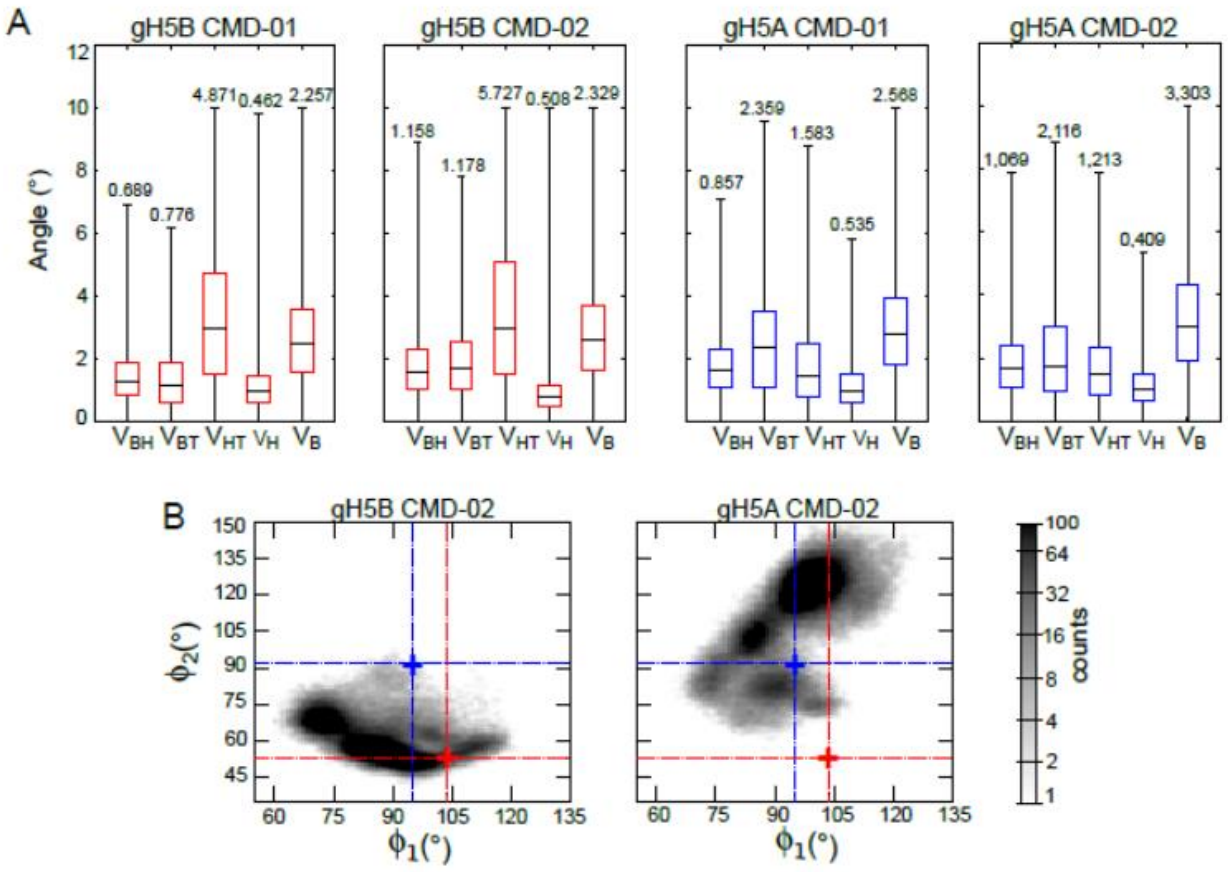


Figure 3.1: Conformational flexibility of the free gH5. **A-** Closed conformation gH5B. **B-** Open conformation gH5A. On the left, the two conformations (chains B and A in PDB id: 1HST) are shown in A and B respectively. Proteins are shown in cartoon representation and colored according to secondary structure: α helices in orange, β sheets in green, and unstructured regions in gray. The conformational space of gH5 is described by the angles Φ_1 and Φ_2 (see Methods). The following vectors are shown: v_H along the axis of helix α_3, v_B threading through the β sheet; v_{BH} connecting the centers of v_H and v_B; v_{BT} connecting the center of v_B with the β turn; v_{HT} connecting the center of v_H with the β turn. Φ_1 is the angle between v_H and v_{HT}; Φ_2 is the angle between v_{BH} and v_{BT}. For clarity, v_H and v_B are shown in red and longer than their actual definition marked with black thin lines. All other vectors are shown in blue and their endpoints as black spheres. On the right,

two-dimensional histograms of the sampling of the Φ_1/ Φ_2 conformational space for the corresponding gH5 conformation during CMD and AMD simulations (Table 3.1) are shown. The red and blue crosses mark the Φ_1 and Φ_2 values in the crystal structure for gH5A and gH5B respectively. See also Figures 3.2 and 3.3. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

In the CMD simulations, gH5B opened partially to a transient conformation characterized by an increase of Φ_2 to about 70°– 80° and a decrease of Φ_1 (Figures 3.1A, 3.2B and C). In AMD simulations, gH5B opened to either partially or fully open conformations (Figures 3.1A and 3.2C). At the lowest boost (AMD-01), gH5B remained in the closed conformation for most of the time but opened irreversibly after ~165 ns. Interestingly, at intermediate boosts (AMD-02, AMD-03), mainly reversible transitions between the closed and partially open conformations occurred, whereas at high boosts (AMD-04, AMD-05), reversible (on the 200 ns timescale of the AMD simulations) transitions to fully open conformations were observed. On the other hand, the open form, gH5A, adopted conformations characterized by Φ_2 values greater than 120° in CMD simulations (Figure 3.1B, 3.2B and C). These differed from that observed in the crystal structure with turn β_1 packing on the opposite side of the β sheet. Interestingly, in one CMD simulation gH5A closed irreversibly adopting a conformation similar to gH5B (Figure 3.1B). In AMD simulations, gH5A closed partially at the lowest boost (AMD-01) and adopted short lived fully closed conformations at intermediate to high boosts (AMD-02, AMD-03, AMD-04, AMD-05).



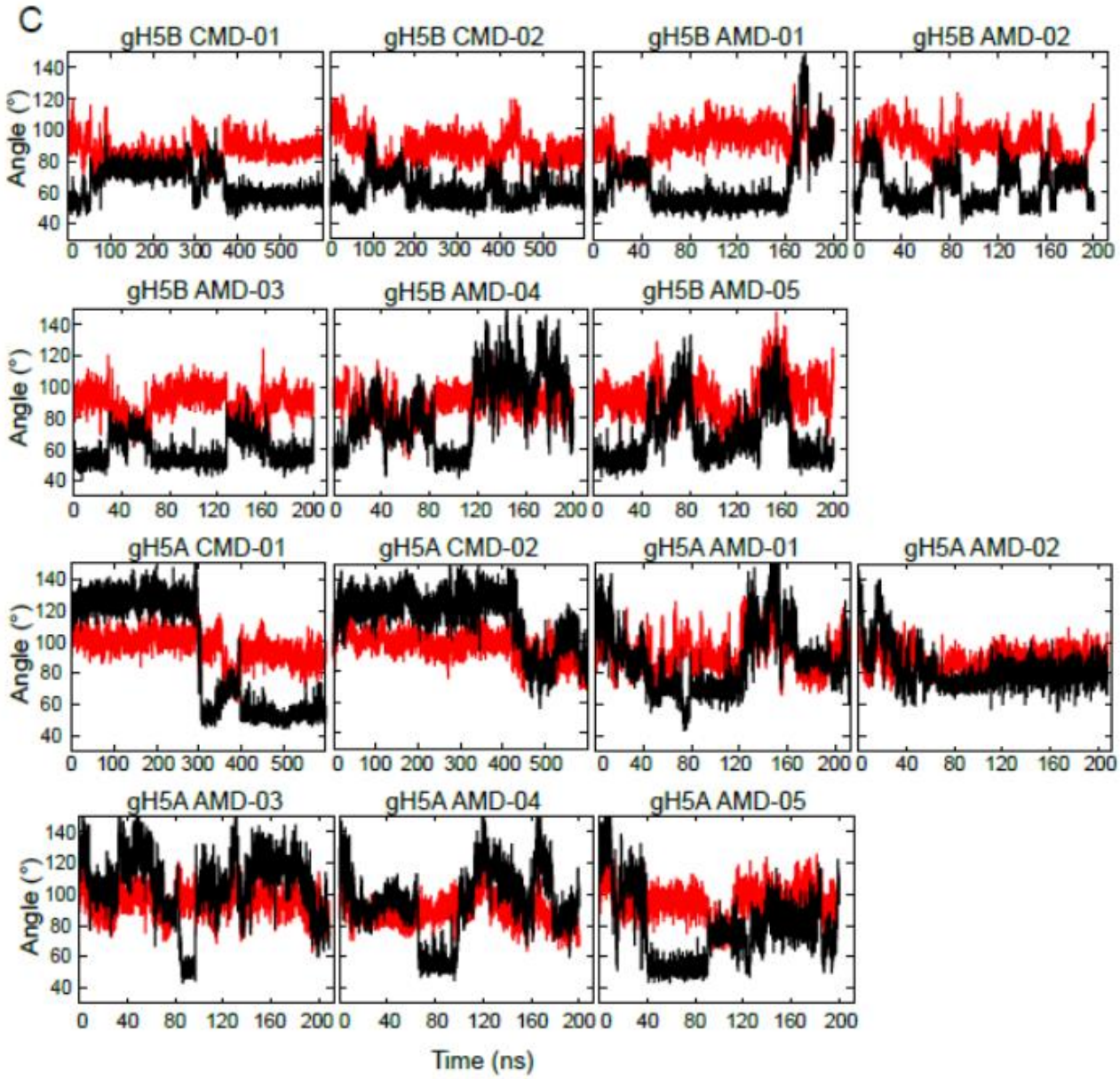


Figure 3.2: Conformational plasticity of gH5. **A-** Variance of the vectors used to calculate the Φ_1 and Φ_2 angle calculations. **B-** Two-dimensional histograms of the sampling of the Φ_1 / Φ_2 conformational space during the CMD-02 simulations of gH5B and gH5A. **C-** Time courses for the Φ_1 and Φ_2 angles during the simulations of gH5. See also Figure 3.1. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

In conclusion, on the timescale of the simulations, both reversible and irreversible transitions between the different conformations of gH5 for both starting structures were observed. A partially closed conformation characterized by a defined range of Φ_2 values (65°–80°) was identified. Furthermore, the open conformation showed an ensemble of conformations spanning a wide range of Φ_2 values. Based on these findings, it can be proposed that the unbound gH5 has a measurable preference for the closed conformation. The closed form is characterized by hydrophobic interactions between residues in turn β_1 (V87, A89) and residues in the helix α_3 as well as hydrogen bonds between polar sidechains (R47 in the helix α_2, Q83 in the sheet β_1) and the backbone of turn β_1 (see

Methods and Figure 3.3). Of these residues, only A89 is conserved in H5 and human H1 LH proteins suggesting that the LH sequence may influence the equilibrium between the two conformations. As residues in the turn β_1 have been proposed to be important for nucleosome binding (16, 45), it is possible that the changes in the equilibrium between the 2 conformations may result in different LH-nucleosome binding geometries.

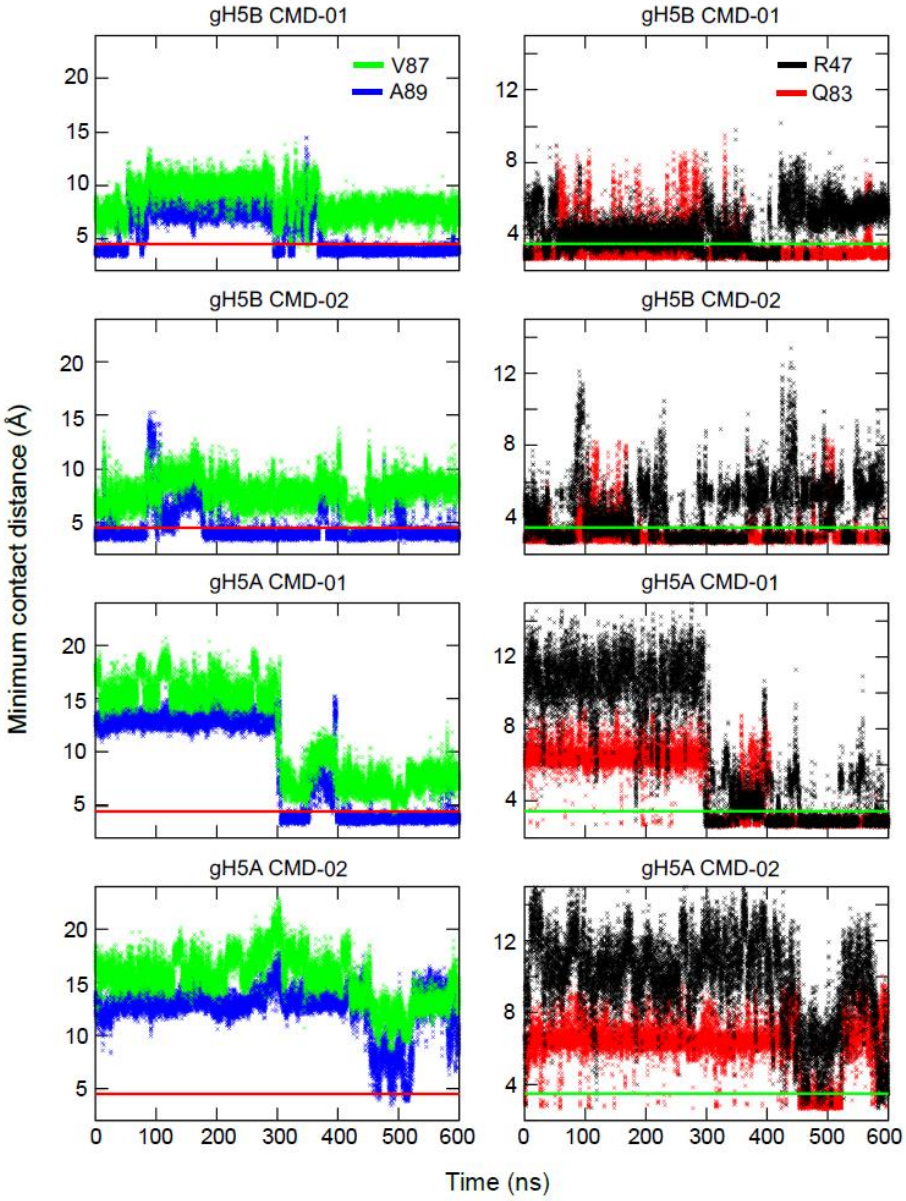


Figure 3.3: Interactions stabilizing the closed gH5B conformation. Time evolution of the minimum contact distances between the V87/A89 methyl groups and helix α_3 non-hydrogen atoms and between R47/Q83 side-chain N and O atoms and the backbone N and O atoms of the turn β_1 during CMD simulations are plotted (see Methods). The threshold distances for hydrophobic interactions (4.5 Å) and hydrogen bonds (3.5 Å) are shown as red and green lines respectively. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

3.3.2 Open LH conformation forms a more rigid LH – nucleosome complex structure

To characterize the binding mode of gH5 to the nucleosome, a schematic representation was adopted for which the nucleosome was aligned with the dyad axis perpendicular to the view plane and L-DNA1 and L-DNA2 on the left and right side, respectively (Figure 3.4). Then, the DNA grooves were numbered as follows: 0 = the minor groove at the dyad, −1, +1 = the neighboring major grooves of N-DNA toward L-DNA1 and L-DNA2, respectively, −2, +2 = the major grooves of L-DNA1 and L-DNA2 at the junction with N-DNA, −3, +3 = the following minor grooves of L-DNA1 and L-DNA2, respectively (Figure 3.4). In the off-dyad binding mode, the LH helix α_3 binds in the major groove −1, the turn β_1 interacts with groove −1 and −2, and the loop l_1 interacts with groove −2 (Figure 3.4A). In the on-dyad binding mode, helix α_3 binds in the minor groove −3, turn β_1 in groove 0, and the loop l_1 interacts with the groove 0 (Figure 3.4B).

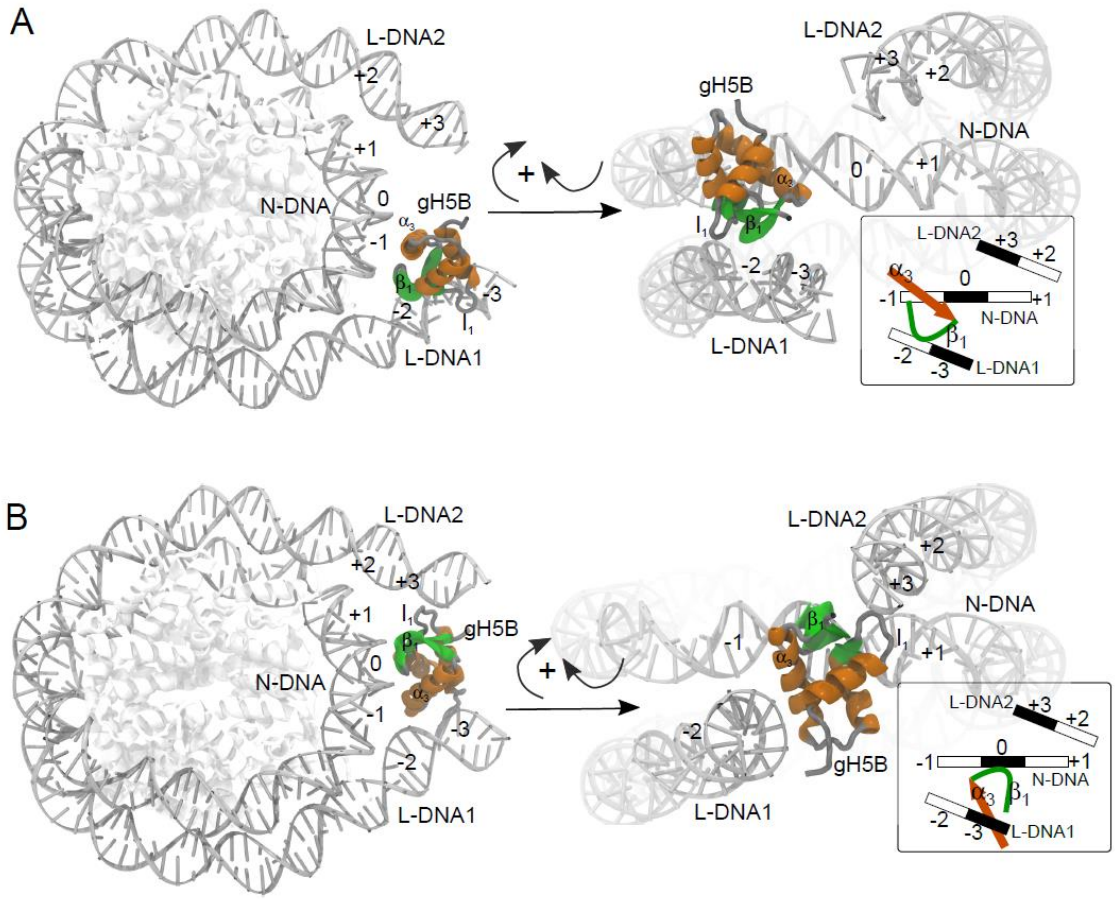


Figure 3.4: LH – nucleosome complex configurations. **A-** The off-dyad configuration proposed from BD docking by Pachov et al. (73) **B-** The on-dyad configuration revealed in the crystal structure of Zhou et al. (7). The images on the right show the nucleosome aligned with the dyad axis pointing towards the viewer and were obtained by 2 rotations, first vertical and second horizontal, marked by the 2 curved arrows. The insets show schematic representations of the gH5-nucleosome binding configurations. The numbers represent the DNA grooves as follows: 0 = the minor groove at the dyad, -1, +1 = the neighboring major grooves of N-DNA

towards L-DNA1 and L-DNA2, respectively, -2, +2 = the major grooves of L-DNA1 and L-DNA2, respectively, at the junction with N-DNA, -3, +3 = the following minor grooves of L-DNA1 and L-DNA2, respectively. Helix α_3 is shown as an arrow oriented from the N- to the C-terminus, whereas the turn β_1 is shown as a curved line representing the closed conformation. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137)

To study the orientation and dynamics of the off-dyad bound gH5 in the simulations of the LH – nucleosome complex, the number of contacts formed between different secondary structure elements of gH5 and the DNA was monitored. In addition, the θ_1 and θ_2 angles were defined (Figure 3.5A and B) to describe the rocking and tumbling motions, respectively, of helix α_3 in the major groove of the N-DNA (see Materials and Methods). For this, a reference coordinate system was defined using the vectors v_N^1 and v_N^2. v_N^1 was defined along the dyad axis and v_N^2 in a direction approximately orthogonal to the dyad axis. Neither of these vectors was sensitive to the intrinsic internal fluctuations (Figure 3.6A). From CMD simulations, it was found that the pattern of contacts between gH5 and the nucleosome depends on the gH5 conformation. The closed gH5B formed more contacts between its loop l_1 and L-DNA1 and fewer contacts between its turn β_1 and L-DNA1 compared to the open gH5A (Figure 3.5C, D and Figure 3.6B). Remarkably, the ranges of sampled θ_1 and θ_2 angles were greater in the CMD simulations of the NUC-gH5B complex compared to the NUC-gH5A form (Figure 3.5C, D and Figure 3.6B) indicating that the open form, gH5A, contributes to a more rigid complex. This suggests that the open gH5A is the preferred conformation of gH5 in the off-dyad configuration of the LH – nucleosome complex. The gH5A-nucleosome off-dyad binding geometry is in agreement with previous experiments that revealed residues involved in H1.0-nucleosome binding (16). H1.0 is the mammalian LH isoform most similar to H5. Interestingly, an off-dyad configuration has also been obtained for the *Drosophila* H1 globular domain (45) but with a different orientation of the LH in which helix α_3 does not dock in the major groove of N-DNA, suggesting that the detailed geometry of the off-dyad configuration may be LH-isoform dependent.

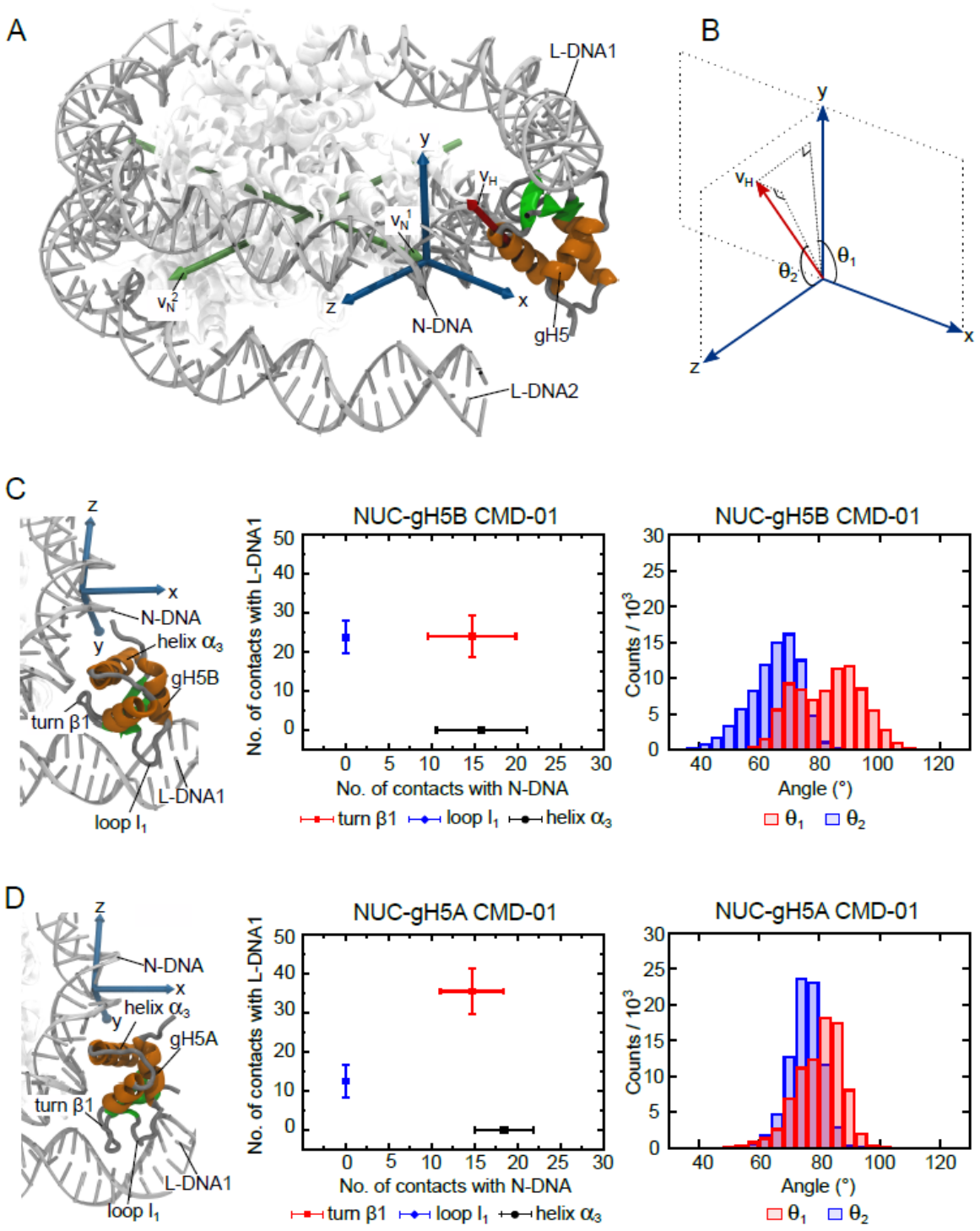


Figure 3.5: Orientation and dynamics of gH5 in the LH – nucleosome complex. **A-** Structure of the off-dyad LH – nucleosome complex. A reference coordinate system, xyz, was constructed using the vectors v_N^1 and v_N^2. These were defined between selected DNA bases to cross as closely as possible to the center of the nucleosome (see Methods). v_N^1 points along the dyad axis. v_N^2 connects two points on opposite sides of the nucleosome DNA, above and below the dyad point, respectively. To construct the coordinate system, v_N^1 was translated on the x axis, the y axis was defined along the cross product of x and v_N^2, and the z axis was defined

along the cross product of x and y. The two linker DNAs (L-DNA1 and L-DNA2), nucleosomal DNA (N-DNA), gH5, and the vector v_H (see Figure 3.1) are labeled. **B-** Schematic representation of the definition of the two angles, θ_1 and θ_2, describing the orientation of helix α_3 of gH5 with respect to N-DNA. θ_1 is the angle between the xy projection of v_H and the x axis. θ_2 is the angle between the yz projection of v_H and the z axis. **C-D** Orientation and dynamics of gH5B (C) and gH5A (D). The number of contacts of three structural regions of gH5 (turn β_1, loop l_1 and helix α_3) with different DNA regions (N-DNA and L-DNA1) and the histograms of θ_1 and θ_2 distributions are plotted. See also Figure 3.6. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

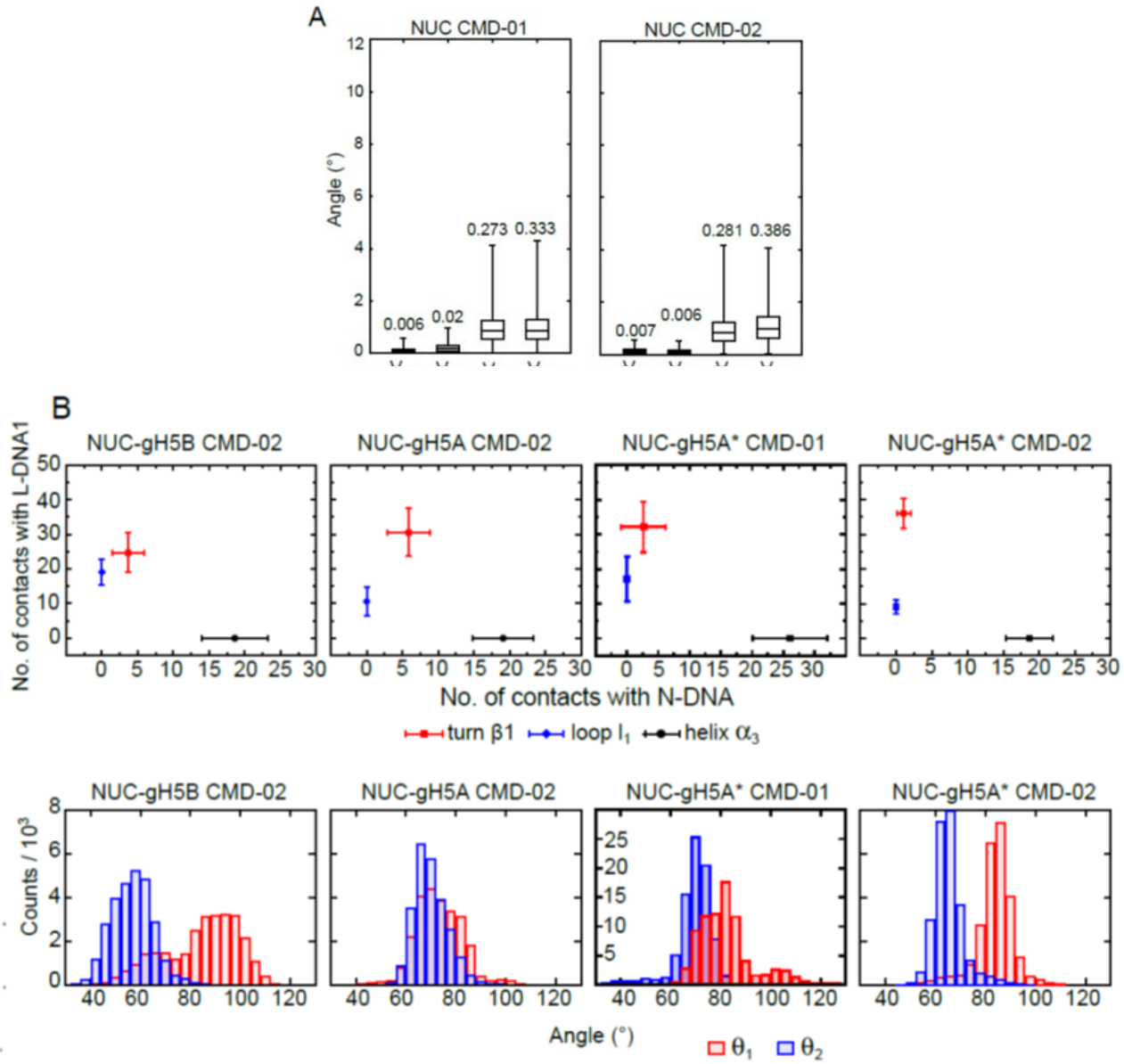


Figure 3.6: LH – nucleosome complex configuration and flexibility. **A-** Variance of the vectors used to define the reference coordinate system for the nucleosome and the L-DNA arms (see Methods) **B-** Orientation and dynamics of gH5 in CMD simulations of the LH – nucleosome complex. First row: Number of contacts between three structural regions of gH5A (turn β_1, loop l_1, and helix α_3) and different DNA regions (N-DNA, L-DNA1). Second row: θ_1 / θ_2 histograms. See also Figure 3.5. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

3.3.3 Closed LH conformation opens in accelerated LH – nucleosome complex simulations

To characterize the conformational dynamics of the LH while bound to the nucleosome, the Φ_1 and Φ_2 angles were monitored (Figure 3.1) during MD simulations of the LH – nucleosome complex (Figure 3.7A, B and 3.8). In the CMD simulations of the NUC-gH5B, gH5B remained closed with a slight increase of both Φ_1 and Φ_2 (Figure 3.7A). In AMD simulations of NUC-gH5B, three open states were observed for gH5 in which both Φ_1 and Φ_2 angles increased by up to 40°. When the boost was low (AMD-01), reversible transitions between closed and open conformations between 45 and 60 ns were occurred (Figure 3.8). In the simulations with higher boosts (AMD-02, AMD-03), Φ_2 increased irreversibly up to 120° (Figure 3.7A and 3.8). Interestingly, opening of gH5B was correlated with an increase in Φ_1 (Figure 3.7A) in contrast to the simulations of the free gH5 (Figure 3.1). In the NUC-gH5A simulations, gH5A adopted a predominant open conformation with Φ_2 larger than 105° and Φ_1 smaller than 110° (Figure 3.7B and 3.8). The values of Φ_2 were similar to those observed in CMD simulations of free gH5A (Figure 3.1) and reflect the packing of the turn β_1 away from helix α_3. Therefore, the extended structure of the β-turn observed in the crystal structure is not stable during the simulations. Importantly, gH5A did not close in any of the simulations of the complex. These findings suggest that the closed conformation of gH5 is not stable in the fully bound complex with the nucleosome in the off-dyad configuration and provide further support for an induced fit mechanism, in which gH5B forms the encounter complex and opens in the fully bound complex.

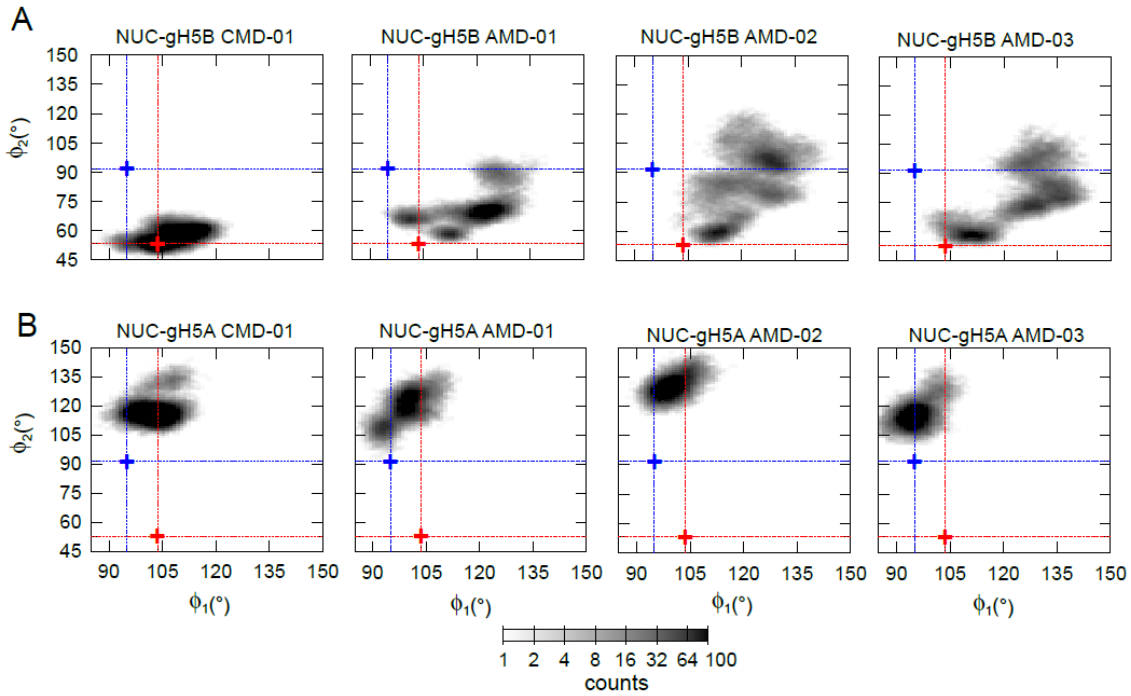


Figure 3.7: Conformational dynamics of gH5 in the LH – nucleosome complex. **A-** NUC-gH5B; **B-** NUC-gH5A. The two-dimensional histograms for the sampling of the Φ_1/Φ_2 conformational space during the

four CMD and AMD simulations (Table 3.1) are shown. The graphs are colored as in Figure 3.1. See also Figure 3.8. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

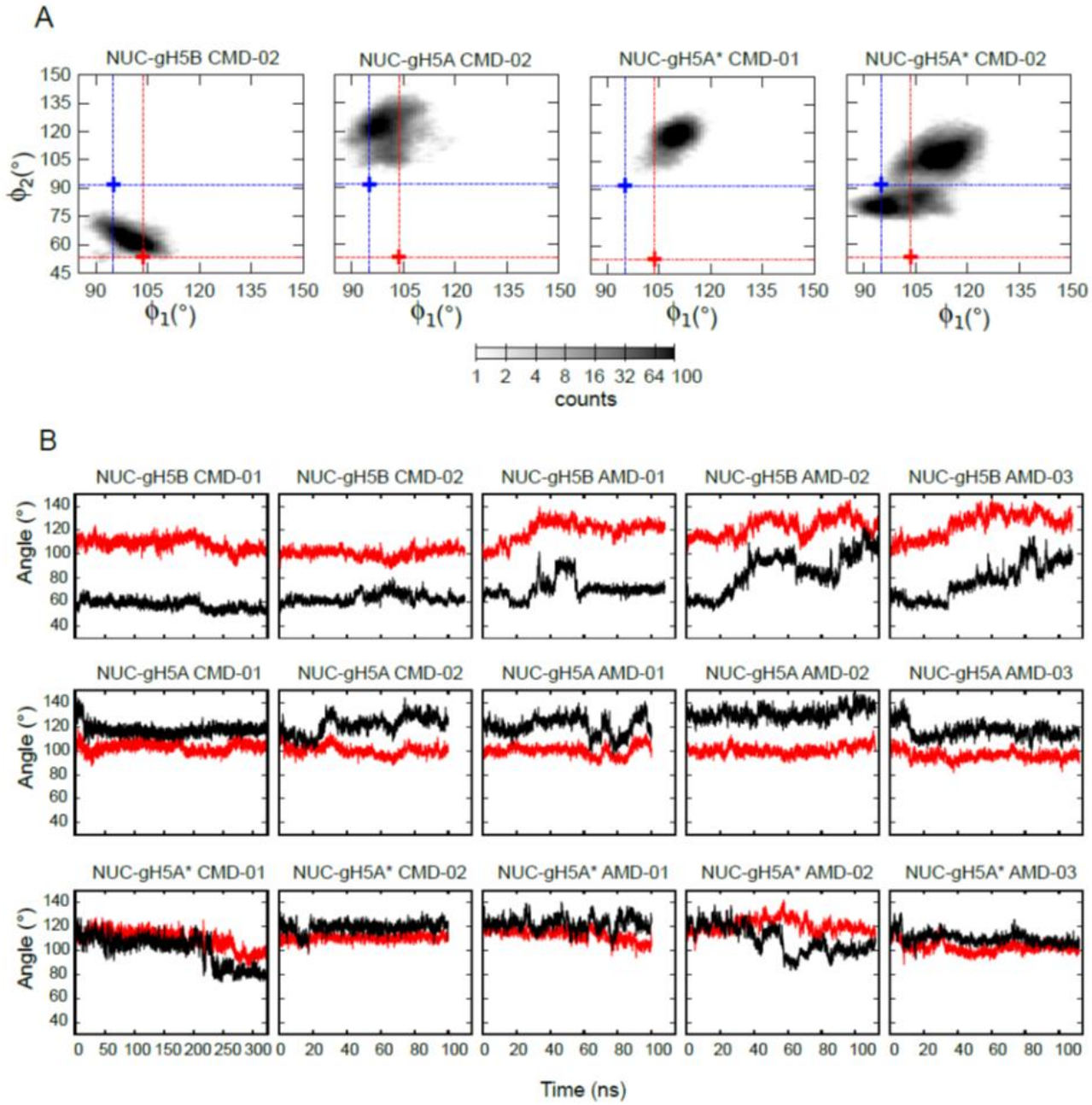


Figure 3.8: Linker histone conformations in the LH – nucleosome complex. **A-** The two-dimensional histograms for the sampling of the Φ_1 / Φ_2 conformational space (see Figure 3.1) during CMD simulations of the LH – nucleosome complex. Time courses for the Φ_1 and Φ_2 angles. See also Figure 3.7. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

3.3.4 Open linker histone conformation interacts with thymidines in the linker DNA

To explore how the open gH5A conformation stabilizes the LH – nucleosome complex, the hydrophobic contacts between turn β_1 of gH5 and thymidines in L-DNA1 were analyzed. It was observed that residues V87 and A89 from gH5A form alternative networks of hydrophobic

interactions with 1 to 3 thymidine bases in L-DNA1 (Figures 3.9A and 3.10). Although these interactions require a higher thymidine content in the L-DNA, the precise position of the bases may vary. Thus, the hydrophobic interactions are only partially DNA sequence specific. In CMD simulations of the NUC-gH5B complex, no hydrophobic contacts between gH5B and L-DNA1 were formed, whereas in the AMD simulations between 7 and 16% of the frames showed at least one such contact (Figures 3.9B and 3.10A). The formation of 1 or 2 hydrophobic contacts between 40 and 55 ns with low boost (AMD-01), and after ~50 ns with higher boost (AMD-02, AMD-03) was correlated with the opening of the gH5B (Figure 3.10B). In contrast, in over 60% of the trajectories of the NUC-gH5A complex (~100% in the simulations with the highest boost), at least one such hydrophobic contact was formed (Figures 3.9B and 3.10C and D). Interestingly, the sampling efficiency of the hydrophobic contacts was greater in the AMD simulations in particular, for the NUC-gH5A* system (Figure 3.10D). These findings indicate that the increased stability of the LH – nucleosome complex with gH5 in the open conformation is due to additional hydrophobic contacts formed between gH5A (V87, A89) and thymidines in L-DNA1. These findings could explain the proposed higher preference of the LH for T-rich regions of DNA (71).

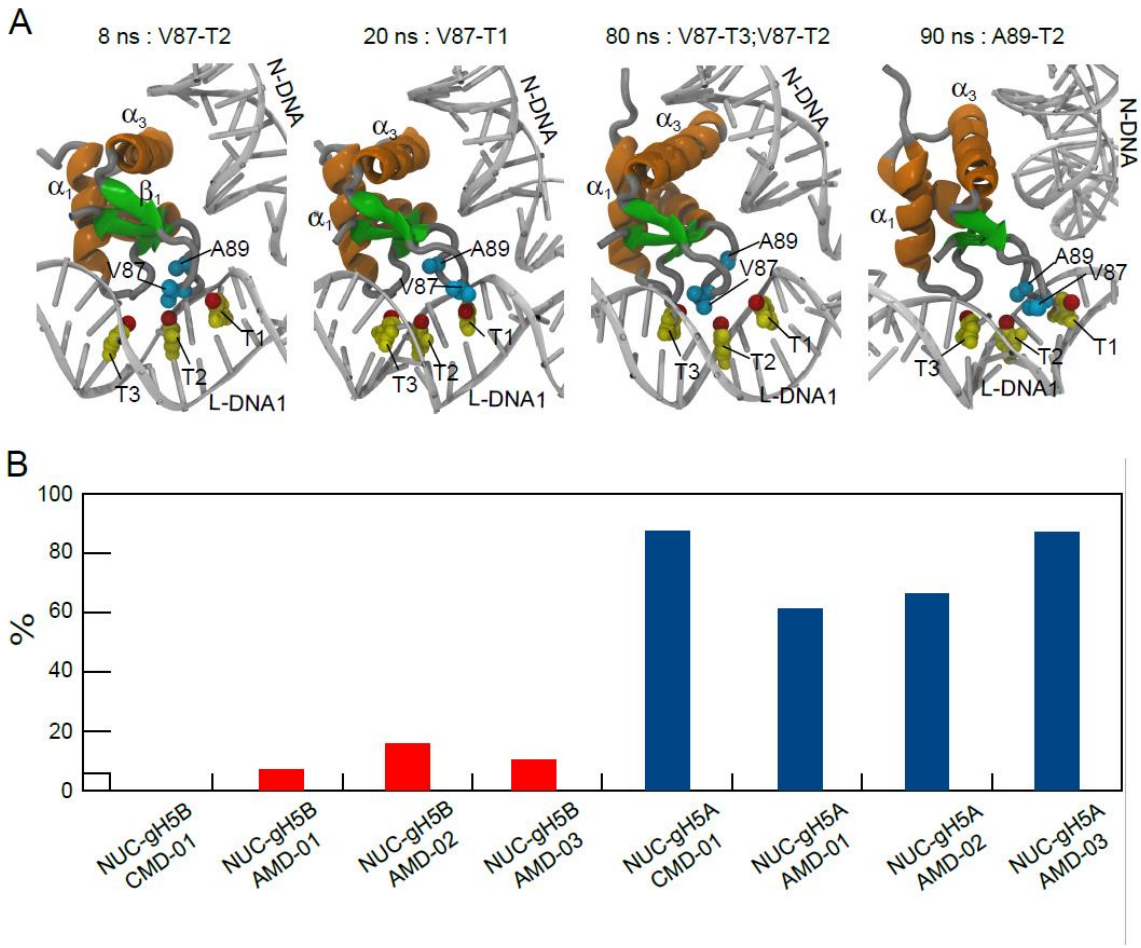


Figure 3.9: Hydrophobic contacts between gH5 and L-DNA1 in the LH – nucleosome complex. **A-** Representative structures from the AMD-01 simulation of the NUC-gH5A complex showing different hydrophobic contacts between V87 and A89 of gH5 and thymidine bases in L-DNA1. Protein residues are

shown in cyan, whereas the thymidine bases are shown in yellow with the methyl group in red. **B-** Percentages of MD trajectory frames in which at least one hydrophobic contact is established for NUC-gH5B (red) and NUC-gH5A (blue) complexes. See also Figure 3.10. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

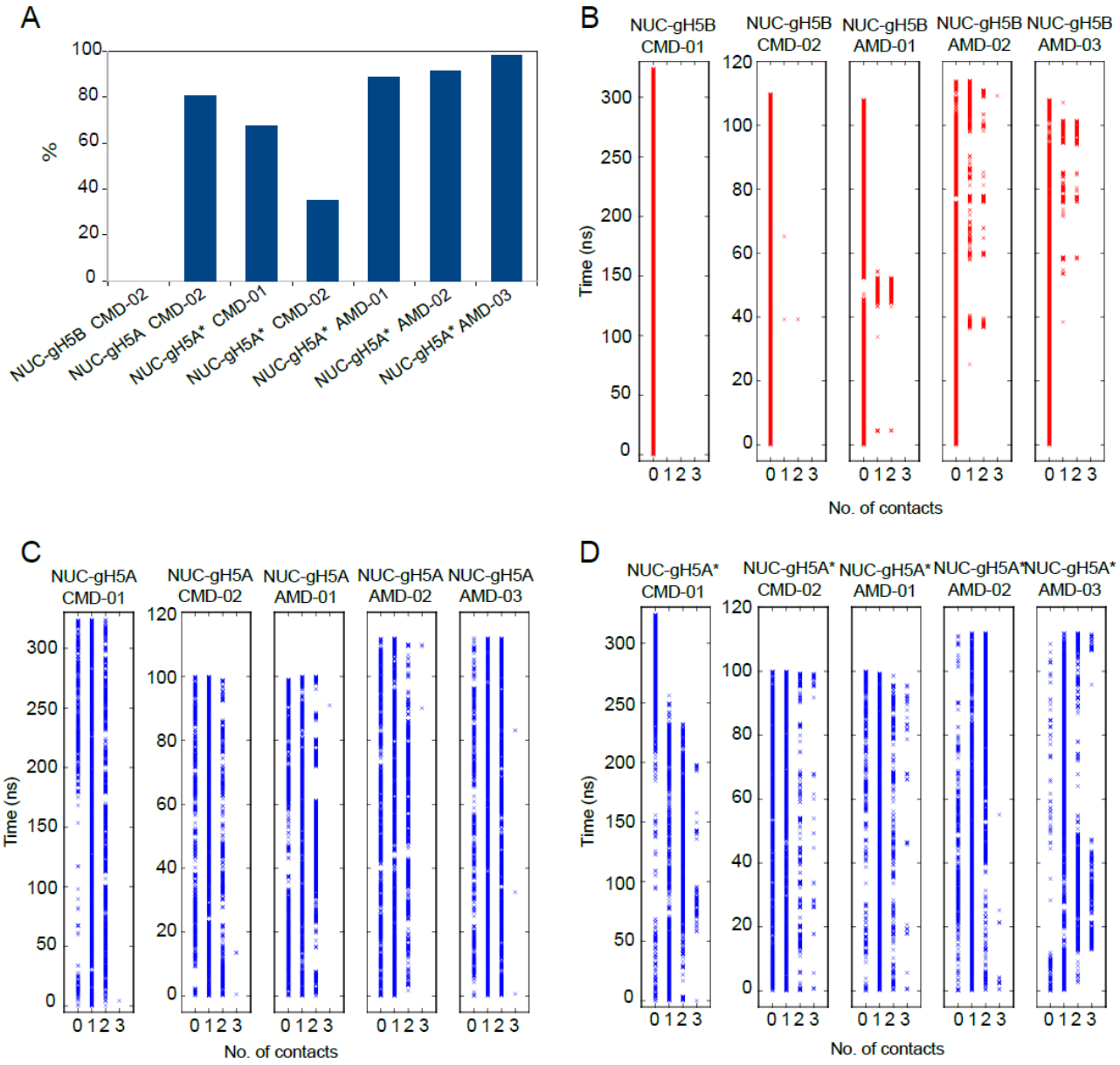


Figure 3.10: Hydrophobic interactions between gH5 and L-DNA1. **A-** The percentage of trajectory in which at least one hydrophobic contact is established between V87 or A89 of gH5 and thymidine bases in L-DNA1. **B-D** Time courses for the number of hydrophobic contacts between gH5 and L-DNA1 during simulations of NUC-gH5B (B), NUC–gH5A (C), and NUC-gH5A* (D). See also Figure 3.9. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

3.3.5 Binding of LH to nucleosome remodels L-DNA dynamics

To study how the binding of the LH influences the dynamics of the L-DNAs, the γ_1 and γ_2 angles were defined which describe the motions of the L-DNAs in the xz and xy planes of the reference coordinate system, respectively (Figures 3.11A, B and see Materials and Methods for details). Then, the essential dynamics from PCA of the CMD simulations were calculated and the γ_1 and γ_2 angles were monitored in the trajectory projections along the first two modes. In the CMD simulation of the free nucleosome, L-DNA1 moved predominantly along a path that is a combination of the two types of motions described by the two angles (Figures 3.11C and 3.12A). On the other hand, L-DNA2 moved predominantly in the xy plane along mode 1 showing little to no variation of γ_1, whereas its motion along mode 2 differed in the 2 independent CMD simulations (Figure 3.11C and 3.12A).

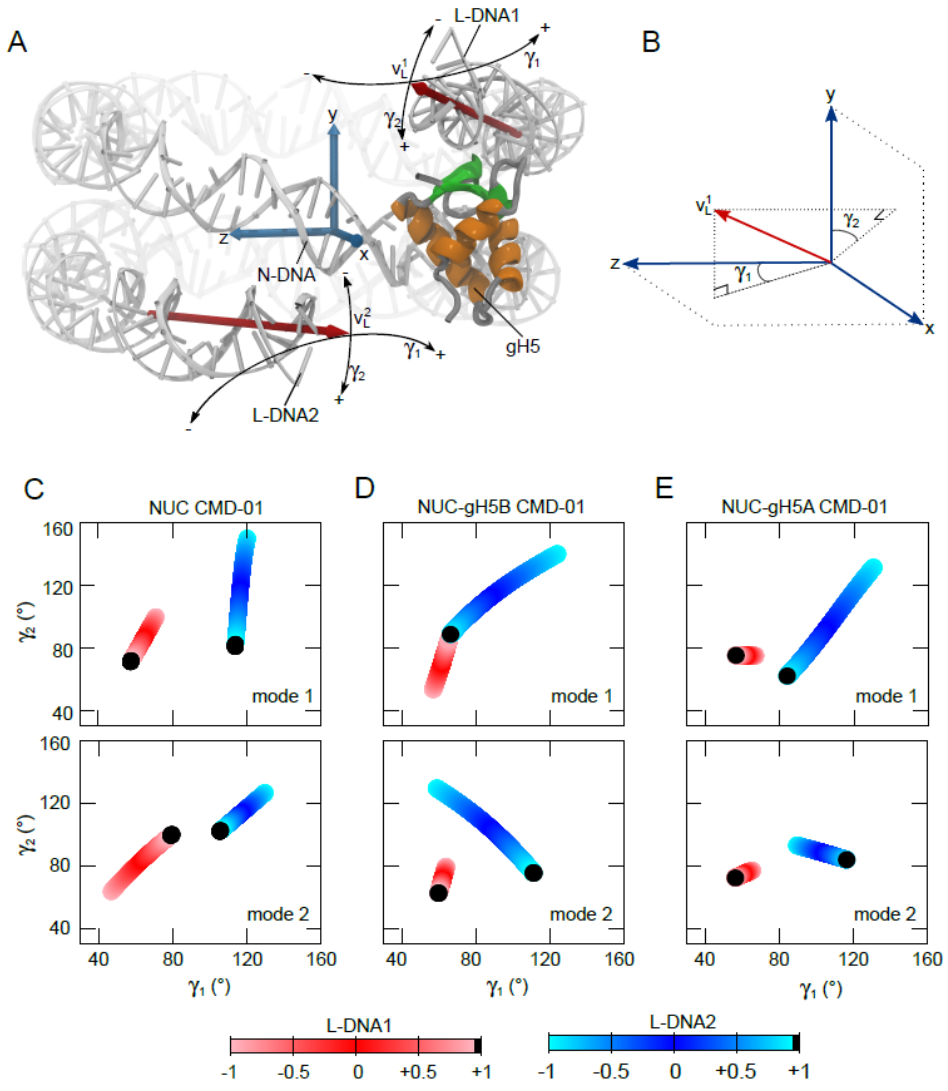


Figure 3.11: Effect of gH5 binding to the nucleosome on L-DNA motions. **A-** Structure of the LH – nucleosome complex showing vectors and angles defining the motions of the L-DNAs. The reference coordinate system xyz is shown in Figure 3.5. The vectors v_L^1 and v_L^2 were defined based on selected DNA

bases to represent the helical axes of L-DNA1 and L-DNA2, respectively. The double headed arrows show the directionality of the L-DNA motions described by the two angles, γ_1 and γ_2 (see Methods for details). **B-** Schematic representation of the definition of γ_1 and γ_2. γ_1 is the angle between the xz projection of v_L^1 or v_L^2 and the z axis, whereas γ_2 is the angle between the xy projection of v_L^1 or v_L^2 and the y axis. **C-E** Motions of the L-DNAs (L-DNA1 in red and L-DNA2 in blue) along the first two essential dynamics modes of the CMD-trajectory. The data from the simulations of the nucleosome (NUC) and the LH – nucleosome complex (NUC-gH5B and NUC-gH5A) are shown in the first, second, and third columns, respectively. See also Figure 3.12. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

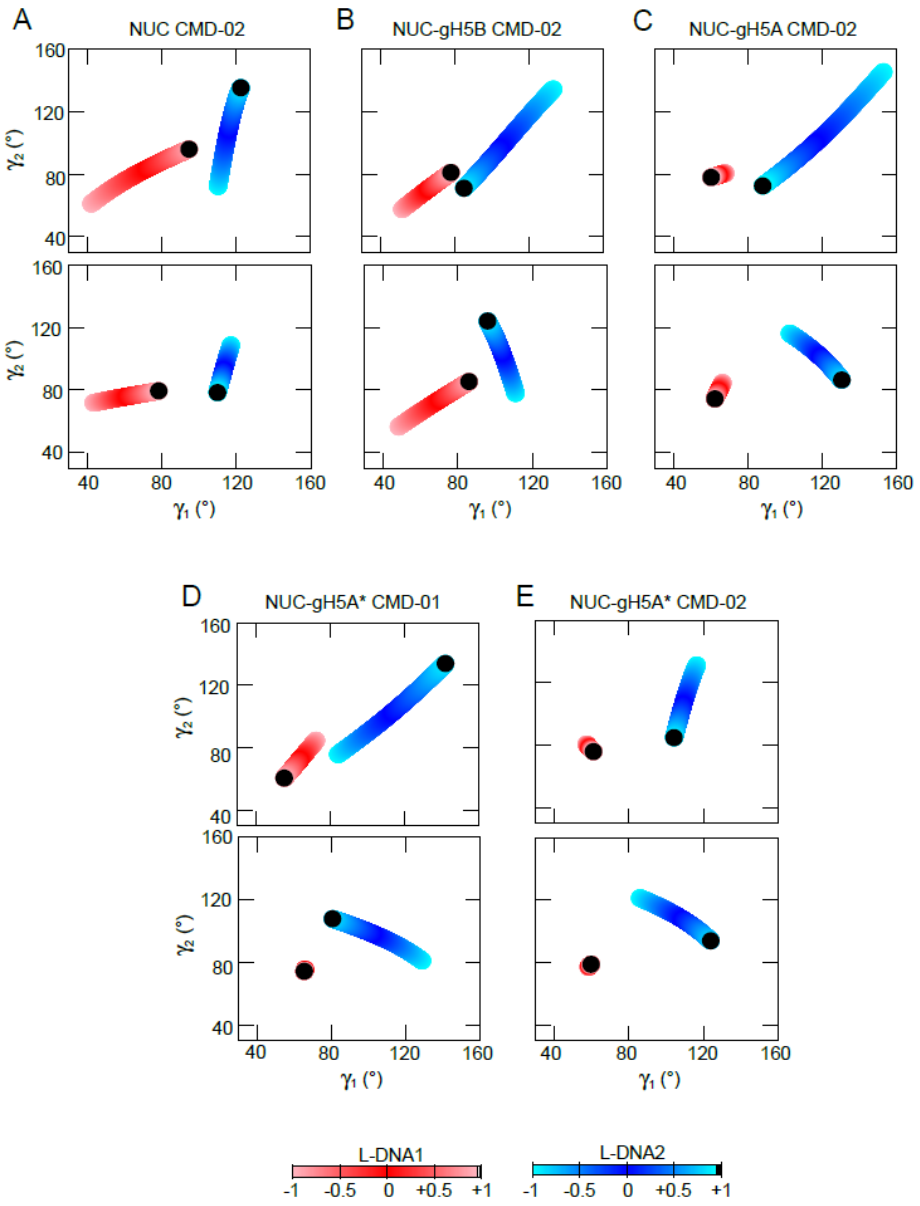


Figure 3.12: Linker DNA dynamics. **A-E** Motions of L-DNAs along the two first essential dynamics modes in CMD simulations of NUC (A), NUC-gH5B (B), NUC–gH5A (C), and NUC-gH5A* (D, E). See also Figure 3.11. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

In CMD simulations of the LH – nucleosome complex, a clear separation between the xy and xz motions along different modes was not observed (Figures 3.11D, E and 3.12). Both L-DNAs sampled predominantly a combined path along all modes, suggesting that the presence of the LH alters the relative timescales of these motions. Interestingly, the closed gH5B conformation only marginally reduced the amplitude of the L-DNA motions (Figures 3.11 and 3.12). In contrast, the open gH5A conformation greatly suppressed the L-DNA1 motion (Figures 3.11E and 3.12C, D and E). These findings are in agreement with the observation that gH5A showed less flexibility in its orientation when bound to the nucleosome compared to gH5B (Figure 3.5), providing further support for the induced fit mechanism described in previous paragraphs. The selective suppression of L-DNA1 in the fully bound gH5-nucleosome complex in the off-dyad configuration, has important implications for the assembly of higher-order chromatin structures (157) and it is analogous to the proposed change in DNA dynamics upon core histone protein binding in the nucleosome (158). Interestingly, L-DNAs used here are asymmetric in sequence, L-DNA2 having a higher GC content, and asymmetric dynamics of the L-DNAs were observed (Figures 3.11A and 3.12A). These findings are in agreement with a recent study showing L-DNA sequence dependent, asymmetric flexibility and unwrapping of the nucleosome (148).

3.3.6 Nucleosome dynamics determine binding mode of LH conformations

To explore the effect of the conformational dynamics on the LH – nucleosome complex assembly, BD simulations were performed with different gH5 and nucleosome conformations. For this, first the electrostatic potential of gH5 was calculated and it was found that it differs between the two conformations. The large positive stripe on gH5A is perturbed on gH5B leading to a more evenly distributed potential (Figure 3.13A and B), suggesting that the gH5-nucleosome encounter complex may differ between gH5A and gH5B. Then, eight representative snapshots from the CMD simulation of the free nucleosome were selected based on the distribution of the γ_1 and γ_2 angles. An increase of γ_1 and a decrease of γ_2 reflect the opening of L-DNA1 along the two types of motion described by the two angles (Figure 3.11A and B). Because in the off-dyad configuration, gH5 binds only to L-DNA1, the conformation of L-DNA1 was varied (Figure 3.13C), keeping the conformation of L-DNA2 fixed (Figure 3.13D). The selection of snapshots was not affected by the limited sampling of the L-DNA dynamics in the 100 ns CMD simulation (Figure 3.14A and B). With this selection, the representative conformational space sampled by L-DNA1 in the absence of the LH was covered. To

evaluate the binding modes resulting from the BD simulations, the scheme described in Figure 3.4 was used.

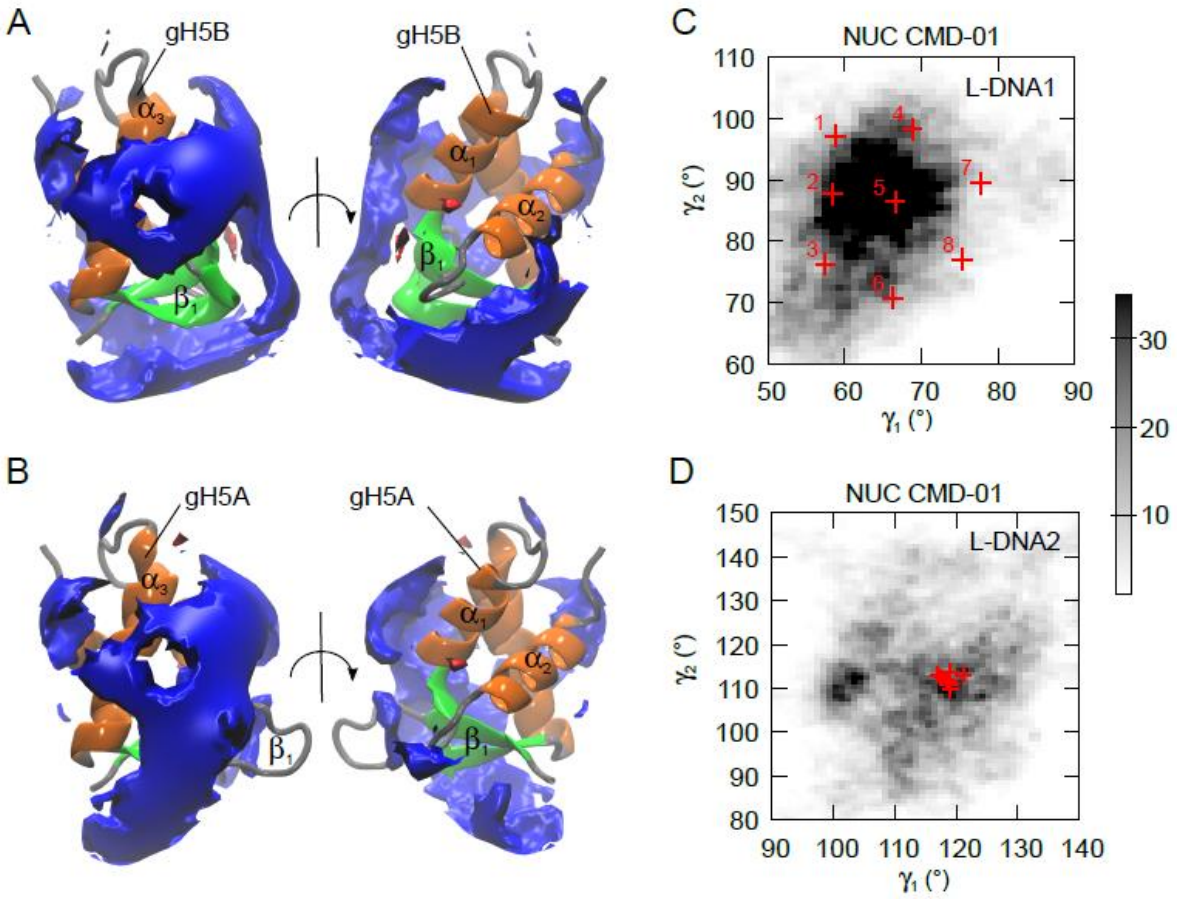


Figure 3.13: Preparation of BD simulations. **A-B** Molecular electrostatic potentials of gH5B (A) and gH5A (B). **C-D** Snapshots from the CMD-01 simulation of the free nucleosome selected for BD simulations (labeled in red) on the γ_1/γ_2 histograms for L-DNA1 (C) and L-DNA2 (D). See also Figure 3.14. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

The closed gH5B conformation formed an off-dyad encounter complex with nucleosome conformations from snapshots 2, 3, 5 and 8 (Table 3.2). As snapshot 5 lies in the center of the γ_1/γ_2 histogram, this finding indicates that the off-dyad configuration is the predominant binding mode for gH5B. Moreover, it shows that closing of L-DNA1 along one direction (lower values of γ_1) and the opening along the other (lower values of γ_2) (snapshots 2 and 3), as well as opening of L-DNA1 along both directions simultaneously (snapshot 8), still permits binding of gH5B in this configuration. These findings confirm previous observations based on NMA and BD simulations (73). The open gH5A conformation formed similar off-dyad encounter complexes in snapshots 3, 6 and 7 (Table 3.2). This indicates that opening of the L-DNA1 in either direction (higher γ_1 or lower γ_2), but not in both simultaneously, is required for the binding of gH5A in this configuration. The BD

simulations also reveal other conformation-dependent configurations of the gH5-nucleosome encounter complex but not the on-dyad configuration (Table 3.2, Figure 3.14C and D). Taken together, these findings indicate that besides the LH conformation, the nucleosome conformational dynamics determine the LH binding configuration. Therefore, it can be proposed that the off-dyad encounter complex forms through a conformational selection mechanism in which different conformations of the LH bind to a subset of specific conformations of the nucleosome. Interestingly, the structure of a chromatin fiber (10) revealed different degrees of L-DNA opening in different regions with asymmetric binding of LH H1. This further suggests that the interplay between LH binding and L-DNA dynamics is important for the higher-order chromatin structures.

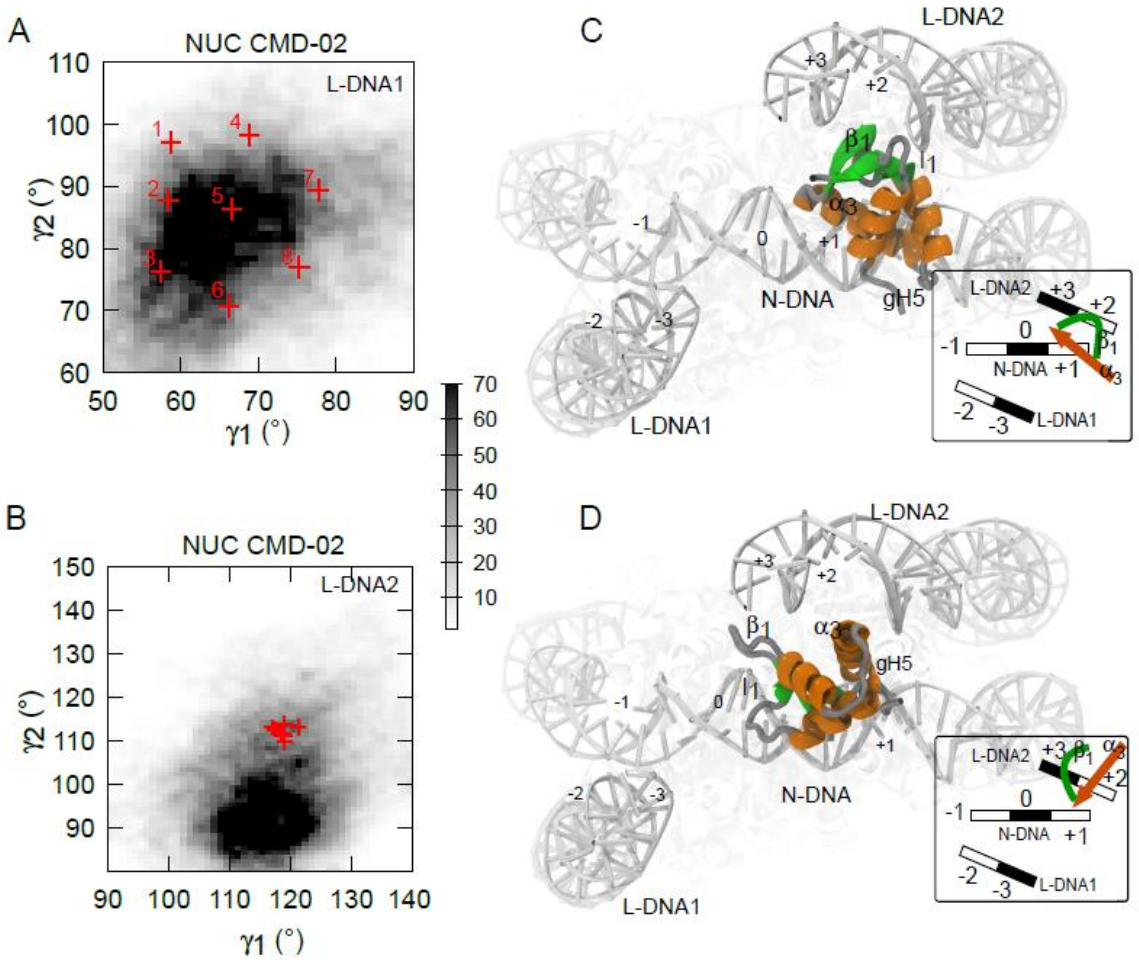


Figure 3.14: BD simulations. **A-B** Positioning of the selected snapshots from the CMD-01 simulation of the free nucleosome (labeled in red) on the γ_1/γ_2 histograms calculated from the CMD-02 simulation for L-DNA1 (A) and L-DNA2 (B). **C-D** Alternative, conformation dependent gH5-nucleosome binding modes in encounter complexes obtained by BD docking simulations of gH5B (C) and gH5A (D). The representation is as in Figure 3.4. See also Figure 3.13. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

Table 3.2: Binding configurations of gH5-nucleosome encounter complexes[#] (Table is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

CMD snapshot	L-DNA1		L-DNA2			gH5B					gH5A				
	γ_1(°)	γ_2(°)	γ_1(°)	γ_2(°)	N[&]	%[&]	α_3[§]	β_1[§]	I_1[§]	N[&]	%[&]	α_3[§]	β_1[§]	I_1[§]
1	58.7	97.0	119.0	109.7	6.5	34	+1 ↙	+1,+2	+2	6.3	49	+2 ↙	+1,+3	none
						31	+1 ↙	+1,+2	+2		20	+2 ↙	+1,+3	none
2	58.4	87.6	118.3	111.8	2.3	35	-2 ↙	-2	-1	3.4	70	+1 ↙	+1,+2	+3
						22	-1 ↘	-1,-2	-2		14	0 ↓	-1,-3	-3
3	57.4	76.2	118.7	111.5	4.3	42	-1 ↘	-1,-2	-2	2.6	42	0,-1 ↘	-1,-3	-2
						30	+1 ↙	+1,+2	+2		25	0 ↓	+1,+3	+2
4	68.8	98.2	119.0	110.9	2.9	63	+1 ↙	+1,+2	+2	2.7	54	0 ↓	-1,-3	-2,-3
						14	+3 ↙	0,+3	+1		16	-1,-2 ↘	-1,-3	none
5	66.5	86.4	121.2	112.9	4.3	30	-1 ↘	-1,-2	-2	4.2	45	+1 ↙	+1	+3
						24	-1 ↘	-1,-2	-2		31	+1 ↙	+1,+3	+2
6	67.2	77.2	118.6	110.0	3.4	46	+1,+2 ↙	+1	0	4.4	45	0,-1 ↘	-1,-3	-2
						27	+1 ↙	+1,+2	+2		22	+1 ↙	+1,+2	+3
7	77.7	89.4	117.0	112.9	2.7	26	+1 ↙	+1,+2	+2	5.6	47	0,-1 ↘	-1,-3	-2
						24	+1 ↙	+1,+2	+2		45	0,-1 ↘	-1,-3	-2
8	75.2	76.9	118.9	113.7	5.9	52	-1 ↘	-1,-2	-2	3.5	49	-1,-2 ↘	-1,-2	none
						19	-1 ↘	-1,-2	-2		24	+1,+2 ↙	+1,+2	+1

[#]The data was obtained from BD docking simulations performed to 8 different nucleosome conformations selected from the CMD simulation of the free nucleosome (Figure 3.13 C, D) (see Methods).

[§]In these columns: the numbers represent the DNA grooves (Figure 3.4) on the nucleosome in contact with each structural element of gH5; the arrows show the orientation of the helix α_3 of gH5 relative to the helical axis of N-DNA when the nucleosome is aligned with the dyad axis perpendicular to the view plane (Figure 3.4); the off-dyad binding mode is shown in red.

[&]N is the total number of complexes in each BD simulation divided by 10^3; the percentage of the given BD cluster members in the total number of complexes is given in the % column for clusters 1 and 2, respectively.

3.3.7 LH – nucleosome complex assembly simulations reveal experimentally-determined configurations

Up to this point, my consideration was focused on the off-dyad chromatosome configuration which originally was proposed based on BD simulations (73). Interestingly, the recent crystal structure of the LH – nucleosome complex revealed the closed gH5B conformation bound in an on-dyad configuration contacting both L-DNA arms (Figure 3.4B) (7). Therefore, there is an apparent contradiction between the experimental and simulation-based configurations. However, it should be noted that the sequence of the nucleosome in the crystal structure differs from that used in BD simulations (73). To test whether the DNA sequence may influence the LH – nucleosome complex configuration, the identical original protocol was applied based on NMA and BD simulations to dock gH5B to the nucleosome taken from the new structure. With the very closed nucleosome conformation from the crystal structure, the reference bound complex was not reproduced (Table 3.3). However, this is not surprising because it is unlikely that a diffusional encounter complex is formed with a tightly closed conformation of the nucleosome. Remarkably, when gH5B was docked to a nucleosome structure opened slightly along the lowest frequency mode obtained from NMA (see Methods), the on-dyad was the obtained configuration in the two topmost ranked clusters (Table 3.3). Besides showing that BD simulations accurately describe LH – nucleosome complex configurations, these findings suggest that the LH binding mode to the nucleosome may depend on DNA sequence as well as histone sequence. Therefore, it can be proposed that both the off-dyad and on-dyad configurations are possible upon binding of gH5 to different nucleosomes with different nucleic acid sequences. A higher GC content around the dyad combined with a higher AT content in the L-DNA may favor the off-dyad configuration in which arginines from the third helix of gH5 form direct contacts with bases in the major groove of nucleosomal DNA (Figure 3.15). It can be proposed that a single LH isoform is able to bind to nucleosomes in different configurations depending on the sequence and conformation which is supported by evidence for both on- and off-dyad binding modes for both gH1 (10, 144, 159, 160) and gH5 (7, 73).

Table 3.3: Binding configurations of encounter complexes[#] (Table is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

Structures/ Models	L-DNA1		L-DNA2		Linker histone	Reference				Docking results			
	γ_1(°)	γ_2(°)	γ_1(°)	γ_2(°)		α_3	β_1	l_1	N $\cdot 10^6$	%	α_3	β_1	l_1
Pachov et al. (73)	55.8	64.4	126.9	87.1	gH5B	-1 ↘	-1,-2	-2					
					gH5A	-1 ↘	-1,-2	-2					
Zhou et al. (7) conformation 0	85.8	117.1	76.0	62.5	gH5B	-3 ↖	0	+3	2.6	49	-3,0 →	-3,+3	none
										24	none →	0,+3	+3
Zhou et al. (7) conformation 1	86.8	107.4	85.5	61.9	gH5B				7.6	28	-3 ↖	0	+3
										23	-3 ↖	0	+3
Zhou et al. (7) conformation 2	87.8	98.5	94.0	61.4	gH5B				2.1	50	0,-3 ↗	0,+3	+3
										12	-3 ↖	0,+3	+3

[#]The data was obtained from BD docking simulations of gH5B to the nucleosome taken from the crystal structure of Zhou et al. (7) ; “conformation 0” is the crystallographic conformation; “conformation 1” and “conformation 2” are open conformations along the lowest frequency mode calculated using NMA (see Methods for details); all other notations and colors are explained in the footnotes of Table 3.2.

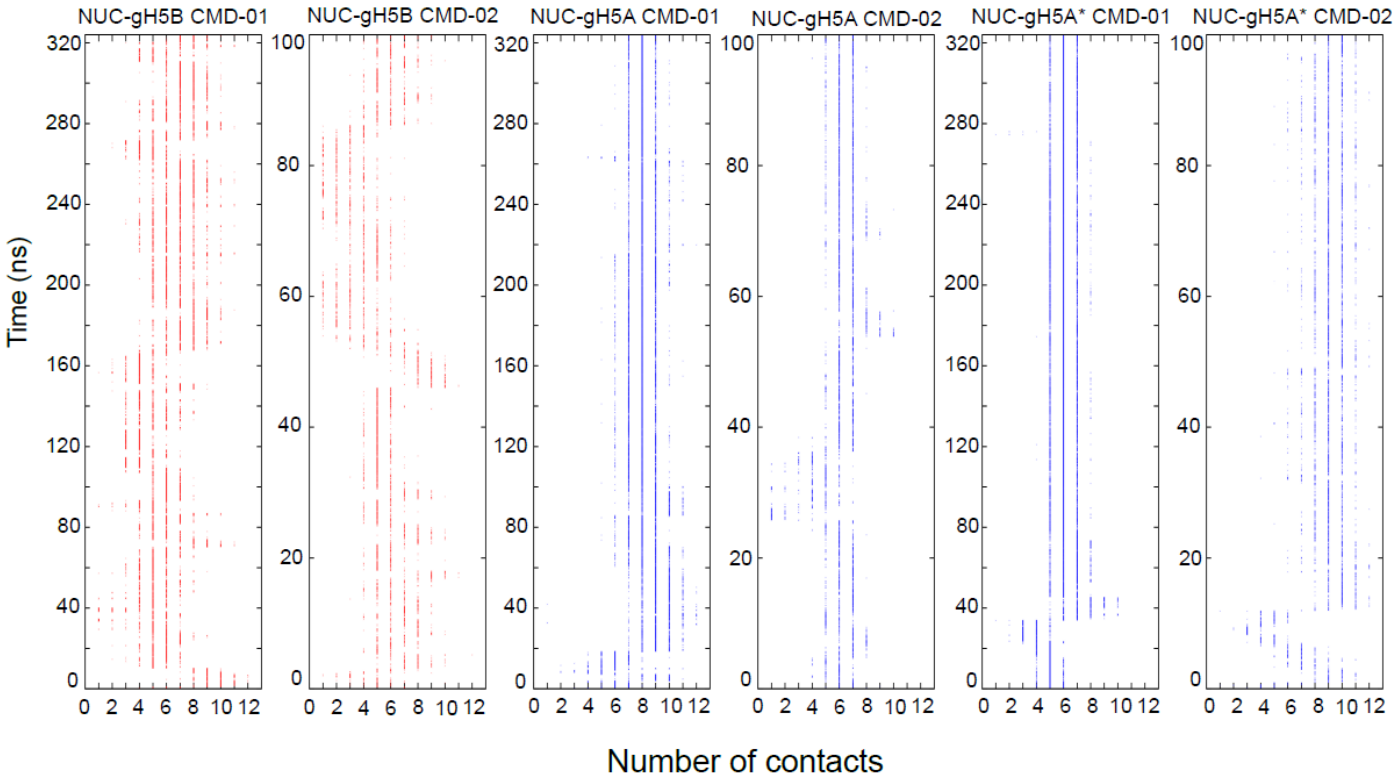


Figure 3.15: Interactions of arginines from LH with the N-DNA major groove. Time evolution of the contacts between R73 and R74 of gH5 and the N-DNA bases in chromatosome simulations. For details see Methods. (Figure is re-published under Open Access CC BY license from Öztürk M.A. et al. (137))

3.4 Concluding discussion

In this work, a series of classical and accelerated MD and BD simulations was performed to explore the dynamic nature of LH - nucleosome binding and chromatosome formation. In the MD simulations, it was found that gH5 has the ability to switch from open to closed conformations and vice versa in solution. Interestingly, the free gH5 has a measurable preference for the closed form which is stabilized by a series of hydrophobic and hydrophilic interactions that involve residues from the turn β_1. However, the open conformation stabilized the off-dyad encounter complex and significantly reduced the L-DNA motion through hydrophobic interactions with thymidines in the nearby L-DNA. This could explain the higher preference of the LH for T-rich regions (71), and provides further support for experimental observations (7). Moreover, the closed conformation opened in accelerated MD simulations of the LH – nucleosome complex. Based on these findings, an induced fit mechanism for the formation of the off-dyad chromatosome configuration can be proposed. On the other hand, it was shown that the conformational plasticity of the nucleosome provides a framework for conformational selection during chromatosome assemby. Therefore, an interplay between induced fit and conformational selection mechanisms contributes to alternative chromatosome configurations which further affect the higher order chromatin structure. Finally, it

was show that when docking the closed gH5 conformation to the alternative DNA sequence used to solve the most recent crystal structure of the gH5-nucleosome complex, the experimentally determined on-dyad binding mode of gH5 was observed. This opens up the possibility that, besides nucleosome and LH conformational plasticity, the DNA sequence may play a role in the chromatosome assembly without necessarily affecting the DNA binding affinity.

One potential limitation of the study may arise from not considering the highly flexible N- and C-terminal tails of the LH and the core histone proteins. It is challenging to sufficiently sample the conformational space of such highly flexible regions in molecular dynamics simulations. Especially the effect of the C-terminal tail of LH proteins may be of particular interest for future studies because, although it does not appear to affect the primary binding geometry around the dyad (144), it does affect the secondary positioning of LH proteins around the linker DNA (144, 159, 161) and the diversity of higher-order chromatin arrangements (144) through mechanisms that may involve DNA-mediated folding (162). The core histone H2A tails have been shown to affect the binding affinity of LH to the nucleosome (45). However, a recent long simulation of a free nucleosome (39) shows no significant overlap between the core histone H3 tails and the LH binding region, consistent with NMR data showing that the H3 tails are unaffected by binding of an H1 construct (163). This further suggests that the core histone tails may have little effect on the binding geometry of the LH whereas they may affect binding affinity through an induced fit mechanism in which the tails wrap around the LH after the initial binding. In conclusion, the chromatosome assembly pathways and final configurations may be significantly more complex than previously thought and further experimental and computational studies are necessary to elucidate them in the context of higher order chromatin structures.

4

Dependence of chromatosome structure on linker histone sequence and post-translational modifications

This chapter is adapted from the submitted research article “Dependence of chromatosome structure on linker histone sequence and post-translational modification” authored by Öztürk M.A., Cojocaru V. and Wade R. C.

4.1 Purpose of research

In Chapter 3, the structural plasticity of the linker histone, nucleosome and LH – nucleosome complex was investigated. By using nucleosome structures representing the conformational space of the L-DNA arms, it was shown that alternative chromatosome configurations were possible upon LH – nucleosome binding. Also, it was found that, both on- and off-dyad chromatosome configurations were possible depending on LH sequence. In this chapter, it is aimed to determine the LH residues which specifically impact the chromatosome configuration. Additionally, effects of post-translational modifications (PTM) on LH globular domain – nucleosome binding are investigated.

LH proteins play a key role in higher order structuring of chromatin for the packing of DNA in eukaryotic cells and in the regulation of genomic function (30, 164). The common fruit fly (*Drosophila melanogaster*) has a single somatic isoform of the linker histone (H1) (165). It is thus a useful model organism for investigating the effects of the LH on nucleosome compaction and the structure of the chromatosome. In 1993, the first crystal structure of the LH GD was reported (PDB id: 1HST, 2.6 Å resolution (5)). Despite the recent determination of the crystal structures of LH GD-nucleosome complexes (PDB id: 4QLC, 3.5 Å resolution (7); PDB id: 5NL0, 5.4 Å resolution (9)), the structural

determinants of chromatosome formation are still not well understood. In two studies by Zhou et al. (7, 8), the authors reported that *G. gallus* gH5 binds on-dyad to a nucleosome with a Widom 601 DNA sequence, whereas *D. melanogaster* gH1 binds off-dyad to the same nucleosome. Interestingly, in a follow-up study, by using low resolution spin labeling experimental constraints Zhou et al. (17) suggested that the off-dyad binding mode of the *G. gallus* gH5 to the nucleosome could be switched to an on-dyad binding mode by introducing a penta-mutation in the *G. gallus* gH5. Thus, it is important to understand the sequence dependence of the structure of the chromatosome which can be composed of various LH isoform and nucleosome sequence combinations.

Various experiments suggest specific effects of LH variants on DNA binding and chromatin condensation. Orrego et al. (166) reported up to 19-fold differences in affinity to chromatin for LH H1 variants and Clausell et al. (167) obtained similar results from atomic force microscopy. Brown and colleagues used mutagenesis and fluorescence recovery after bleaching (FRAP) to map the regions affecting chromatin-binding affinity in H1.1 - H1.5 and to identify distinct nucleosome binding surfaces in H1c and H10 (44, 146). It was also found that individual LH variants can trigger apoptosis (168) and are differentially expressed during stem cell differentiation, cell cycle progression and proliferation (169, 170). The specificities and genomic distribution of LH variants was recently reviewed by Kowalski et al. (171, 48) and Millán-Ariño et al. (49). These data suggest that LH variants may have distinct functions because of different nucleosome interaction and chromatin compaction mechanisms.

In Chapter 1.5, LH PTMs (methylation, acetylation, ADPribosylation, ubiquitination, formylation and PARylation) and their phenotypic effects were introduced. In order to understand the phenotypic effect of each LH PTM, it is crucial to determine effects of LH PTMs on chromatin compaction. For this purpose, computational methods can be applied.

A range of computational approaches has been used to model and simulate LH - nucleosome complexes. Mesoscale simulations have been applied to explore the influence of LH concentration, conformation and nucleosome interactions on chromatin structure as well as the dependence of LH - chromatin interactions on salt concentration (63, 143, 157, 172–174). Most approaches to obtain atomic-detail structures have employed computational docking subject to experimental constraints (14, 70, 144). Most recently, Zhou et al. (17, 45) used HADDOCK (69) and Bednar et al. (9) used Autodock Vina (175) to determine structures of LH GD–nucleosome complexes based on experimental constraints. In Chapter 3, I have shown that Brownian dynamics (BD) rigid-body docking can be used for electrostatically driven macromolecular docking to generate diffusional encounter complexes (137) and could be used without experimental constraints to generate structures of *G. gallus* gH5-nucleosome encounter complexes that were consistent with the available experimental data (73). Then, atomic-detail molecular dynamics (MD) simulations were performed

starting from the BD encounter complexes, that by taking LH GD and nucleosome flexibility into account, revealed a binding mechanism involving conformational selection and induced fit (137). In the bound complex with an off-dyad position of *G. gallus* gH5, it was found that the gH5 β_1 loop V78 makes hydrophobic contacts with the DNA and stabilizes the complex (137). There are exchanges of positive with hydrophobic residues at three positions in the β_1 loop of the LH between *G. gallus* gH5 and *D. melanogaster gH1* sequences (Figures 4.1 and 4.2) suggesting that mutants with single point mutations on the β_1 loop could help to understand the determinants of chromatosome structure.

While atomic detail and coarse-grained molecular dynamics simulations have been applied to study the effects of PTMs of core histone tails on protein binding (172, 176) on nucleosome structure (177, 178) and on internucleosome interactions (172) no such studies have yet been reported for variants or PTMs of LHs. In this study, the BD docking approach is applied to investigate the effects of sequence variation and PTMs on the binding configurations of *G. gallus* gH5 and *D. melanogaster* gH1 to the nucleosome. The computational efficiency of the BD approach allows to consider a number of mutations and PTMs. Moreover, docking is performed for different nucleosome conformations, allowing the relation between LH binding mode and nucleosome opening to be explored. The disordered N- and C-terminal domains of the LH are not included in the models as it has been shown that they do not affect the location of the GD LH on the nucleosome (9, 17) although the C-terminus affects the affinity (147).

First, the BD docking protocol is validated by testing its reproduction of crystallographic structures of chromatosomes. Then, single point mutations are introduced into both LHs, and by docking the mutants to nucleosome structures, residues that switch chromatosome configurations were identified. Furthermore, the effects of *D. melanogaster* gH1 PTMs on LH – nucleosome binding and the distribution of the chromatosome structural ensemble are analyzed.

```
SP|P02259|H5_CHICKEN     22   SHPTYSEMIAAAIRAEKSRGGSSRQSIQKYIKSHYKVGHNA-DLQIKLSIRRLLAAGVLKQTKGVGASGSFRLAK   97
SP|P02255|H1_DROSOPHILA  45   SHPPTQQMIDAAIKNLKERGGSSLLAIKKYITATYKVDAQKLAPFIKKYLKSLVVNGKLIQTKGKGASGSFKLS-  118
SP|P22844|H1_XENONPUS    25   -HPKYSDMILAAVQAEKSRSGSSRQSIQKYIKNHYKVGENA-DSQIKLSIKRLVTSGALKQTKGVGASGSFRLAK   97
                               **  .:** **::  *.*.***  :*:***.  ***. :     **  :: *:. * * **** ******:*:
```

Figure 4.1: Sequence alignment of the globular domains (GD) of the *G. gallus* (chicken) H5, *D. melanogaster* H1 and *X. laevis* H1 isoforms. The sequence identity of the three LH GD structures is 45%. The secondary structure of the GDs is shown above the alignment. Uniprot accession numbers are given at the beginning of each row. Residues that are mutated in *G. gallus* gH5 and *D. melanogaster* gH1 in this work are shown in red. Residues that are post-translationally modified in *D. melanogaster* gH1 are shown in blue (see Figure 4.2). Note that Zhou et al. used a *D. melanogaster* gH1 construct which has core stabilizing mutations shown in magenta (8). For consistency, the same construct was used in simulations for the reference WT *D. melanogaster* gH1.

4.2 Materials and methods

Five sets of systems for BD docking simulations were prepared, see Table 4.1. Each system consisted of a LH GD structure and a nucleosome structure to which the LH GD was docked:

Table 4.1: Systems used in BD docking simulations**.** The 5 different simulation systems and the details of their structural components are given. See Figure 4.5 for a comparison of the three different DNA sequences in the nucleosomes studied.

Nucleosome structures	DNA sequence	Core histones	Number of nucleosome conformations for docking	L-DNA length (bp)	LH globular domain with conformation in parentheses	BD simulations
Crystal structure (PDB id: 4QLC, Zhou et al. (7)) and structures from NMA	Widom 601	*D. melanogaster*	3	10	*G. gallus* gH5 (closed)	protocol validation (Figure 4.6A)
Crystal structure (PDB id: 5NL0, Bednar et al. (9)) and structures from NMA	Widom 601L	*X. laevis*	3	26	*X. laevis* gH1 (closed)	protocol validation (Figure 4.6B)
MD snapshots (Öztürk et al. (137) Based on PDB ids: 1KX5 and 1ZBB	palindromic *H. sapiens* X chromosome α-satellite sequence	*X. laevis*	8	10	*G. gallus* gH5 (closed)	gH5 mutants: V80K K82I K85V V87K (Figure 4.7A)
ditto	ditto	ditto	ditto	ditto	*D. melanogaster* gH1 (same as Zhou et al. (8))	gH1 mutants: K102V I104K K107 K109V (Figure 4.7B)
ditto	ditto	ditto	ditto	ditto	*D. melanogaster* gH1 (same as Zhou et al. (8))	gH1 PTMs: K58 dimethylation S66 phosphorylation S67 phosphorylation K72 dimethylation (Figure 4.7C)

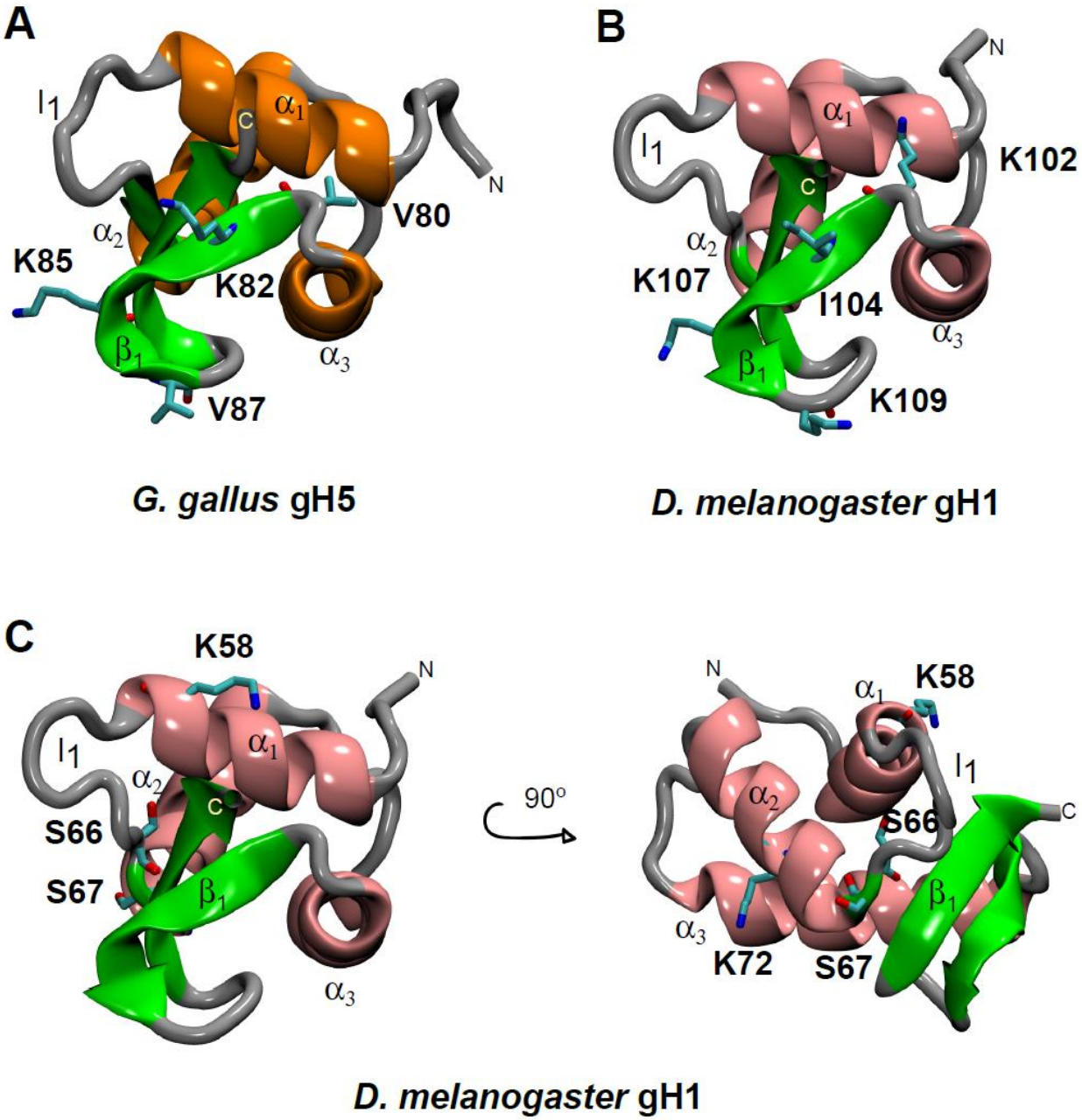


Figure 4.2: Structures of the LH GDs studied. **A-** *G. gallus* gH5 showing positions of mutated residues: V80, K82, K85 and V87. **B-** *D. melanogaster* gH1 showing positions of mutated residues: K102, I104, K107 and K109. **C-** *D. melanogaster* gH1 in two orientations showing sites of PTMs: K72dimethylation, S67phosphorylation, S66phosphorylation and K58dimethylation. LH proteins are shown in cartoon representation and colored according to secondary structure: α helices in orange or pink, β sheets in green and unstructured regions in gray. Mutated side chains are shown in stick representation with coloring by atom-type.

4.2.1 LH – nucleosome complex structures

Recently, two crystal structures were reported for LH GD – nucleosome complexes (PDB ids: 4QLC (7) and 5NL0 (9)). To confirm the validity of the computational protocol, these structures were used as control test systems. From each structure, two PDB files were created, one for the nucleosome and one for the LH GD. Conformational variability of the nucleosome was considered as done previously (73) by generating a set of structures by performing an elastic network normal mode analysis (NMA) using the NOMAD - Ref server (105). For the nucleosome structure from PDB id: 5NL0 the following parameters were used: number of modes to calculate: 106; distance weight parameter for elastic constant: 5 Å; elastic network model cutoff for mode calculation: 10 Å; average RMSD in output trajectories from the initial structure: 3 Å; calculation method: all atom and automatic. For the nucleosome from PDB id: 4QLC, the same nucleosome structures obtained with the default NOMAD – Ref server parameters as used in previous study were applied (137) which were, number of modes to calculate: 16; distance weight parameter for elastic constant: 5 Å; elastic network model cutoff for mode calculation: 10 Å; average RMSD in output trajectories from the initial structure: 1 Å; calculation method: all atom and automatic. The output structures of the nucleosomes were named mode 7_0 (crystal structure), mode 7_1 and mode 7_2 and correspond to snapshots along the lowest frequency mode (mode 7) with increasingly more open L-DNA arms (See Table 4.2).

As homology modeled LH structures were used to fit the LH densities in the recent crystal structures, the LH structures extracted from these PDB files were refined using the GalaxyRefine web-server tool (179) to increase the structural quality of the side chains of the LHs by using the ‘mild relaxation only’ option. The GalaxyRefine tool rebuilds sidechains and performs sidechain repacking and structure relaxation with a molecular dynamics simulation-based protocol. The tool was ranked best for improving the local structure quality in the CASP10 assessment (179). In all refinements, the all-atom RMSD of the input and output LH structures of the GalaxyRefine tool was below 2 Å.

Table 4.2: L-DNA opening angles are given for each L-DNA arm for the conformations derived from NMA using the nucleosome from the crystal structure with PDB id: 4QLC determined by Zhou et al. (1) and PDB id: 5NL0 determined by Bednar et al. (2). The conformations were generated along the first internal motion mode (mode 7, modes 1-6 correspond to rigid body translation and rotation). Mode 7_0 corresponds to the crystal structure. Mode 7_1 and mode 7_2 represent increasingly more open structures of the nucleosome. These three structures were used in BD docking to obtain the results given in Figure 4.6. See Figure 4.3 for the definitions of the angles.

Nucleosome structure	L-DNA1 γ_1(°)	L-DNA1 γ_2(°)	L-DNA2 γ_1(°)	L-DNA2 γ_2(°)
Crystal structure (4QLC) (Mode 7_0)	56.0	98.9	111.2	78.3
Mode 7_1	57.1	99.8	105.7	77.4
Mode 7_2	58.5	101.9	99.9	76.2
Crystal structure (5NL0) (Mode 7_0)	60.5	84.5	108.7	83.7
Mode 7_1	62.4	85.2	105.6	82.5
Mode 7_2	64.1	85.9	102.6	81.3

4.2.2 Apo-nucleosome structures

The 8 snapshots from the MD simulation of an apo-nucleosome previously used in Chapter 3 for BD rigid body docking simulations were used. In Chapter 3, I showed that the BD docking to these snapshots resulted in similar LH binding configurations to those obtained for nucleosome structures generated by normal mode analysis (NMA) by Pachov et al. (73). These nucleosome structures were derived from the crystal structures with PDB id: 1KX5 (1.9 Å resolution) (43) for the nucleosome core particle, and with PDB id: 1ZBB (9 Å resolution) (66) for the 10 bp extensions of each L-DNA. Snapshots of nucleosome structures from MD simulations were prepared previously (137) by using the following procedure: The N-DNA was extended with two L-DNA arms and core histone tails were removed. Nucleosome dynamics were simulated for 100 ns by standard molecular dynamics simulation. After clustering of structures from the trajectory, 8 different snapshots were selected to cover the conformational space of the nucleosome, in which the L-DNA2 arm was in a highly populated conformation and the conformation of the L-DNA1 arm varied (for details of the nucleosome structures, see Figures 4.3 – 4.4 and Table 4.3. The 8 nucleosome structures have different L-DNA1 arm opening and closing angles: snapshot 5 represents the average structure and snapshots 6, 7 and 8 have a more open L-DNA1 arm and snapshots 1, 2, 3 and 4 have a more closed L-DNA1 arm (see Figures 4.3 – 4.4 and Table 4.3).

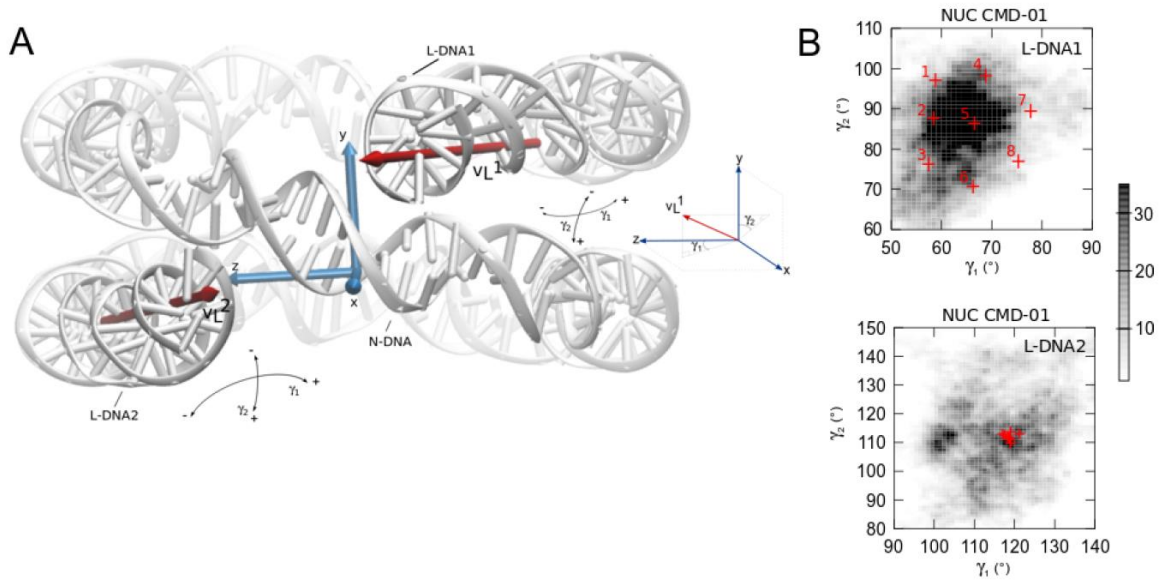


Figure 4.3: Calculation of L-DNA opening angles. **A-** The vectors v_L^1 and v_L^2 were defined based on selected DNA bases to represent the helical axes of L-DNA1 and L-DNA2, respectively, Öztürk et al. (137). The double headed arrows show the directionality of the L-DNA motions described by the two angles (γ_1 and γ_2). **B-** 8 Snapshots selected from a 100 ns standard MD simulation of the free nucleosome selected for BD simulations (labeled in red) on the γ_1 and γ_2 histograms for L-DNA1 and L-DNA2, Öztürk et al. (137). Snapshots 1-4 have more closed conformations of the nucleosome and snapshots 6-8 have more open conformations of the nucleosome compared to snapshot 5, see Figure 4.4. (Figure 4.3 is re-printed from Öztürk et al. (137) under Open access CC BY license.)

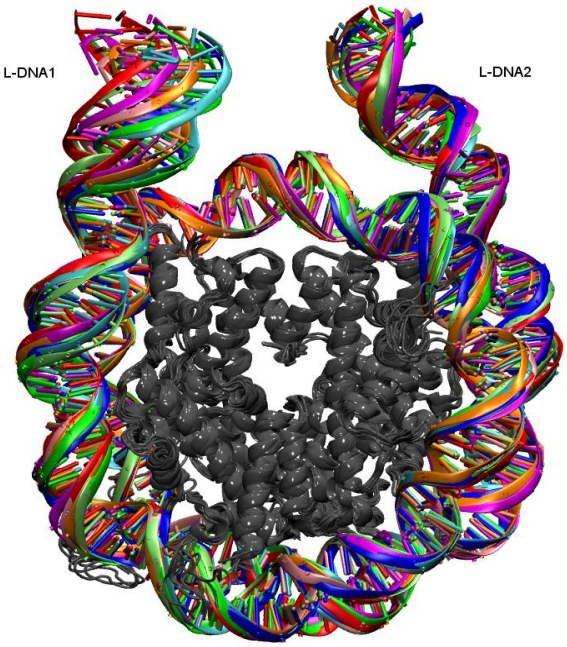


Figure 4.4: Superposition of the 8 nucleosome snapshots selected from 100 ns standard MD simulation of the nucleosome for BD simulations (see Figure 4.3). The DNA is colored according to snapshot and the core histones are shown in cartoon representation in gray. Snapshots 1-4 (1-blue, 2-green, 3-cyan, 4-lime) have more closed conformations of the nucleosome and snapshots 6-8 (6-pale pink, 7-red, 8-magenta) have more open conformations of the nucleosome compared to Snapshot 5 (orange).

Table 4.3: L-DNA opening angles are given for each L-DNA arm for the nucleosome structures used for the BD docking simulations with results given in Figure 4.7. The L-DNA1 arm is in a relatively closed form in nucleosome structures 1, 2, 3 and 4 and in a more open form for nucleosome structures 6, 7 and 8 (see Figures 4.3 and 4.4 for further details.).

Nucleosome structure	L-DNA1 γ_1(°)	L-DNA1 γ_2(°)	L-DNA2 γ_1(°)	L-DNA2 γ_2(°)
1	58.7	97.0	119.0	109.7
2	58.4	87.6	118.3	111.8
3	57.4	76.2	118.7	111.5
4	68.8	98.2	119.0	110.9
5	66.5	86.4	121.2	112.9
6	67.2	77.2	118.6	110.0
7	77.7	89.4	117.0	112.9
8	75.2	76.9	118.9	113.7

4.2.2.1 DNA sequences of the nucleosomes used

Structure based pairwise sequence alignments of the DNA in the nucleosomes used in this study are given below.

```
4QLC          L-DNA2 ---------------ACTGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGT        45
5NL0          L-DNA2 ACTACGTAATATTGGCCAGCTAGGATATCACAATCCCGGTGCCGAGGCCGCTCAATTGGT        60
                             * *   *      * ****************************

4QLC                 CGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGC       105
5NL0                 CGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGGAATCCGTACGTGCGTTTAAGCGG       120
                     ************************************   *     *     ***** **

4QLC                 CAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTGCATGTAA       165
5NL0                 TGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTGTGATATCCTAGCT       180
                           *   *     * *        **                              *

4QLC                 GT--------------- L-DNA1 167
5NL0                 GGCCAATATTACGTAGT L-DNA1 197
                     *


4QLC          L-DNA2 ACTGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAG        60
1KX5/1ZBB     L-DNA2 ACTGGCCGCCCTGGAGAATCACCTGCAGATACTACCAAAAGTGTATTTGGAAACTGCTCC        60
                     ******************** *              *  ** **    * ** *   **

4QLC                 CACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCC       120
1KX5/1ZBB            ATCAAAAGGCATGTTCAGCTGGAATCCAGCTGAACATGCCTTTTGATGGAGCAGTTTCCA       120
                       *         *  *    *   *    *      *           * * * *   *

4QLC                 CTAGTCTCCAGGCACGTGTCAGATATATACATCCTGTGCATGTAAGT L-DNA1             167
1KX5/1ZBB            AATACACTTTTGGTAGTATCTGCAGGTTACATCCTGTGCATGTAAGT L-DNA1             167
                           *   ** ** *     ********************
```

```
5NL0          L-DNA2 ACTACGTAATATTGGCCAGCTAGGATATCACAATCCCGGTGCCGAGGCCGCTCAATTGGT        60
1KX5/1ZBB     L-DNA2 ---------------ACTGGCCGCCCTGGAGAATCACCTGCAGATACTACCAAAAGTGTA        60
                                     * *   *      * **** *             *  ** **

5NL0                 CGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGGAATCCGTACGTGCGTTTAAGCGG       120
1KX5/1ZBB            TTTGGAAACTGCTCCATCAAAAGGCATGTTCAGCTGGAATCCAGCTGAACATGCCTTTTG       120
                       * ** *   **    *       * *  *    *******    *  * *       *

5NL0                 TGCTAGAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTGTGATATCCTAGCT       180
1KX5/1ZBB            ATGGAGCAGTTTCCAAATACACTTTTGGTAGTATCTGCAGGTTACATCCTGTGCATGTAA       167
                         **   * ** *    ** **  *             *  *         * *

5NL0                 GGCCAATATTACGTAGT L-DNA1 197
1KX5/1ZBB            GT--------------- L-DNA1 167
                     *
```

Figure 4.5: Structure based pairwise sequence alignments of the DNA in the nucleosomes used in this study. The systems studied are listed in Table 4.1 and referred to by the PDB identifier. The L-DNA1 and L-DNA2 linker DNA arms are labeled by bold and the nucleotides at the dyad point are highlighted in green. Pairwise alignments are shown as 4QLC and 5NL0 are similar (51% sequence identity) while 1KX5/1ZBB is rather different in sequence (42% sequence identity to 4QLC and 31% to 5NL0).

4.2.3 LH globular domain structure

The refined *G. gallus* gH5 crystal structure (PDB id: 1HST, chain B (5)) was used for docking to the apo-nucleosome structures. The V80K, K82I, K85V and V87K mutations were introduced into *G. gallus* gH5. The structure of *D. melanogaster* gH1 as reported in Zhou et al. (2013) (8) was kindly obtained from Yawen Bai and the K102V, I104K, K107V and K109V mutations were introduced into *D. melanogaster* gH1 by using the PyMOL molecular modeling software (180) (Figures 4.1, 4.2A and 4.2B). After introduction of the mutations, the GalaxyRefine structure refinement webserver tool (179) was used as described above to refine the structure of each mutant.

Additionally, the K72dimethylation, S67phosphorylation, S66phosphorylation and K58dimethylation PTMs (Figure 4.2C) (87) were introduced into the *D. melanogaster* gH1 by applying the PyTMs plugin in PyMOL (180, 181) to the refined WT structure. As the GalaxyRefine web server only accepts standard amino acids, partial atomic charges and radii of the post-translationally modified residues were obtained from previously published studies (182, 183) and added manually to the PQR files generated for these structures without further refinement.

4.2.4 BD simulations

For BD simulations, polar hydrogen atoms were added to the structures by using the PDB2PQR 2.1.1 web-server (155) and partial atomic charges and atomic radii were assigned by using the AMBER99 force field (184). For all structures, the molecular electrostatic potentials were calculated by using APBS 1.4 (156) to solve the non-linear Poisson–Boltzmann equation with a 1 Å grid spacing. Input parameters were a temperature of 298.15 K, solvent and solute relative dielectric

constants of 78.54 and 2, respectively, and an ionic strength of 100 mM. The van der Waals surface was used to define the dielectric boundary. BD simulations were performed by using the SDA 7 software (130) with electrostatic interaction forces and neglecting short-range interactions. In order to prevent overlap of structures, an excluded volume criterion was applied. Effective charges were assigned to charged residues on the protein and to P atoms on the DNA using the ECM program (133). The BD trajectories were started randomly on a sphere at a center-to-center distance of ~185 Å and stopped at a center-to-center distance of ~204 Å. For each system, 20000 BD trajectories were generated. My test runs for 10000 BD trajectories and for different initial random number seeds result in similar cluster configurations and population percentages. The following two geometric conditions were used to define the diffusional encounter complexes: (i) the geometric center-to-center distance of LH and the nucleosome < 98 Å, and (ii) the nucleosome dyad point and LH center separation < 40 Å. The coordinates and interaction energies of the complex were recorded if the RMSD to the previously recorded complexes was > 1 Å and the interaction energy was within the 5000 most favorable energy complexes recorded. A complex with RMSD < 1 Å to a previously recorded complex but lower energy was recorded as a substitute of that complex; higher energy complexes were added to the count of occurrence of the closest recorded complex with lower energy. Finally, the top 5000 lowest energy structures were clustered into 10 groups which were ranked according to cluster size, taking the number of counts for each complex recorded into account.

4.2.5 Analysis of docked encounter complexes

The configuration of the LH on the nucleosome was classified for the representative structures of the first two largest clusters of encounter complexes with the highest populations obtained in each docking simulation by applying the following procedure: The nucleosome dyad axis was aligned perpendicular to the viewing plane and the DNA grooves were labeled. The minor groove on the dyad was labeled as 0, the neighboring major grooves of N-DNA toward L-DNA1 and L-DNA2 were labeled as -1 and +1, respectively. The adjacent major grooves on the L-DNA1 and L-DNA2 were labeled as -2 and +2, respectively, and so on to the ends of the L-DNA arms. The DNA groove contacts of the structural elements to the LH (α_3, β_1 and l_1) were computed for the representative structure of each docking cluster and represented by a vector. The orientation of the α_3 helix of LH was determined and an arrow added to the vector to represent the direction of the vector from the N- to the C-terminus of the α_3 helix. An X sign was used when the α_3 helix vector was perpendicular to the viewing plane. See Figure 4.8A for an example of the analysis for the crystal structure, PDB id: 4QLC, in the configuration (-3 ↑, 0, +3).

LH configurations on the nucleosome that differed for at least two of four of the structural determinants given in the vector with respect to the WT LH configuration and had a cluster population

above 25% were considered as configuration shifts compared to WT LH. The PyMOL software (180) was used to quantify hydrogen-bonding (with a distance criterion of 3.2 - 3.6 Å) between the LH and the nucleosome structure.

4.3 Results and discussion

4.3.1 BD simulations recapitulate experimentally determined configurations of LH – nucleosome complexes

First, the ability of the combined LH structural refinement and BD docking protocol to reproduce the experimentally determined LH-nucleosome complex structures was tested. Zhou et al. (2015) published the first crystal structure of the LH – nucleosome complex of *G. gallus* gH5 binding to a nucleosome with a Widom 601 sequence in an on-dyad binding mode (PDB id: 4QLC) (7). The protocol of structural refinement of the LH followed by BD rigid-body docking was applied. For comparison, it should be born in mind that the docking protocols generate diffusional encounter complexes which are expected to be close to but not identical to the bound structures studied experimentally. In particular, the rigid-body docked complexes are expected to be looser and will lack optimization of short-range hydrogen-bonds and hydrophobic contacts. Therefore, the structures were compared by using a classification of the binding configurations based on LH-nucleosome contacts, rather than commonly used measures based on root mean squared deviations. The docking results are given in Table 4.4 and Figure 4.6. In the current docking simulations, and in the previous BD docking study (137), WT *G. gallus* gH5 binds to the nucleosome from the 4QLC structure in an on-dyad configuration. The orientation of gH5 corresponds to that in the crystal structure in the largest encounter complex cluster (cluster 1) obtained by docking gH5 to the nucleosome of the LH – nucleosome complex crystal structure PDB id: 4QLC (mode 7_0) and to the slightly more open mode 7_1 structure. It should be noted that for the same system in the previous docking simulations (137) a LH refinement protocol and some opening of the nucleosome were not applied, as represented here by the mode 7_1 and mode 7_2 snapshots, but was necessary to allow access of the LH to the nucleosome dyad axis and to reproduce the crystallographic binding mode. This opening of the nucleosome was not necessary for the refined LH structure to bind in the crystallographic binding mode although binding in this orientation was facilitated by the slight opening in the mode 7_1 structure.

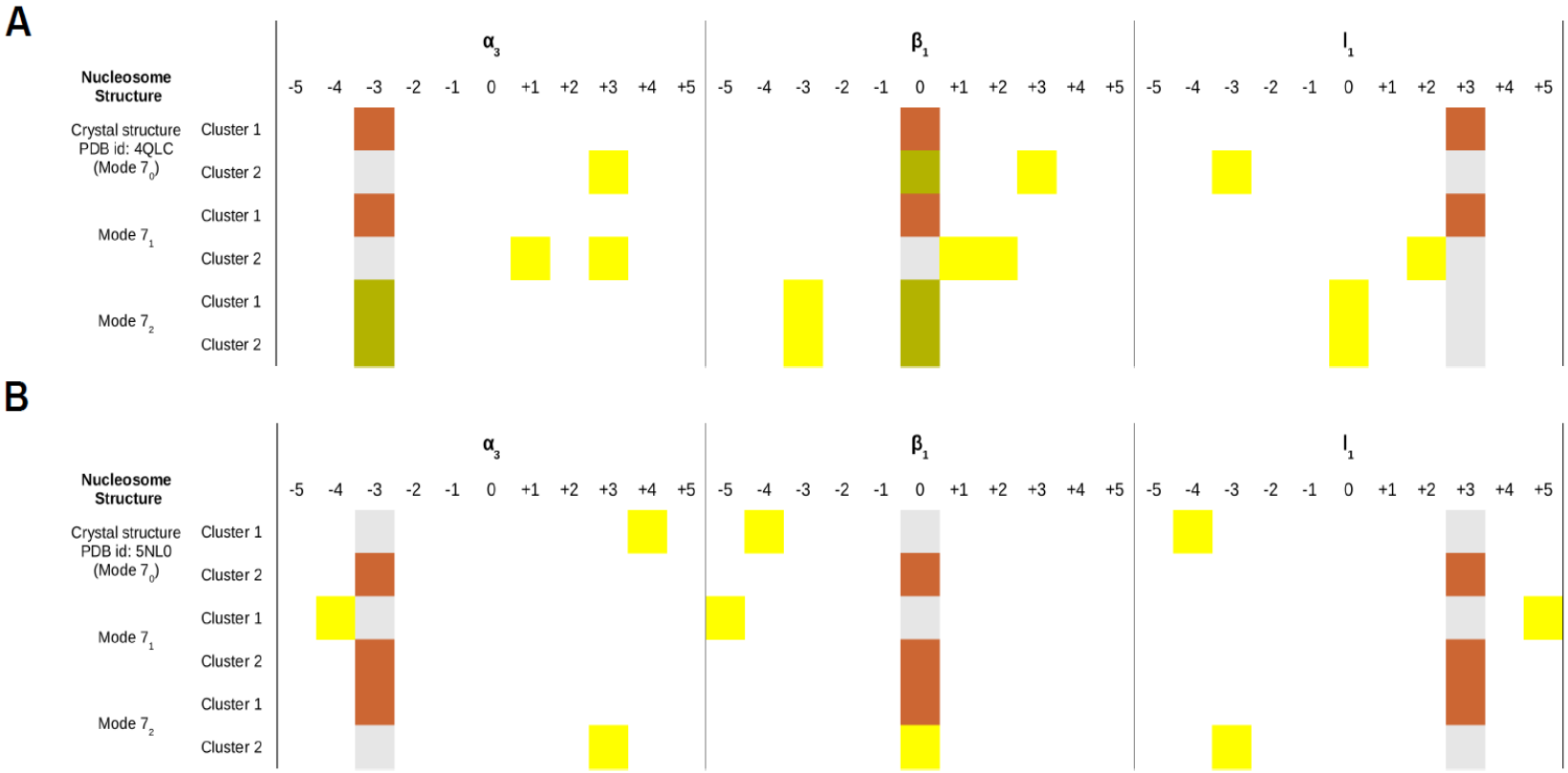


Figure 4.6: Comparison of BD docked LH GD-nucleosome encounter complexes with crystal structures of the complexes. **A-** Docking of WT *G. gallus* gH5 to the nucleosome (PDBid: 4QLC, Zhou et al. **(7)**), **B-** Docking of WT *X. laevis* gH1 to the nucleosome (PDB id: 5NL0, Bednar et al. **(8)**). For each system, the orientations of the representative structures of the largest two clusters of docked encounter complexes are given for docking of the LH GD to the nucleosome crystal structure (mode 7_0) and two structures generated by normal mode analysis with slightly opened L-DNA arms (modes 7_1 and 7_2) (See Methods and Table 4.2 for details). The DNA grooves on the nucleosome in contact with each structural element of the LH (α_3, β_1 and l_1) are given in the respective columns (see Figures 4.2 and 4.8A). Color code: The DNA groove contacts of the α_3, β_1 and l_1 elements are given in gray for the crystal structures, in orange when BD results match with the DNA groove contacts of the crystal structures for all three structural elements, in green when the BD results partially (only for 1 or 2 structural elements) match with the crystal structures, and in yellow when the BD results have different DNA contacts from the crystal structures. For encounter complex cluster populations, see Table 4.4.

Table 4.4: BD docking simulations of LH binding to the nucleosome in the LH – nucleosome complex crystal structures, PDB ids: 4QLC (7) and 5NL0 (9). The orientations of the representative structures of the largest two clusters of encounter complexes obtained for WT *G. gallus* gH5 docking to the Zhou et al. (PDBid: 4QLC) nucleosome and WT *X. laevis* gH1 docking to the Bednar et al. nucleosome (PDB id: 5NL0) are given for docking to the nucleosome crystal structure (Mode 7_0) and two structures (Mode 7_1 and Mode 7_2) with slightly opened LDNA arms.

Nucleosome structure (*)	gH5 WT					Nucleosome structure (*)	H1 WT				
	N	%	α_3	β_1	l_1		N	%	α_3	β_1	l_1
Reference (4QLC)			-3 ↑	0	+3	Reference (5NL0)			-3 ↑	0	+3
Crystal structure (4QLC) (Mode 7_0)	2.0	35	-3 ↑	0	+3	Crystal structure (5NL0) (Mode 7_0)	1.5	39	+4 ↖	-4	-4
		31	+3 ↘	0, +3	-3			27	-3 ↑	0	+3
Mode 7_1	1.4	79	-3 ↑	0	+3	Mode 7_1	1.2	46	-4 →	-5	+5
		8	+1, +3 ↑	+1, +2	+2			34	-3 ↑	0	+3
Mode 7_2	2.0	36	-3 ↑	0, -3	0	Mode 7_2	1.0	85	-3 ↑	0	+3
		28	-3 ↑	0, -3	0			8	+3 ↓	0	-3

The DNA grooves on the nucleosome in contact with each structural element of LH (α_3, β_1 and l_1) are given in the respective columns (See Figure 4.2A and 4.8A). The arrows show the orientation of the LH α_3-helix when the nucleosome is aligned perpendicular to the viewing plane. N is the total number of encounter complexes that satisfy the docking criteria in each BD simulation, divided by 10^6. The % of these encounter complexes in the two largest BD clusters is given in the % column for clusters 1 (above) and 2 (below). (*) Normal Mode Analysis of an elastic network model of the nucleosome was performed. The crystal structure corresponds to mode 7_0 and modes 7_1 and 7_2 are structures along the slowest mode (mode 7) that have more open L-DNA arms than the crystal structure (See Methods for details).

Bednar et al. published the crystal structure of *X. laevis* gH1 binding to a palindromic Widom 601L nucleosome in an on-dyad mode (PDB id: 5NL0) (9). By applying the LH refinement and docking protocol, it was possible to reproduce the configuration observed in the crystal structure (modes 7_0-7_2) in the first or second encounter complex cluster (-3 ↑ ,0 +3) (Figure 4.8C and Table 4.4). Interestingly, the number of encounter complexes observed in each docking simulation was somewhat lower than for the gH5 simulations (1 - 1.5 million compared to 1.4 - 2.0 million), indicating that the LH binding site was less accessible, possibly because of the longer L-DNA arms (26 bp vs 10 bp). Consistently, in the docking simulations to the crystal structure and mode 7_1 of the nucleosome, the α_3 helix of the LH binds to the nucleosome DNA grooves +4 and -4, respectively. When the L-DNA arms open further in the mode 7_2 nucleosome structure, the LH can approach closer to the LH core and the LH α_3 helix binds predominantly to nucleosome DNA groove -3, as

observed in the crystal structure. This indicates that further conformational relaxation of the LH and nucleosome should stabilize these on-dyad binding modes.

Summarizing, the diffusional encounter complex structures generated by BD docking simulations are largely consistent with the crystallographic results of Zhou et al. and Bednar et al. (7, 9) for 2 different LH - nucleosome systems. Previously both on- and off-dyad chromatosome configurations were obtained by BD docking simulations using nucleosome structures generated by normal mode analysis and by MD simulation (73, 137). Therefore, the BD docking approach was applied to investigate the effects of mutations and post-translational modifications on LH – nucleosome binding configurations.

4.3.2 Single point mutations in the LH globular domain can significantly affect chromatosome structure

BD docking results for *G. gallus* gH5 and *D. melanogaster* gH1 binding to the 8 representative nucleosome structures generated by MD simulation (Figures 4.3 and 4.4) are given in Figure 4.7. The nucleosome structures have different extents of opening and closing of the L-DNA1 arm with respect to the average simulated apo-nucleosome structure (snapshot 5); snapshots 1, 2, 3 and 4 are more closed structures and snapshots 6, 7 and 8 are more open structures (see Figures 4.3 and 4.4). The binding mode of the largest encounter complex cluster obtained for the LH variants is compared with that for the WT LH GD for each of the nucleosome structures (See Tables 4.5 and 4.6 for the results for the two largest encounter complex clusters and their populations). Mutant LH-nucleosome configurations that differ significantly from the configurations of the WT LH GD are highlighted in yellow whereas those that are conserved are highlighted in orange in Figure 4.7; grey indicates the configuration obtained from docking the WT LH GD and green indicates a partial configuration similarity (only for 1 or 2 structural elements) with the WT LH.

As found before (73, 137), all docked configurations of WT gH5 (with a closed loop) to these nucleosome structures are off-dyad and in the (-1 ↘, -1 -2, -2) (Figure 4.8D) configuration for the nucleosome structure close to the crystal structure, snapshot 5, and the most open structure, snapshot 8. The off-dyad LH binding mode is similar to that identified from NMR (8, 17) and cryo-EM (10) experiments. For the other snapshots, the alternative off-dyad docking position (+1 ↖, +1 +2, +2) dominates, as observed previously (137).

For the gH5 mutants, both off-dyad and on-dyad configurations are observed when considering all 8 nucleosome snapshots (Table 4.2). Interestingly, the on-dyad binding for WT and mutant *G. gallus* gH5 is not observed in open nucleosome conformations, suggesting that L-DNA opening is important for the distinction between off- and on-dyad binding modes of the LH. Additionally, the *G. gallus* gH5 K82I mutation resulted in 4 on-dyad binding modes out of a total of 8

docking simulations. Considering that the gH5 K82 residue is located at the beginning of the β_1 loop of the LH (Figure 4.2A), the unit charge reduction from the isoleucine substitution could reduce contact with the L-DNA arms, and making the on-dyad configuration more preferable for this mutant. Moreover, no binding to L-DNA1 is observed for WT and mutant gH5 docking in the BD simulations for the more closed nucleosome snapshots. For the most open conformers (snapshots 7 and 8), mainly the off-dyad mode with binding to L-DNA1 is observed, due to the opening of the L-DNA1 arm. Remarkably, mutation of gH5 K85 (which is conserved as lysine at the corresponding position in gH1 and *X. laevis* H1) to valine revealed an off-dyad L-DNA2 binding mode (+1 ↖, +1 +2, +2) in 7 out 8 docking simulations, indicating that, in addition to L-DNA opening angles, LH sequence is also a determinant of the chromatosome configuration.

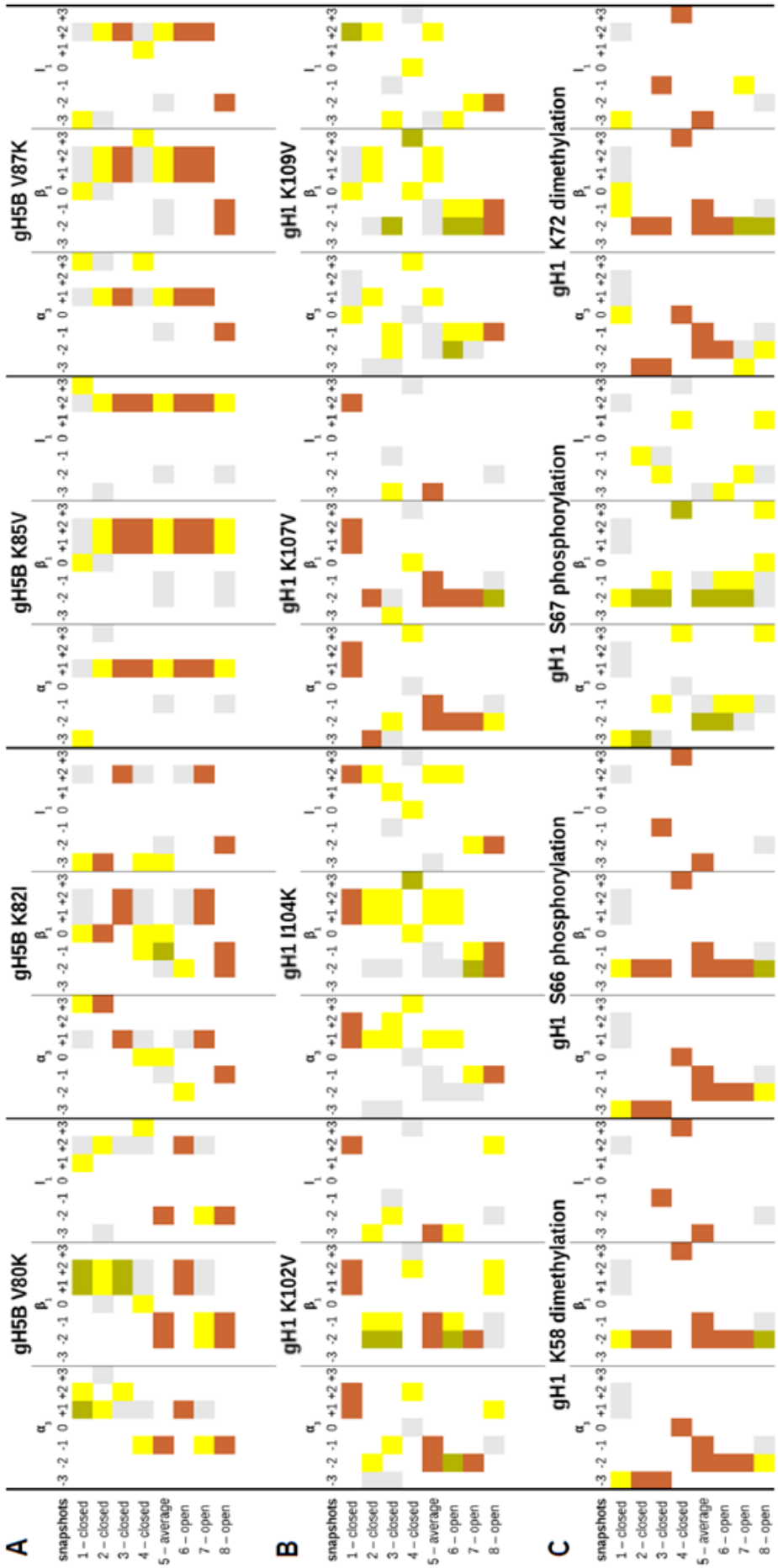


Figure 4.7: Comparison of the BD docked configurations of *G. gallus* gH5 and *D. melanogaster* gH1 variants to those for WT LH GDs upon docking to 8 representative nucleosome structures. **A-** WT and mutant *G. gallus* gH5, **B-** WT and mutant *D. melanogaster* gH1, **C-** WT and post-translationally modified *D. melanogaster* gH1. The figures show the orientations of the representative structures of the largest cluster of encounter complexes from docking each GD LH to 8 representative structures of the nucleosome from an MD simulation started from PDB id:1KX5 (57) (Table 1). The coloring scheme is the same as in Figure 4.6. For cluster populations, see Tables 4.5 – 4.6.

Table 4.5: Docked configurations of WT and mutant *G. gallus* gH5 – nucleosome encounter complexes. The two largest encounter complexes are listed. Docking was performed to the 8 representative structures of the nucleosome from an MD simulation started from PDB id: 1KX5. For details, see Table 4.1.

Snapshot	gH5B WT					gH5B V80K					gH5B K82I					gH5B K85V					gH5B V87K				
	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1
1	2.7	68	+1 ↖	+1,+2	+2	3.8	46	+1 +2 X	+1,+2	+1	7.6	51	+3 ↓	0	-3	5.3	46	-3 ↑	0	+3	4.5	94	+3 ↓	0	-3
		29	-3 ↑	0	+2		35	+1 ↖	+1,+2	+2		9	+3 ↓	0	-3		23	0 ↙	0	+2		3	+3 ↓	0	-3
2	2.8	42	+3 ↓	0	-3	4.5	82	+1 ↑	+1,+2	+2	4.2	26	+3 ↓	0	-3	7.3	59	+1 ↖	+1,+2	+2	3.3	62	+1 ↖	+1,+2	+2
		42	+1 ↖	+1,+2	+2		8	+1, +2 ↙	+1,+2	+3		26	+3 ↓	0	-3		32	+1 ↖	+1,+2	+2		28	-2 ↘	-1,-2	-3
3	5.6	53	+1 ↖	+1,+2	+2	5.7	39	+2 X	+1,+2	no	2.1	59	+1 ↖	+1,+2	+2	6.4	38	+1 ↖	+1,+2	+2	5.4	95	+1 ↖	+1,+2	+2
		37	+1 ↖	+1,+2	+2		20	+2 X	+1,+2	no		28	0 ↓	0, -1	-3		25	+1 ↖	+1,+2	+2		2	-2 ↘	-1,-2	-3
4	3.0	67	+1 ↖	+1,+2	+2	2.1	39	-1 ↑	0	+3	4.1	40	0 ↓	0, -1	-3	3.9	48	+1 ↖	+1,+2	+2	3.9	48	+3 ↙	+3	+1
		17	0, +3 ↓	0, -1	no		28	+3 ↓	0	no		17	0 ↓	0, -1	-3		25	0 ↙	+1	+3		40	+1 ↖	+1,+2	+2
5	4.9	38	-1 ↘	-1,-2	-2	3.7	28	-1 ↘	-1,-2	-2	4.8	45	0 ↓	0, -1	-3	13.9	37	+1 ↖	+1,+2	+2	7.5	48	+1 ↖	+1,+2	+2
		24	-1 ↘	-1,-2	-2		21	+1, +2↙	+1,+2	no		14	0 ↓	0	-3		25	+1 ↖	+1,+2	+2		31	0 ↓	0	-3
6	3.3	58	+1 ↖	+1,+2	+2	5.4	61	+1↖	+1,+2	+2	1.6	85	-2 ←	-2	no	4.3	86	+1 ↖	+1,+2	+2	4.1	73	+1 ↖	+1,+2	+2
		29	+1 ↖	+1,+2	+2		16	+1↖	+1,+2	+2		6	+2 ←	+2	no		6	+1,+2 ↖	+1,+2	no		15	0 ↑	0	+2
7	7.4	27	+1 ↖	+1,+2	+2	9.4	41	-1 ↘	-1,-2	-2	0.7	42	+1↖	+1	+2	8.1	26	+1 ↖	+1,+2	+2	3.1	96	+1 ↖	+1,+2	+2
		20	+1 ↖	+1,+2	+2		37	-1 ↘	-1,-2	-2		31	-1 ↓	0	no		24	+1 ↖	+1,+2	+2		2	-1 ↘	-1,-2	-2
8	5.0	42	-1 ↘	-1,-2	-2	5.6	74	-1 ↘	-1,-2	-2	2.6	33	-1 ↘	-1,-2	-2	6.1	44	+1 ↖	+1,+2	+2	5.0	42	-1 ↘	-1,-2	-2
		28	-1 ↘	-1,-2	-2		10	-1 ↙	-1	-2		23	-1 ↘	-1,-2	-2		16	-1 ↘	-1,-2	-2		28	-1 ↘	-1,-2	-2

Table 4.6: Docked configurations of WT and mutant *D. melanogaster* gH1 – nucleosome encounter complexes. The two largest encounter complexes are listed. Docking was performed to the 8 representative structures of the nucleosome from an MD simulation started from PDB id: 1KX5. For details, see Table 4.1.

Snapshot	dH1 WT					dH1 K102V					dH1 I104K					dH1 K107V					dH1 K109V				
	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1	N	%	α_3	β_1	l_1
1	2.3	25	+1↖	+1, +2	+2	1.9	50	+1↖	+1,+2	+2	2.4	66	+1↖	+1,+2	+2	1.1	39	+1↖	+1,+2	+2	1.6	47	0 ↑	0	+2
		21	-3 ↖	-3	no		21	0 ↓	-1,-3	-3		9	+1↖	+1,+2	+2		25	0, -1 ↓	-1,-2	no		28	0 ↑	0	+2
2	3.3	97	-3 ←	-2	no	1.8	49	-2 ↗	-1,-2	-3	3.9	72	+1↖	+1,+2	+2	1.2	57	-3 ↙	-2	no	1.8	70	+1↖	+1,+2	+2
		2	-1 ↙	-1,-2	-3		20	+1↖	+1,+2	+2		15	-2 →	-1,-2	-2		34	-2 ↗	-3	-3		15	-2 ↗	-1, -2	-3
3	3.9	60	-3 ←	-2	-1	2.7	27	-1 ↙	-1,-2	-2	4.4	65	+1,+2 ↓	+1,+2	0	1.7	55	-2 ↗	-3	-3	2.2	54	-1, -2 ↗	-2	-3
		31	-1 ↙	-1,-2	-2		26	-1, -2 ↖	-1,-2	0		21	-1 ↘	-1,-2	-2		22	0, -1 ↙	-1	-3		23	+2 ↓	+1,+2	no
4	2.7	83	0 ↖	+3	+3	1.3	47	+2 →	+2	no	3.1	92	+3 ↙	0,+3	0	0.9	42	+3 ↙	0	no	0.9	69	+3 ↙	0,+3	0
		4	-3 ↖	-2	-1		20	-3 ↖	-1	+3		6	+1↖	+1,+2	+2		25	-3 ↗	-1	no		14	-1 ↗	0	+3
5	4.0	59	-1, -2 ↗	-1, -2	-3	4.9	36	-1, -2 ↗	-1,-2	-3	4.5	51	+1↖	+1,+2	+2	3.5	22	-1, -2 ↗	-1,-2	-3	1.8	53	+1↖	+1,+2	+2
		27	-2 ↖	-2	no		28	-1, -2 ↗	-1,-2	-3		35	-1, -2 →	-1,-2	-2		21	-1, -2 ↗	-1,-2	-3		28	-1, -2 ↗	-1,-2	-3
6	1.9	63	-2 ←	-2	no	2.1	55	-2 ↗	-1,-2	-3	5.3	95	+1↖	+1,+2	+2	3.9	55	-2 ←	-2	no	2.3	32	-1, -2 ↗	-1,-2	-3
		29	-1,-2 ↗	-1,-2	-3		13	-1, -2 ↗	-1,-2	-3		1	-1 ↘	-1,-2	-2		24	-1, -2 ↗	-1,-2	-3		32	+1↖	+1,+2	+2
7	1.5	49	-2 ←	-2	no	1.9	74	-2 ←	-2	no	5.8	27	-1 ↘	-1,-2	-2	2.6	44	-2 ←	-2	no	1.9	78	-1 ↘	-1,-2	-2
		22	-3 ↖	-2	-1		14	+1↖	+1,+2	+2		24	-1 ↘	-1,-2	-2		30	-2 ←	-2	no		12	-1 ↘	-1,-2	-2
8	4.0	75	-1 ↘	-1,-2	-2	2.2	64	+1↖	+1,+2	+2	3.6	61	-1 ↘	-1,-2	-2	3.3	29	-2 ←	-2	no	1.8	77	-1 ↘	-1,-2	-2
		8	-1 ↘	-1,-2	-2		19	-1 ↘	-1,-2	-2		28	-1 ↘	-1,-2	-2		28	-1, -2 ↗	-1,-2	-3		9	-1,-2 ↘	-1,-2	-3

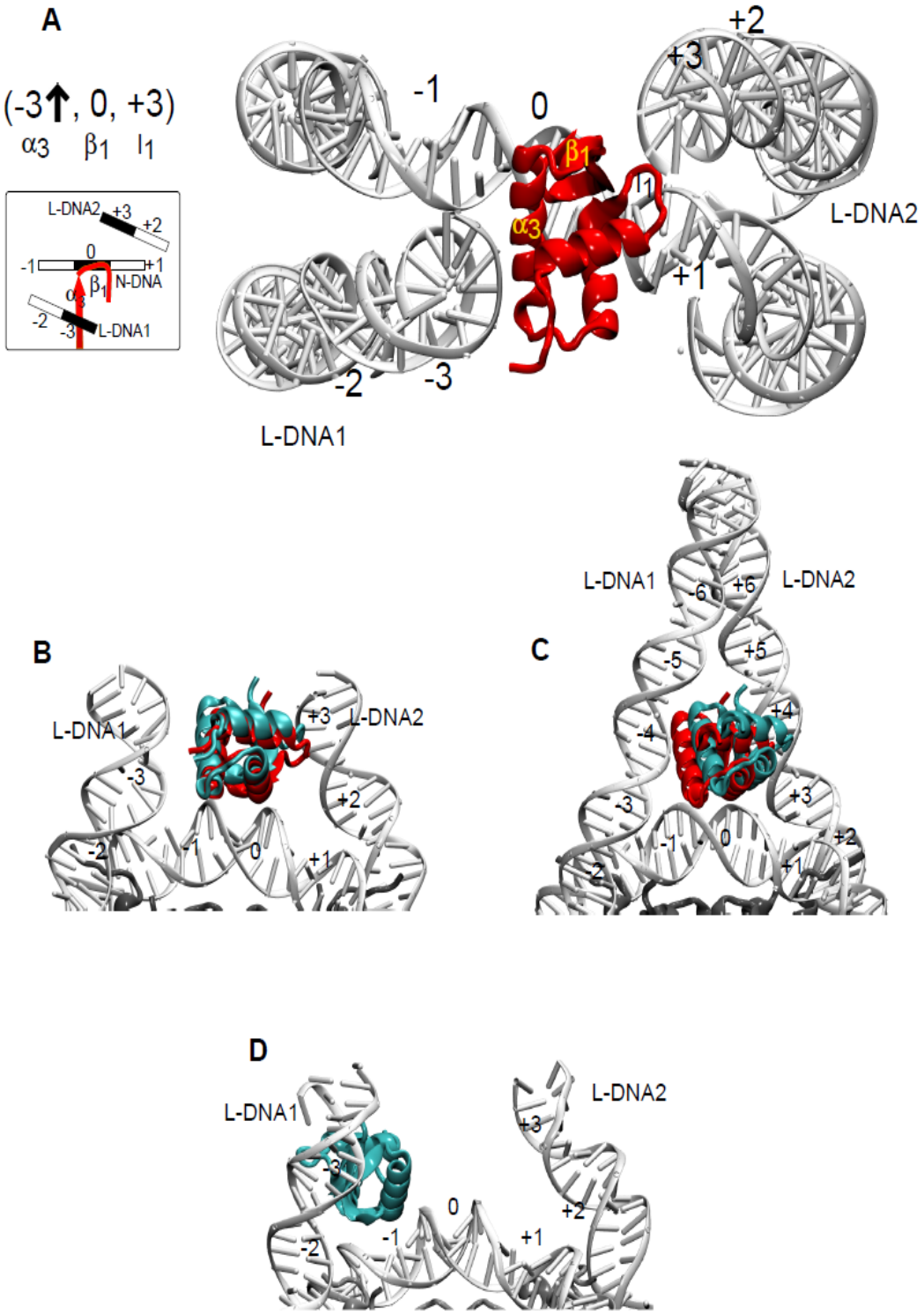


Figure 4.8: Representative LH - nucleosome encounter complexes from BD docking simulations. LHs are shown in cartoon representation in red for reference crystal structures, and cyan for docking results. **A-** Crystal structure of the complex formed by *G. gallus* gH5 bound to a 147 bp Widom 601 DNA sequence nucleosome (PDB id: 4QLC) (7) The classification of the configuration as (-3↑, 0, +3) (on-dyad) is illustrated. **B-** Representative structure from the largest diffusional encounter complex cluster (cluster 1) from docking of *G. gallus* gH5 (residues 24-98) to the mode 7_1 structure of the nucleosome derived by NMA from the crystal structure PDB id: 4QLC (7) shown in A. Compared to the position in the crystal structure (red), the gH5 has a

Cα RMSD of 3.6 Å and the same docked on-dyad configuration (-3↑, 0, +3). **C-** Representative structure from encounter complex cluster 2 for *X. laevis* gH1 docked to the nucleosome structure from PDB id: 5NL0 (9) Compared to the position in the crystal structure (red), the gH1 has a Cα RMSD of 5.5 Å and the same docked on-dyad configuration (-3↑, 0, +3). **D-** Representative structure from the encounter complex cluster with the greatest population (cluster 1) from docking WT *G. gallus* gH5 to snapshot 5 from MD simulation of the nucleosome (see Table 4.1 and Figure 4.4), which represents the average structure in the simulation. The docked configuration is (-1 ↘, -1 -2, -2) and off-dyad.

Table 4.7: Docking configurations of WT and post-translationally modified *D. melanogaster* gH1 – nucleosome encounter complexes. The two largest encounter complexes are listed. Docking was performed to 8 representative structures of the nucleosome from an MD simulation started from PDB id:1KX5. For details, see Table 4.1.

Snapshot	gH1 WT					gH1 K58 dimethylation					gH1 S66 phosphorylation					gH1 S67 phosphorylation					gH1 K72 dimethylation				
	N	%	α_1	β_1	l_1	N	%	α_1	β_1	l_1	N	%	α_1	β_1	l_1	N	%	α_1	β_1	l_1	N	%	α_1	β_1	l_1
1	2.3	25	+1 +2 ↗	+1, +2	+2	2.3	27	-3 ←	-2	no	1.9	35	-3 ←	-2	no	1.4	57	-3 ←	-2	no	2.2	47	0 ↓	0, -1	-3
		21	-3 ↖	-3	no		19	0 ↓	0, -1	-3		14	0 ↓	0, -3	no		13	0 ↙	0, -1	-3		33	0, -1 ↙	-1, -3	-3
2	3.3	97	-3 ←	-2	no	3.7	90	-3 ←	-2	no	3.4	97	-3 ←	-2	no	3.0	43	-3 ←	-2	-1	3.8	96	-3 ←	-2	no
		2	-1 ↙	-1,-2	-3		5	-3 ←	-2	no		1	-3 ←	-2	no		24	-3 ←	-2	-1		3	0 ↓	0, -1	-3
3	3.9	60	-3 ←	-2	-1	3.4	65	-3 ←	-2	-1	2.8	47	-3 ←	-2	-1	2.0	36	-1 ↘	-1, -2	-2	3.9	82	-3 ←	-2	-1
		31	-1 ↙	-1,-2	-2		19	-1 ↘	-1, -2	-2		40	-1 ↘	-1, -2	-2		22	-2 ↖	-2	-1		7	-1, -2 ↓	-1, -2	-2
4	2.7	83	0 ↖	+3	+3	2.3	73	0 ↖	+3	+3	1.8	67	0 ↑	+3	+3	1.6	30	+3 ↙	+3	+1	2.6	70	0 ↖	+3	+3
		4	-3 ↖	-2	-1		6	-3 ↖	-1, -2	-1		13	-3 ↙	0, -3	0		29	+3 ↙	+3	+1		11	0 ↑	+3	+3
5	4.0	59	-1, -2 ↗	-1, -2	-3	3.1	60	-1,-2 ↗	-1,-2	-3	2.7	50	-1,-2 ↗	-1,-2	-3	2.2	59	-2 ←	-2	no	2.3	36	-1,-2 ↗	-1, -2	-3
		27	-2 ↖	-2	no		23	-2 ←	-2	no		30	-2 ←	-2	no		25	+1, +2 ↓	+1,+2	0		20	-1,-2 ↗	-1, -2	-3
6	1.9	63	-2 ←	-2	no	1.7	71	-2 ←	-2	no	1.5	72	-2 ←	-2	no	1.1	33	-1,-2 ↗	-1, -2	-3	2.0	35	-2 ←	-2	no
		29	-1,-2 ↗	-1,-2	-3		20	-1,-2 ↗	-1, -2	-3		16	-1,-2 ↗	-1, -2	-3		23	+2, +3 ↙	0, +3	0, +3		18	-3 ↖	-1, -2	-1
7	1.5	49	-2 ←	-2	no	1.7	55	-2 ←	-2	no	1.4	44	-2 ←	-2	no	1.1	45	-1 ↘	-1, -2	-2	1.9	46	-3 ↖	-2	-1
		22	-3 ↖	-2	-1		25	-3 ↖	-2	-1		36	-3 ↖	-2	-1		34	-3 ↖	-2	-1		39	-3 ↖	-2	-1
8	4.0	75	-1 ↘	-1,-2	-2	3.9	59	-2 ↖	-2	no	3.7	50	-2 ↖	-2	no	4.0	26	+3 ↙	0, +3	+1	3.2	47	-2 ↖	-2	no
		8	-1 ↘	-1,-2	-2		24	-2 ↖	-2	no		21	-2 ↖	-2	no		21	+3 ↙	0, +3	+1		11	-1 ↘	-1, -2	-2

WT gH1 overall adopts a greater diversity of configurations than WT gH5, but all are off-dyad except for the on-dyad docking mode observed for the most closed nucleosome structure, snapshot 4 and for docking of the K109V mutant to a closed nucleosome structure, snapshot 1 (Figure 4.7). Interestingly, all gH1 mutants, except K102V, bind on-dyad to snapshot 4. Furthermore, compared to the WT and mutant gH5 simulations, more of the docking poses for WT and mutant gH1 are off-dyad binding to the L-DNA1 arm. These results suggest that *G. gallus* gH5 and *D. melanogaster* gH1 have distinct nucleosome binding preferences. For most snapshots, the number of encounter complexes recorded is lower for gH1 than gH5, indicating lower accessibility to the nucleosome which also correlates with the higher accessible surface area of gH1 compared to gH5 (3998 $Å^2$ vs 3810 $Å^2$). For the most open structure, snapshot 8, both gH1 and gH5 bind predominantly in the same off-dyad (-1 ↘, -1 -2, -2) configuration.

The effect of LH mutations on chromatosome configuration varies amongst the different snapshots of the nucleosome. For the *G. gallus* gH5 mutants, docking to nucleosome snapshots 1, 4 and 5 resulted in a major configuration shift compared to *G. gallus* gH5 WT for all the mutants. For the *D. melanogaster* gH1 mutants, the LH configuration was most affected (with shifts for all four mutants) compared to WT *D. melanogaster* gH1 for nucleosome snapshot 4, the snapshot with the most closed conformation of the nucleosome. On the other hand, for some snapshots, there were very few shifts in LH-nucleosome configuration upon mutation. For *G. gallus* gH5, only one mutant showed a shift for nucleosome snapshots 6, 7 and 8 whereas for *D. melanogaster* gH1, two mutants showed a shift in nucleosome snapshots 1 and 6. The results show that point mutations may result in a range of changes to LH-nucleosome binding configuration that are dependent on L-DNA opening. The results for gH5 mutants indicate that chromatosome formation for the more open nucleosome structures may be less sensitive to gH5 sequence, which would have implications for LH binding mechanisms in chromatin, the formation of chromatin structure and the phenotypic effects of mutations on LHs.

The applied point mutations involved either the introduction or the removal of a +1 charge from the total +11e charge of the two LHs by the exchange of a lysine residue with a hydrophobic residue. Each single point mutation had a significant effect on LH docking to at least one of the 8 different nucleosome structures. This observation is consistent with the idea that LH – nucleosome recognition is strongly affected by electrostatic interaction forces. For *G. gallus* gH5, the total number of configuration changes in the first encounter complex clusters for docking to the 8 different nucleosomes is 5 (V80K and K85V) and 4 (K82I and V87K), whereas for *D. melanogaster* gH1, it is 6 (K109V and I104K) and 3 (K102V and K107V). Previously, by using BD and molecular dynamics simulations, it was shown that *G. gallus* gH5 V87 makes hydrophobic contacts with nucleosome thymine methyl groups in the off-dyad binding mode that are enhanced by induced fit and the adoption of a loop-out conformation of the gH5 (137). While the present rigid-body docking results indicate that the V80, K82, K85 and V87 residues of the *G. gallus* gH5 and the corresponding K102, I104, K107 and K109 residues of *D. melanogaster* gH1 are important for nucleosome recognition, it can be anticipated that the mutations will also affect stabilization of the chromatosome complex by induced fit.

Analysis of hydrogen bonds (H-bonds) between the LH and the phosphate backbone of the nucleosomal DNA indicates that WT and mutant *D. melanogaster* gH1 generally make fewer hydrogen bonds in the encounter complexes compared to WT and mutant *G. gallus* gH5. Summing over the 8 different docking simulations and over the 2 encounter complex clusters with the highest populations, WT *D. melanogaster* gH1 makes 27 hydrogen bonds whereas *G. gallus* gH5 makes 35 (Tables 4.8 and 4.9). Interestingly, in WT *D. melanogaster* gH1, the residues making the most hydrogen bonds are K92 and K95 on the α_3 helix that can bind to alternative DNA grooves on the

nucleosome (Table 4.9). On the other hand, in *G. gallus* gH5, most of the hydrogen bonds formed in docking simulations are made by R47 and R94 on the α_2 helix and β sheet, respectively (Table 4.8). These hydrogen-bonding differences indicate that different LH isoforms may have different nucleosome recognition mechanisms.

Table 4.8: The number of hydrogen bonds formed between nucleosomal DNA and wild-type and mutant *G. gallus* gH5 are summed over 8 different *G. gallus* gH5 - nucleosome docking simulations. Occurrences of 6 or more hydrogen bonds to a *G. gallus* gH5 residue are highlighted in bold. Some residues have more than one hydrogen bond in the docked position.

Residues	gH5 WT	V80K	K82I	K85V	V87K
S24	0	0	1	0	0
R37	0	0	0	1	0
R42	3	3	3	2	0
R47	**8**	**9**	0	**10**	3
Q48	2	0	0	2	0
K52	0	**6**	0	0	0
K55	0	1	2	0	0
K59	0	0	2	0	0
K69	2	5	**8**	3	**12**
R73	0	2	2	3	1
R74	2	2	0	1	0
A78	0	1	0	0	0
K82	0	1	1	4	0
K85	2	1	2	3	2
K87	0	0	0	0	1
S90	0	1	0	0	1
R94	**14**	2	0	0	5
K97	2	4	0	2	2
Total	**35**	**38**	**21**	**31**	**27**

The introduction of single point mutations in the LHs also resulted in significant changes in hydrogen bonding with the nucleosome. Summing over the 8 nucleosome structures of the two encounter complex clusters with the highest populations, the residues that make more than six hydrogen bonds with the nucleosome in *G. gallus* gH5 mutants are R47 (9 H-bonds made by V80K mutant and 10 H-bonds made by K85V mutant), K52 (6 H-bonds made by V80K mutant), and K69 (8 H-bonds made by K82I mutant and 12 H-bonds made by V87K mutant) (Table 4.8). For *D. melanogaster* gH1 docking, the corresponding residues are K72 (6 H-bonds made by K107V mutant

and 7 H-bonds made by K109V mutant) and K91 (6 H-bonds made by K102V mutant) (Table 4.9). These results indicate that the hydrogen bonding network of LH – nucleosome interaction is sensitive to point mutations. Remarkably, the I104K mutant of *D. melanogaster* gH1 makes far fewer hydrogen-bonds (18 in 8 simulations) compared to the other mutants and PTMs (Table 4.9). Interestingly, a significant shift in configuration for *D. melanogaster* gH1 (6 of 8 encounter complexes shifted compared to WT) is also observed for this mutant, which suggests that H-bonding is important for the LH – nucleosome configuration.

In certain LH mutant and nucleosome combinations, single point mutations on the LH are able to switch the LH binding mode from *D. melanogaster* gH1 to WT *G. gallus* gH5 or vice versa. For example, in docking of the I104K mutant of *D. melanogaster* gH1 to nucleosome conformations 1 and 6, the representative structures from the largest encounter complexes are similar to the configurations for WT *G. gallus* gH5 docking to the same nucleosome conformations (+1 ↖, +1 +2, +2). Furthermore, docking the K109V mutant of the *D. melanogaster* gH1 to nucleosome conformation 6 yielded similar configurations to WT *G. gallus* gH5 for docking to the same nucleosome conformations (+1 ↖, +1 +2, +2) (Tables 4.5 and 4.6). In addition, docking of the K82I mutant of the *G. gallus* gH5 to nucleosome conformation 6 (-2 ←, -2, no) resulted in similar configurations to WT *D. melanogaster* gH1 (Tables 4.5 and 4.6). These results indicate that single point mutations may switch the chromatosome configurations of different LH isoforms.

Even though the sequences of *D. melanogaster* gH1 and *G. gallus* gH5 share only 49 % sequence identity (Figure 4.1), both WT LH GDs docked in the same off-dyad mode to the most open nucleosome conformation, snapshot 8, as (-1 ↘, -1 -2, -2) (Tables 4.5 and 4.6). This shows that, apart from the amino acid sequence of the LH, L-DNA opening of the nucleosome affects the chromatosome configuration. It suggests that open nucleosome conformations may be able to bind LH proteins off-dyad non-specifically, and that subsequently more specific on- and off-dyad configurations are formed upon LH-induced nucleosome closing.

4.3.3 PTMs of *D. melanogaster* gH1 can modulate LH – nucleosome binding

In addition to single point mutations, docking results revealed that PTMs can also switch the configuration of *D. melanogaster* gH1 – nucleosome binding. Four known PTMs, two lysine dimethylations and two serine phosphorylations, were investigated. Dimethylation interferes with salt-link formation and phosphorylation introduces negative charge. The number of significant shifts in chromatosome configurations (Figure 4.8) upon introducing PTMs is 8 (S67phosphorylation), 3 (K72dimethylation), and 2 (S66phosphorylation, K58dimethylation). In WT *D. melanogaster* gH1, K58 is on the α_1 helix and has very limited contacts with nucleosomal DNA in the docked encounter

complexes (Table 4.9). Thus, it is not surprising that dimethylation of K58 has a modest effect on nucleosome binding. S66 points towards the LH core and not to the surface as S67 does, which could explain the limited shifts in configuration observed upon S66phosphorylation (Figure 4.2C). On the other hand, S67 and K72 are both on the α_2 helix (Figure 4.2C) and introduction of these PTMs on the interaction surface of LH – nucleosome affects the LH binding pose and the number of H-bonds made by the neighboring residues (Table 4.9).

Apart from S67phosphorylation, all PTMs resulted in an on-dyad binding mode to the most closed nucleosome conformation, snapshot 4, as observed for WT *D. melanogaster* gH1. Overall, though, for a total of 40 docking encounter complexes for WT gH1 and gH1 with PTMs, 31 resulted in off-dyad binding to L-DNA and only 3 in off-dyad binding to L-DNA2.

For all four PTMs, the number and nature of the hydrogen bonds with the nucleosome compared to WT *D. melanogaster* gH1 is affected for the 2 encounter complexes with the highest populations. In WT *D. melanogaster* gH1, the majority of the hydrogen bonds with the nucleosome are made by K92 and K95 (6 H-bonds each). For gH1 with PTMs, the most H-bonds are made by R63 (9 H-bonds, K72dimethylation, 6 H-bonds, S66phosphorylation), K91 (8 H-bonds, S67phosphorylation; 6 H-bonds, S66phosphorylation), and K107 (6 H-bonds, K72dimethylation) (Table 4.9). Unlike the point mutants studied, the gH1 variants with PTMs bind differently to the most open nucleosome conformation, snapshot 8. This may be due to their greater size which reduces steric accessibility to the N-DNA and results in encounter complexes further away from the dyad axis. The high variation in hydrogen bonding upon introducing PTMs suggests that each post-translationally modified LH could have unique nucleosome interaction features and thus may have a distinct regulatory effect on chromatin compaction and gene regulation.

Table 4.9: The number of hydrogen bonds formed between nucleosomal DNA and wild-type, mutant and post-translationally modified *D. melanogaster* gH1 summed over 8 different *D. melanogaster* gH1 - nucleosome docking simulations. Occurrences of 6 or more hydrogen bonds to a *D. melanogaster gH1* residue are highlighted in bold. Some residues have more than one hydrogen bond in the docked position.

Residues	WT gH1	K102V	I104K	K107V	K109V	K58Dme	S66Phos	S67Phos	K72Dme
K61	0	1	0	0	0	0	0	0	0
R63	3	3	4	5	3	4	**6**	5	**9**
S66	0	0	1	0	0	0	0	0	1
S67	0	1	0	0	0	0	1	4	0
L68	0	0	0	2	0	0	0	0	0
K72	1	3	3	**6**	**7**	3	4	0	0
K73	3	3	0	2	0	2	1	5	0
Q84	0	0	1	0	0	0	0	0	0
K85	0	0	0	1	1	1	1	1	2
K90	0	0	0	0	0	0	1	0	0
K91	4	**6**	0	1	4	5	**6**	**8**	4
K92	**6**	3	1	4	2	2	1	1	1
K95	**6**	3	0	4	3	4	4	3	4
I104	0	0	1	0	0	0	0	0	0
Q105	0	0	0	0	0	0	1	0	0
T106	0	0	1	0	0	0	0	0	1
K107	3	2	4	0	3	**6**	3	5	**6**
K109	1	0	1	1	0	1	1	1	2
A111	0	0	0	1	0	0	0	0	0
G113	0	0	0	1	0	0	0	0	0
K116	0	0	1	0	0	0	0	0	0
Total	**27**	**25**	**18**	**28**	**23**	**28**	**30**	**30**	**30**

4.4 Concluding discussion

By BD docking of refined structures of LH GDs to nucleosome structures, the crystal structures of chromatosomes determined by Zhou et al. (7) and Bednar et al. (9) were recapitulated. These results confirm that BD rigid body docking is a valid tool for studying LH – nucleosome binding configurations and can be used without prior knowledge of the structural constraints of the complex structure. Previous MD simulations of the chromatosome structure in Chapter 3 suggested that both conformational selection and induced fit facilitate the LH – nucleosome complex structure (137). Thus, it should be born in mind that a complete understanding of chromatosome complexation by LH mutants will require further MD simulations to investigate the stability of the fully bound mutant complexes formed from the diffusional encounter complexes generated by BD docking.

The results of BD docking simulations indicate that the chromatosome configuration is sensitive to single point mutations and PTMs in the GD of LHs. It was shown that mutations changing

the charge on *G. gallus* gH5 residues V80, K82, K85 and V87 and on *D. melanogaster* gH1 K102, I104, K107 and K109 around the LH β turn significantly affect the LH configuration. Considering the diversity of species of the LHs used in recent experimental studies of LH - nucleosome complexes, results indicate that a systematic comparison of chromatosome configurations for different LH and nucleosome sequences and single point mutations is necessary to understand the distribution of the chromatosome structural ensemble and its effect on function. Moreover, experiments to investigate the structural ensemble in solution, such as hydroxyl radical footprinting or NMR, are important to complement crystallographic data. In higher eukaryotes, having a chromatosome structural ensemble could facilitate the ability of one LH isoform to substitute for other LH isoforms, for example as indicated by recent experimental studies showing that a single LH isoform knock out is not lethal in mice (55).

Currently, there is a significant interest in determining the phenotypic effects of core histone tail PTMs. Here, it was shown that LH PTMs may alter the chromatosome structural ensemble, which may impact higher order chromatin structure and possibly gene expression profiles. It was found that S67phosphorylation and K72dimethylation cause the most significant shifts in chromatosome configuration whereas S66phosphorylation and K58dimethylation have modest effects. Applying single point mutations like K72R to mimic dimethylation (185) and S67E for phosphorylation (186) of *D. melanogaster* gH1 could be a promising experimental approach to understand the phenotypic effects of these PTMs.

The results of our BD docking simulations indicate that the chromatosome configuration is sensitive to single point mutations and PTMs in the GD of LHs. I show that mutations changing the charge on *G. gallus* gH5 residues V80, K82, K85 and V87 and on *D. melanogaster* gH1 residues K102, I104, K107 and K109 around the LH β turn significantly affect the LH configuration. The results show that both electrostatic and steric effects of the mutations and PTMs significantly influence the LH – nucleosome configuration. The computed LH GD – nucleosome interaction energies in the diffusional encounter complexes vary within a few kT in the different configurations. Thus, other mutations and PTMs on the nucleosome binding faces of the LH GDs can be expected to affect LH – nucleosome configuration to varying extents.

In conclusion, by applying BD docking simulations, it was found that the chromatosome structural ensemble is sensitive to specific LH mutations and PTMs, which may have implications for the effects of LH binding on chromatin structure and function.

5

Computation of FRAP recovery times for linker histone – chromatin binding on the basis of Brownian dynamics simulations

5.1 Purpose of research

Proteins make various interactions and they diffuse through different parts of the cell to perform their molecular functions. For the understanding of *in vivo* biological processes, it is crucial to determine the binding and unbinding features of proteins inside the cell. The development of photobleaching methodologies about 40 years ago enabled researchers to measure diffusion constants of biomolecules in membranes (187, 188). Later, following the discovery of green fluorescent protein (GFP) and advances in the fluorescence recovery after photobleaching (FRAP) methodology (189), it became possible to understand the binding and unbinding of GFP-attached proteins inside cells. Both qualitative and quantitative analyses of the kinetics of various biological systems have been reported with these experimental approaches (190). Among these studies, there has been significant interest in understanding the binding of the linker histone (LH) protein to chromatin (16, 191, 192).

Previously, Lele et al. used FRAP to study LH - chromatin binding and found that LH diffusion through the nucleus does not play a significant role in the bleach recovery rate of the LH but it is instead determined by binding and unbinding events of the LH (192). By measuring the ratio of the bound and free concentrations of the LH at steady state, and using mathematical modeling of reaction-dominant conditions (bound / free LH exchange is slower than free LH diffusion), the authors derived a dissociation rate constant (k_{off}) = 0.0131 s^{-1} and an association rate constant (k_{on}) = 0.14 s^{-1} for GFP-H1.1 binding to chromatin in NIH3T3 cells. However, concurrently, Beaudouin et al.

showed that for five H1-PAGFP isoforms binding to chromatin, intermolecular diffusion limits the very transient interactions and it can slow down fluorescence redistributions in NRK cells (193). The authors reported an upper limit for the equilibrium dissociation constant K_d of 70 nM. Subsequently, Stasevich et al. used different domains of LH for FRAP analysis and they found that there are slow and fast binding states (191). In the extensive study of Brown et al. (16), 41 different single and double residue mutants were generated in the globular domain of mouse H1.0 (gH1.0) and GFP protein was attached on the C-terminus of the LH. FRAP experiments were conducted on H1.0 – GFP injected mouse BALB/c 3T3 cell lines for each gH1.0 mutant. The authors ontained recovery half time (t_{50}) values for mutants that differ from wild-type LH. Then, they used this information to identify LH residues in the LH - nucleosome binding interface and thereby generate a model of the LH – nucleosome complex (16). Previously, it has been shown that macromolecular crowding affects enzyme kinetics, association of proteins and protein - DNA interactions (194, 195). Thus, the crowding effect inside the nucleus should be accounted for in *in vitro* and *in silico* studies of the isolated molecular interactions for proper comparison with *in vivo* studies.

Brownian dynamics (BD) simulation is a valuable tool to study the diffusional binding kinetics of biomolecules and has been previously used to predict k_{on} values and the structures of diffusional encounter complexes of macromolecular complexes (26, 137, 196). Here, by performing BD simulations of LH - nucleosome binding, k_{on} and k_{off} values were calculated for nucleosome binding by wild-type (WT) and mutant gH1.0. After using a scaling factor to compensate for the crowded cellular environment, reaction dominant conditions were used to compute FRAP recovery plots and t_{50} values for LH - chromatin binding.

5.2 Methods and theory

k_{on} and k_{off} values for the binding of different mouse gH1.0 mutants to a model nucleosome were computed by using the following BD simulation protocol.

5.2.1 Structure preparation

In order to compare to the FRAP experiments reported by Brown et al. (16) mentioned above, the structure of the mouse H1.0 globular domain (residues 24-97) was modeled by using the Swiss-homology-modeling web server (197) with the *Gallus gallus* gH5 structure (PDB id: 1HST, Chain B, 2.6 Å resolution) (5) as a template. As 3 of the 41 mutants reported by Brown et al. (16) were outside the modeled globular domain structure, the remaining 38 point mutations in the globular domain were generated by using the PyMOL molecular modeling software (180) and selecting the first optimized rotamer of the mutated side chains suggested by the program.

The nucleosome structure used was the same as modeled by Pachov et al. (73), which showed a good agreement with the data from Brown et al. (16) for the interaction interface of the LH

and the nucleosome. The nucleosome structure consisted of the nucleosome core particle taken from PDB id: 1KX5, (1.9 Å resolution) (43) and the two 10 bp linker DNA (L-DNA) extensions taken from PDB id: 1ZBB, (9 Å resolution) (66).

5.2.2 BD simulations to compute association rate constants and binding free energies

The association rate constant of each mutant mouse gH1.0 binding to the nucleosome was calculated by using the SDA 7 software (130) with the following parameters: The PDB2PQR 2.0.0 web-server was used to add polar hydrogen atoms to the structures (198). The AMBER99 force field (184) was used to assign the partial atomic charges and atomic radii. The molecular electrostatic potentials were calculated by using the APBS 1.4 software to solve the non-linear Poisson–Boltzmann equation with a 1 Å grid spacing (156). The temperature was 300 K, the solvent and solute dielectric constants were 78 and 1, respectively, and the ionic strength was 150 mM. The van der Waals surface was used to define the dielectric boundary between solute and solvent.

BD simulations (130) were run with electrostatic interaction and electrostatic desolvation terms computed between the nucleosome and the gH1.0. In order to prevent overlap of the molecules, an excluded volume criterion was applied. The ECM program was used to assign effective charges to charged residues on the protein and to P atoms on the DNA (133).

The BD trajectories were started with random orientations and positions on a sphere at solute center-to-center distance of ~200 Å and stopped at a center-to-center distance of ~400 Å. For each system, 5000 BD trajectories were generated. For the k_{on} calculation, a criterion of 4.5 Å was used to identify LH – nucleosome contacts by aligning each LH mutant structure to off-dyad position by using the LH – nucleosome complex structure reported by Pachov et al. (73) (Figure 5.1) as a reference. From 10 contact pairs defined, when 2 pairs of contact atoms were within 6.5 Å and independent as defined with a distance criterion of 6 Å, the reaction criteria were considered satisfied and the probability of satisfying the reaction criteria (β) was used to derive the k_{on} values (See Chapter 2.3.3). Bootstrap analysis was performed to calculate the standard deviation of k_{on} values.

In docking, the same contact pairs as for k_{on} calculations were used. When 1 pair of contact atoms were within 9.5 Å, the reaction conditions were satisfied and the coordinates and interaction energies of the complex were recorded. If the RMSD to the previously recorded complexes was > 1 Å and the interaction energy was within the 5000 most favorable energy complexes recorded. In the case of a complex with RMSD < 1 Å to a previously recorded complex but with a lower energy, the new complex was recorded as a substitute of the previously recorded complex. In each simulation, the average Gibbs binding free energy (ΔG) of forming a diffusional encounter complex was

calculated by averaging the binding energies of the 5000 complexes from BD outputs mentioned above.

By using Eq. (5.1):

$$k_{off} = k_{on}\, e^{\frac{\Delta G}{RT}} \tag{5.1}$$

it is possible to obtain k_{off} for the LH dissociation from the LH - nucleosome diffusional encounter complex. For 38 different LH mutants, k_{off} values were calculated.

5.2.3 Data analysis

It should be noted that, Lele et al. (192) reported their k_{on} results for LH – chromatin binding in s^{-1} units. In SDA k_{on} values are calculated for LH – nucleosome binding and results obtained in $M^{-1}s^{-1}$ units. These measured and computed k_{on} values with s^{-1} and $M^{-1}s^{-1}$ units can be compared under the assumption of a scaling factor between moles and the number of nucleosomes on the chromatin.

For wild-type LH, the calculated k_{off} value was ~ $2.8\text{x}10^6$ times higher than the experimentally determined k_{off} value for the *in vivo* LH - chromatin complex reported by Lele et al. (192). This large difference in magnitude is likely due to crowding and confinement effects on transfer from *in vitro* dilute conditions to cellular conditions as well as non-diffusive effects due to the difference between the diffusional encounter complex and the bound complex. Next, it was assumed that these effects can be described by a single constant parameter applicable to all LH mutants corresponding to the scaling factor for the wild-type LH: $2.8\text{x}10^6$. Then, to be able to estimate a value corresponding to the experimental k_{off} for each mutant, the computed k_{off} values were divided by this scaling factor for the wild-type protein.

When the reaction binding rate (number of molecules that bind to the binding partner per second) is much lower than the diffusion rate (reaction dominant conditions), the FRAP recovery curve is determined by,

$$F(t) = 1 - e^{-k_{off}\, t} \tag{5.2}$$

For the 38 different LH mutants (16), FRAP t_{50} values were computed by solving the Eq. (5.3):

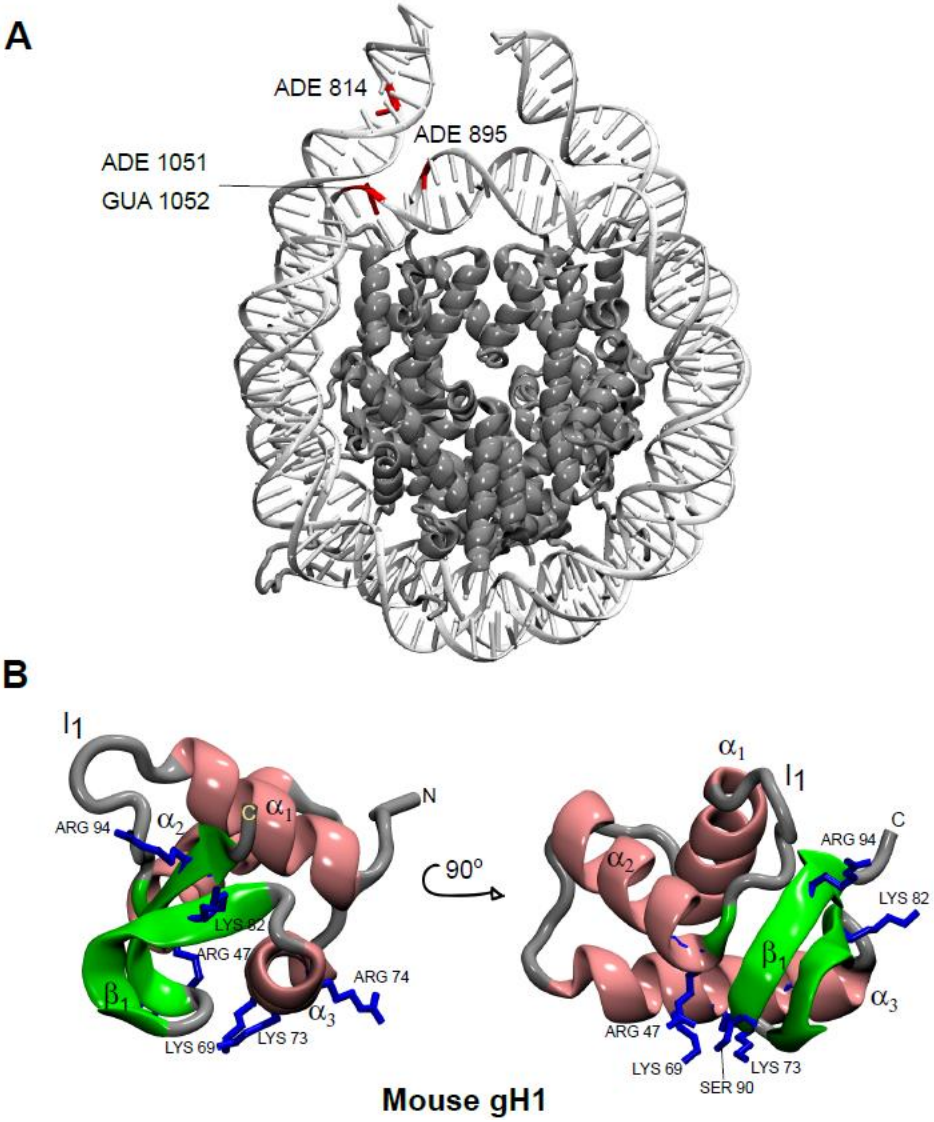
$$0.5 = 1 - e^{-k_{off}\, t_{50}} \tag{5.3}$$

These computed t_{50} were compared with the measured t_{50} values reported by Brown et al. (16).

5.3 Results and discussion

5.3.1 BD-based FRAP t_{50} predictions of LH – nucleosome binding show trends observed in experimental data

Computed FRAP t_{50} recovery values are compared with the experimental t_{50} values for LH mutant-chromatin binding reported by Brown et al. (16) in Figure 5.2. The results for the D65K mutant (Figure 5.1C) were only given in Table 5.1 as the computed t_{50} value is ~40X the measured value.



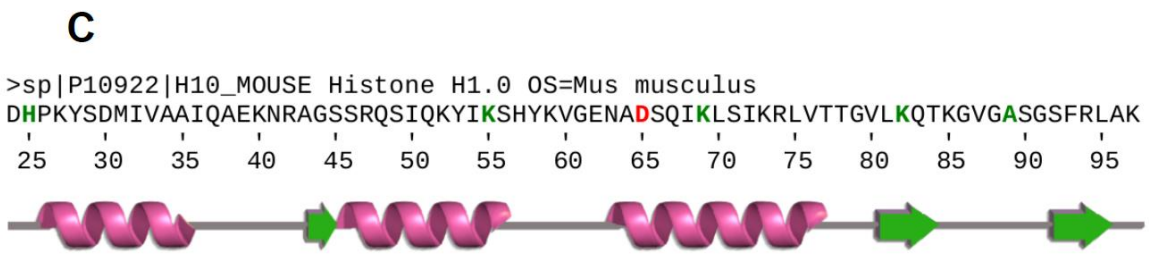


Figure 5.1: Nucleosome and LH structures used in BD simulations. **A-** Structure of the nucleosome used in BD simulations. The DNA (147 bp nucleosomal DNA plus two 10 bp linker DNA arms) is shown in white cartoon representation and the core histones in gray cartoon representation. The nucleotides used to define reaction criteria contacts in association rate constant calculations for the off-dyad gH1.0 binding mode are indicated in red. **B-** Structure of mouse gH1.0 used in BD simulations. The globular domain of the LH is shown in cartoon representation and colored according to secondary structure: α helices in pink, β sheets in green and unstructured regions in gray. Residues used to define reaction criteria contacts in association rate constant calculations are indicated in blue stick representation. **C-** The sequence and the secondary structure of the globular domain of the LH used in BD simulations. The five residues exhibiting outlier behavior in the comparison of computed and experimental values of t_{50} (see Figure 5.2) are highlighted in green. D65K mutant (red) is neglected in analyses as the computed value was ~40X the measured value (see Table 5.1).

For the remaining 37 mutants measured experimental t_{50} values are in a range of 2 – 76 s whereas computational values vary between 1.5 – 87.4 s. The overall BD prediction of FRAP t_{50} recovery times compared to WT are in good agreement with the experiment except for 7 outliers (deviating more than 3σ from the measured t_{50} value) out of the 37 mutants (Figure 5.2). These outliers are H25G (12.0 vs 50.5, experimental t_{50} in s vs BD based t_{50} prediction in s), H25E (10.0 vs 38.2), K55E (3.0 vs 27.5), K55D (5.0 vs 34.2), K69A (4.0 vs 14.1), K82V (63.0 vs 20.2), A89D (3.0 vs 18.0). Among these mutants, K82V is underestimated whereas the rest of the outliers are overestimated with respect to the experimental t_{50} recovery times (Figure 5.2).

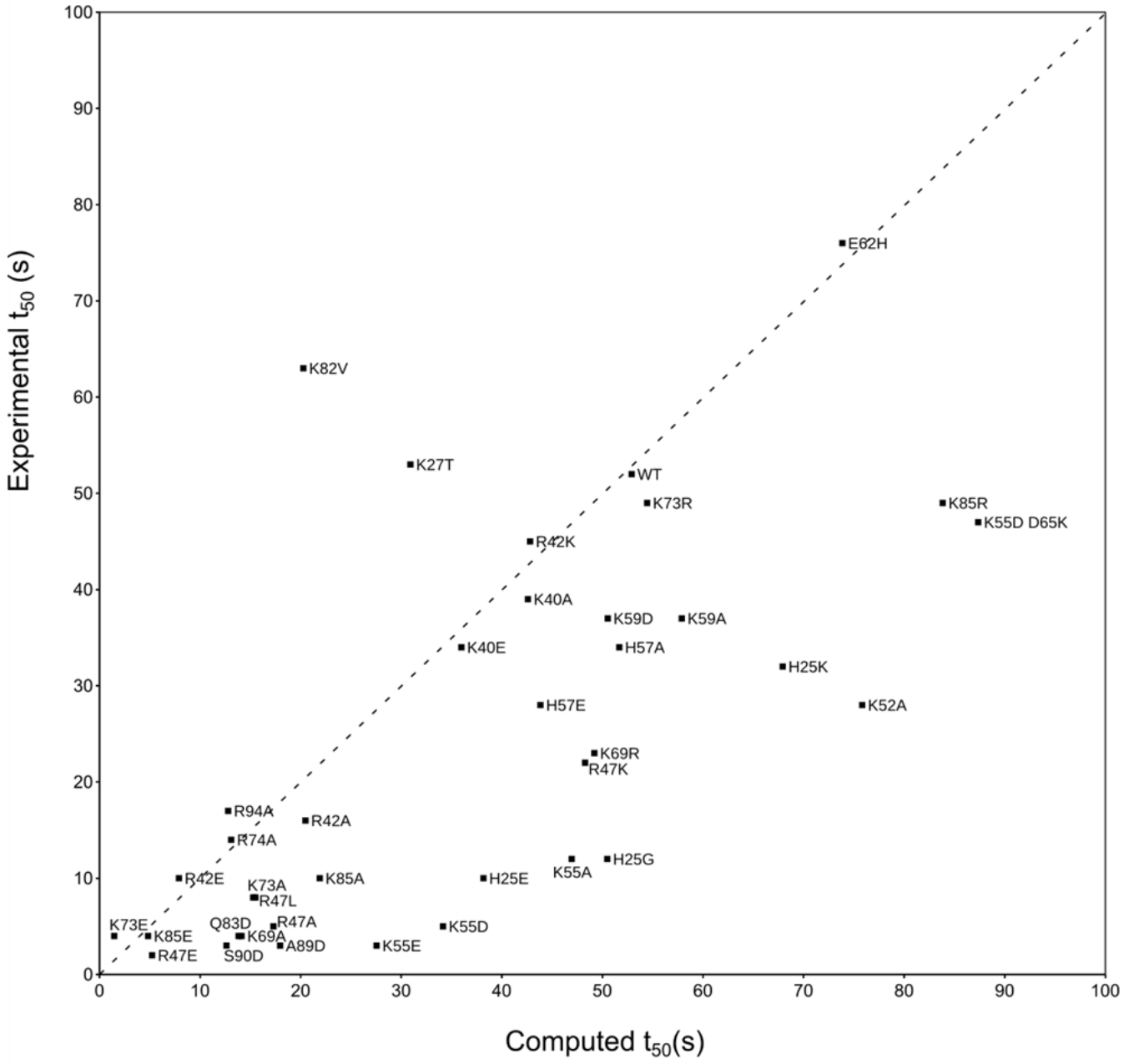


Figure 5.2: Comparison of the BD-based computed FRAP t_{50} recovery times with experiment. The experimental t_{50} recovery times from Brown et al. (16) are plotted against computed FRAP t_{50} recovery times for the WT and 37 single and double point mutant gH1.0 structures. See Table 5.1 for standard deviations (σ).

As the BD simulations are conducted with static protein structures, mutations of the outlier residues could cause higher overall structural deviation from the WT structure compared to mutations of other residues. Interestingly, the positions of these mutations are in close proximity to the α_1, α_2 and α_3 helices or β_1 turn of the LH whose mutations could impact the overall stability of the LH (Figure 5.1C). Remarkably, only K27T, R42E, K73E, K82V and R94A t_{50} recovery values are underestimated in calculations compared to their experimental FRAP t_{50} recovery times whereas the majority of the t_{50} recovery values of the remaining LH mutants are overestimated. Considering that BD calculations were done for LH – nucleosome binding only and crowded nucleus conditions were not applied, such trend can be expected as molecular crowding was previously shown to affect the rate of association (199).

5.3.2 Computed FRAP recovery plot simulations are in good agreement with experimental FRAP recovery plots for mutations to same residue and mutations of the same residue

Previously, Brown at al. (16) compared the FRAP t_{50} recovery plots by mutating the positively charged residues of the LH to alanine resulting in K69A, K73A, K85A (N-DNA contacting residues) and K40A, R42A, K52A, R94A (L-DNA contacting residues), respectively. By using my BD based protocol, the computed t_{50} recovery values of these mutants are plotted (See Methods and Figures 5.3A and B). Remarkably, in simulation generated FRAP t_{50} recovery plots it is possible to reproduce t_{50} values relative to WT in which K69A < K73A < K85A < WT is observed experimentally (Figure 5.3A and Table 5.1). Similarly, except for the K52A mutant, the correct experimental trend of R94A < R42A < K52A < K40A < WT is also obtained in my simulations (Figure 5.3B and Table 5.1). As the K52 residue is at the center of the α_2 helix, a mutation of this residue could be causing a structural distortion of the core packaging (Figure 5.1C).

Furthermore, Brown et al. (16) compared the FRAP t_{50} recovery plots of the various mutations of the same LH residue for R42A, R42E, R42K and for K73A, K73E, K73K. In both these cases, the BD based plots for the FRAP t_{50} recovery times of the given residues agree with the experimental data (Figures 5.3C and D). In simulations, the overall FRAP t_{50} values are reproduced as, R42E < R42A < R42K < WT and K73E < K73A < K73R < WT. This analysis indicates that BD is a valid computational tool to predict the experimental trends in FRAP t_{50} recovery times for LH - chromatin binding.

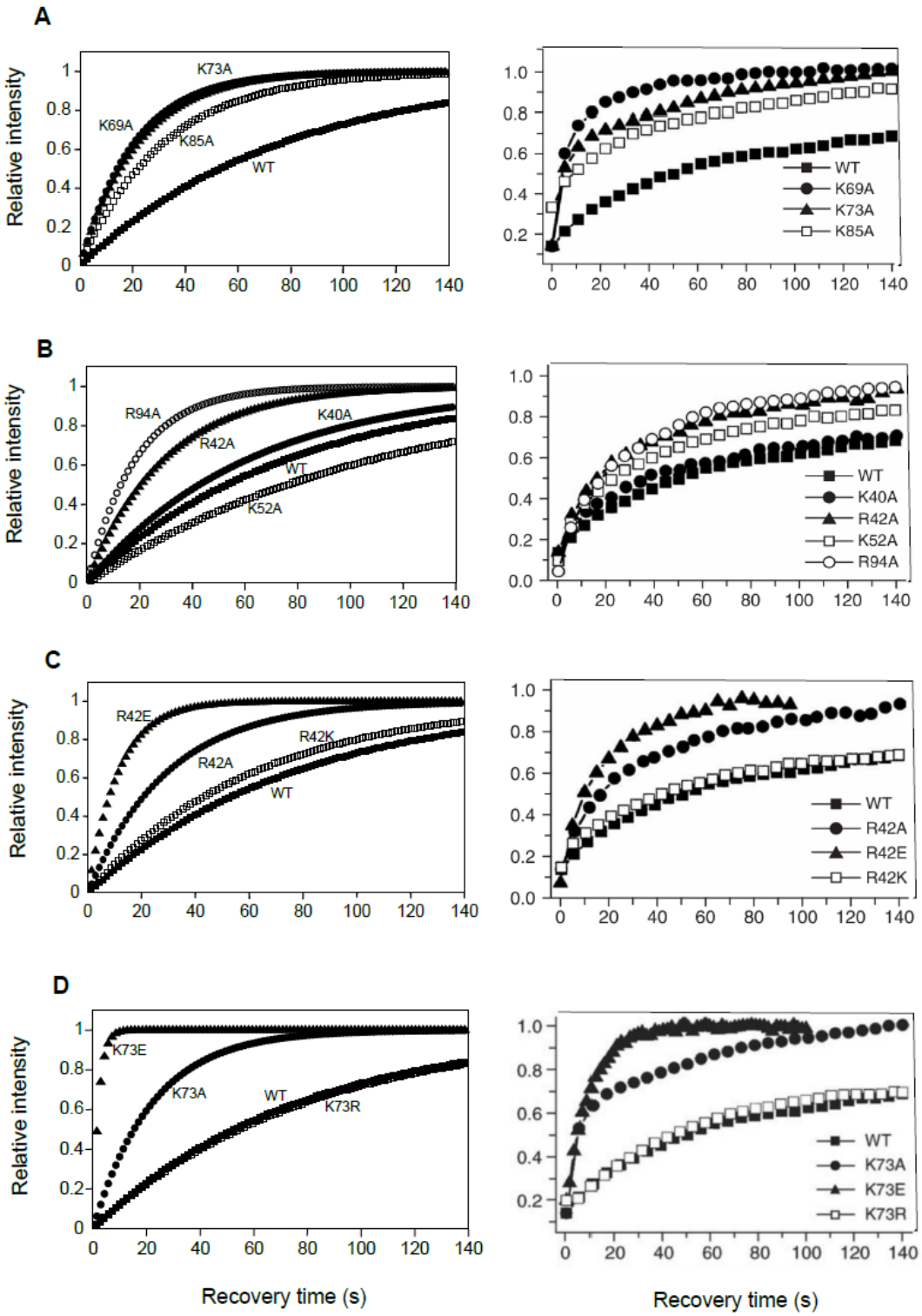


Figure 5.3: Comparison of the BD-based computed FRAP t_{50} recovery time plots with the experimental data from Brown et al. (16). Computed plots are given on the left (this work) and experimental plots are given on the right (re-published from Brown et al. (16) with permission, Figures 1C, 1D, 2A and 2B) for four sets of mutants of mouse H1.0. **A -** K69A, K73A, K85A and WT. **B-** K40A, R42A, K52A, R94A and WT. **C-** R42A, R22E, R42K and WT. **D-** K73A, K73E, K73R and WT. (For the computed relative intensities of each mutant, the same symbol as for the experiments were used.)

5.3.3 BD computed binding free energy correlates with computed association rate constants (k_{on})

Figure 5.4 shows the average binding free energy of 5000 LH - nucleosome encounter complex structures versus the logarithm of the association rate constant of each mutant (Figure 5.4). The results indicate that the average binding free energy correlates with the association rate constant and BD is sensitive enough to differentiate the binding kinetics and binding energy effects of single point mutations. Interestingly, the average binding energy of the LH mutants varies from 7.9 to 13.7 kT and WT, is 13.1 kT. Similarly, the association constant varies around six-fold from 0.0427 to 0.2393 10^{10} x $M^{-1}s^{-1}$ and WT is 0.2040 x 10^{10} x $M^{-1}s^{-1}$. Considering that all BD simulations are done with a single or double point mutant of the LH, almost two-fold variation of the binding energy and six-fold variation of the association constant indicates that single point mutants of the LH could significantly affect its chromatin binding. Furthermore, the high association rate constants overall obtained in simulations show that the binding of the LH to chromatin is facilitated by electrostatic forces. In parallel, the lowest binding affinity and slowest binding was observed for the K73E, K85E, R47E and R42E mutants. As the mutations of positively charged arginine and lysine residues to glutamic acid would reduce the total charge of the LH by 2e-, such change would cause a repulsive effect for LH binding to the negatively charged DNA. Similarly, the opposite effect is observed for the highest binding affinity and fastest binding LH mutants which are K55D D65K, K85R, E62H and H25K, in which the total positive charge is either increased or the charge is redistributed.

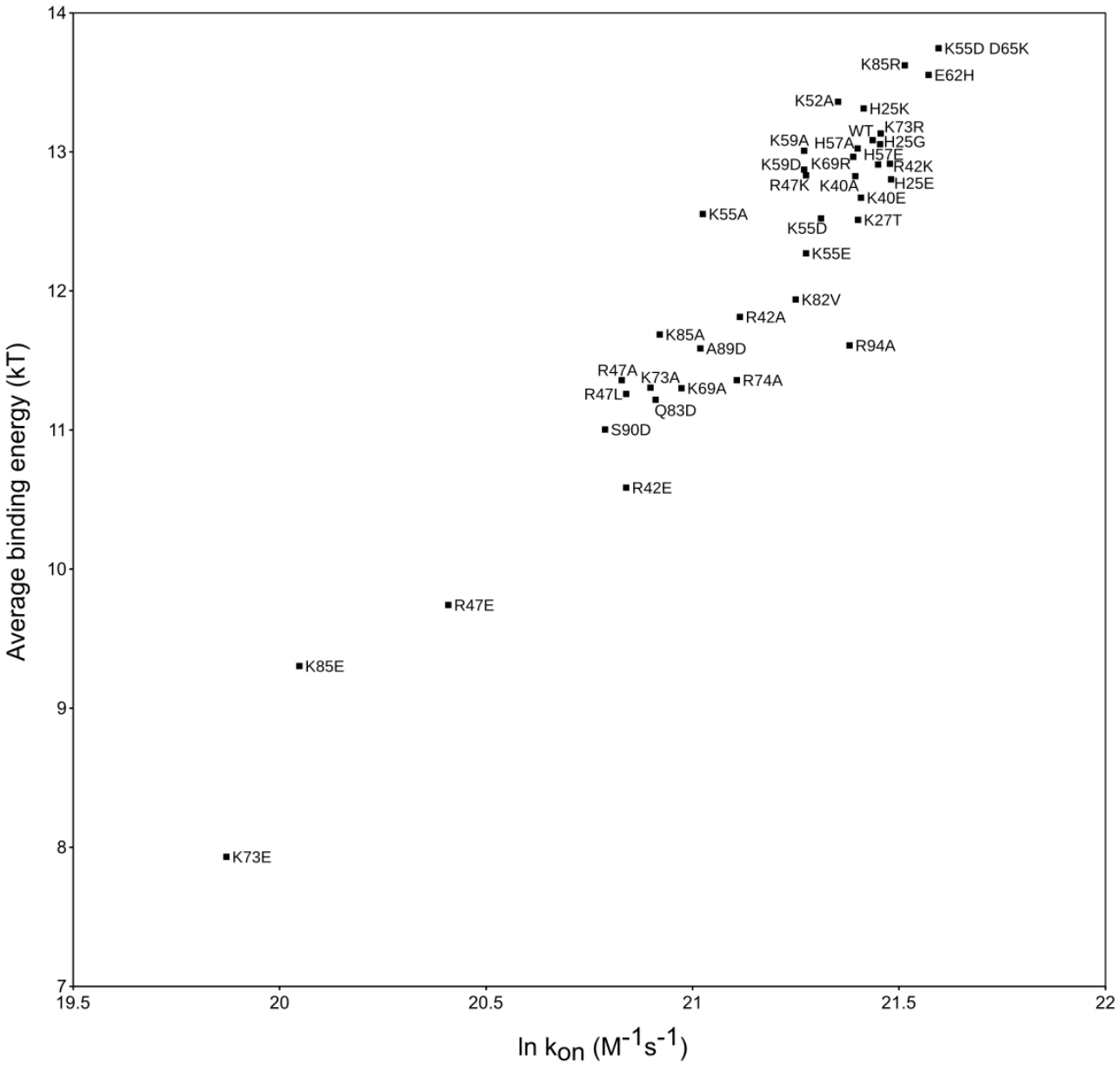


Figure 5.4: BD-based computed LH - nucleosome encounter complex binding free energies correlate (R^2=0.91) with the computed bimolecular association rate constants. The average binding free energy is plotted against ln(k_{on}) for the WT and 37 mutant LH structures. See Table 5.1 for standard deviations.

Table 5.1: BD-based computed values of the Gibbs binding free energy (ΔG), association rate constant (k_{on}), dissociation rate constant (k_{off}) and FRAP t_{50} recovery values related to formation of a diffusional encounter complex between the nucleosome and the mouse gH1.0 are given with FRAP t_{50} values measured by Brown et al. (16).

Protein	ΔG (kT)	k_{on} (10^{10} M^{-1} s^{-1})	scaled k_{off} (s^{-1})	computed t_{50} (s)	reference t_{50} (s)
WT	13.085 ± 0.145	0.2040 ± 0.0103	0.0131 ± 0.0020	52.9 ± 8.1	52 ± 3.2
H25G	13.056 ± 0.143	0.2077 ± 0.0159	0.0137 ± 0.0220	50.5 ± 8.2	12 ± 0.8
H25E	12.803 ± 0.130	0.2133 ± 0.0142	0.0182 ± 0.0027	38.2 ± 5.6	10 ± 1.2
H25K	13.313± 0.189	0.1996 ± 0.0143	0.0102 ± 0.0020	67.9 ± 13.7	32 ± 4.3
K27T	12.512 ± 0.123	0.1969 ± 0.0143	0.0224 ± 0.0032	30.9 ± 4.4	53 ± 4.2
K40A	12.826 ± 0.140	0.1956 ± 0.0139	0.0163 ± 0.0026	42.6 ± 6.7	39 ± 2.8
K40E	12.671 ± 0.141	0.1983 ± 0.0145	0.0193 ± 0.0031	36.0 ± 5.7	34 ± 2.5
R42A	11.814 ± 0.246	0.1479 ± 0.0145	0.0339 ± 0.0090	20.5 ± 5.4	16 ± 1.5
R42E	10.586 ± 0.252	0.1123 ± 0.0111	0.0878 ± 0.0238	7.9 ± 2.1	10 ± 1.1
R42K	12.915 ± 0.146	0.2127 ± 0.0149	0.0162 ± 0.0023	42.8 ± 6.1	45 ± 3.3
R47A	11.359 ± 0.258	0.1111 ± 0.0113	0.0401 ± 0.0023	17.3 ± 4.8	5 ± 0.2
R47E	9.742 ± 0.378	0.0730 ± 0.0087	0.1327 ± 0.0526	5.2 ± 2.1	2 ± 0.3
R47K	12.915 ± 0.146	0.1736 ± 0.0138	0.0144 ± 0.0024	48.3 ± 8.0	22 ± 2.2
R47L	11.260 ± 0.258	0.1123 ± 0.0102	0.0447 ± 0.0122	15.5 ± 4.2	8 ± 0.5
K52A	13.361 ± 0.196	0.1876 ± 0.0141	0.0091 ± 0.0019	75.8 ± 15.9	28 ± 1.6
K55A	12.554 ± 0.217	0.1352 ± 0.0139	0.0148 ± 0.0035	46.9 ± 11.3	12 ± 1.2
K55E	12.527 ± 0.213	0.1736 ± 0.0132	0.0252 ± 0.0057	27.5 ± 6.2	3 ± 0.2
K55D	12.522 ± 0.192	0.1800 ± 0.0132	0.0203 ± 0.0042	34.2 ± 7.0	5 ± 0.2
D65K	15.268 ± 0.116	0.3074 ± 0.0174	0.0022 ± 0.0030	311.6 ± 40.2	8 ± 0.3
K55D D65K	13.746 ± 0.114	0.2393 ± 0.0156	0.0079 ± 0.0010	87.4 ± 11.5	47 ± 3.6
H57A	13.025 ± 0.150	0.1967 ± 0.0139	0.0134 ± 0.0022	51.7 ± 8.6	34 ± 2.8
H57E	12.910 ± 0.156	0.2067 ± 0.0153	0.0158 ± 0.0027	43.8 ± 7.6	28 ± 2.6
K59A	13.009 ± 0.147	0.1728 ± 0.0141	0.0120 ± 0.0020	57.9 ± 9.7	37 ± 4.5
K59D	12.873 ± 0.139	0.1728 ± 0.0134	0.0137 ± 0.0022	50.5 ± 8.0	37 ± 3.4
E62H	13.554 ± 0.180	0.2336 ± 0.0141	0.0094 ± 0.0018	73.9 ± 14.0	76 ± 5.7
K69A	11.301 ± 0.171	0.1284 ± 0.0113	0.0491 ± 0.0094	14.1 ± 2.7	4 ± 0.3
K69R	12.965 ± 0.150	0.1946 ± 0.0142	0.0141 ± 0.0024	49.2 ± 8.2	23 ± 3.2
K73A	11.305 ± 0.187	0.1191 ± 0.0116	0.0454 ± 0.0096	15.3 ± 3.2	8 ± 0.4
K73E	7.931 ± 0.503	0.0427 ± 0.0071	0.4742 ± 0.2512	1.5 ± 0.8	4 ± 0.3
K73R	13.133 ± 0.171	0.2080 ± 0.0143	0.0127 ± 0.0023	54.4 ± 10.0	49 ± 3.9
R74A	11.359 ± 0.258	0.1468 ± 0.0142	0.0530 ± 0.0146	13.1 ± 3.6	14 ± 1.8
K82V	11.939 ± 0.173	0.1693 ± 0.0139	0.0342 ± 0.0065	20.3 ± 3.9	63 ± 4.6
Q83D	11.218 ± 0.201	0.1206 ± 0.0115	0.0501 ± 0.0111	13.8 ± 3.1	4 ± 0.3
K85A	11.687 ± 0.230	0.1218 ± 0.0117	0.0317 ± 0.0079	21.9 ± 5.5	10 ± 1.1
K85E	9.303 ± 0.357	0.0509 ± 0.0077	0.1435 ± 0.0556	4.8 ± 1.9	4 ± 0.3
K85R	13.623 ± 0.155	0.2205 ± 0.0145	0.0083 ± 0.0014	83.8 ± 14.1	49 ± 4.2
A89D	11.587± 0.176	0.1344 ± 0.0122	0.0386 ± 0.0076	18.0 ± 3.6	3 ± 0.2
S90D	11.004 ± 0.298	0.1067 ± 0.0102	0.0549 ± 0.0171	12.6 ± 4.0	3 ± 0.2
R94A	11.609 ± 0.154	0.1929 ± 0.0138	0.0542 ± 0.0092	12.8 ± 2.2	17 ± 1.4

5.4. Concluding discussion

Previously, BD simulations have been used for molecular docking and binding association rate (k_{on}) calculations in various systems (137, 200). In this project, for the first time BD simulations were applied for the prediction of the experimental FRAP t_{50} recovery times. Results provide similar trends to the experimental data of Brown et al. (16) and indicate that BD is a valid tool to compare the *in vivo* chromatin binding kinetics of the LH mutants. Considering that BD simulations in this study are conducted with static LH mutant structures, observed mutant outliers could potentially arise from conformational rearrangements of the LH structures upon introduction of single point mutations.

It should be noted that for the LH chromatin binding in my calculations, the WT k_{off} value was calibrated with the result from Lele et al. (192) which is not the same data set that was used for the comparison of the mutant FRAP t_{50} recovery times to from Brown et al. (16). By applying a scaling factor, the molecular crowding and other effects, which are not directly included in BD simulations were accounted for. For further applications of the methodology in different systems, a similar calibration would be required. Even though it is debated in the literature whether LH – chromatin binding is diffusion limited or not, the results indicate that consideration of intermolecular diffusion explains the main trends due to LH mutation. Thus, explanation of FRAP experiments at the level of intermolecular interactions could be supported by a physical model quantitively.

Consistent with the experiments, results indicate that introduction of positively charged residues or removal of negatively charged residues on LH would increase the binding affinity and increase the FRAP t_{50} values of LH - chromatin binding.

Finally, my analysis showed that by using BD simulations it is possible to obtain predictive binding kinetics parameters for the binding of various mutant structures compared to WT protein. With a proper WT calibration, a BD based approach may be useful for predicting FRAP behavior of other biomolecular interactions.

6

Overall conclusion and outlook

In eukaryotes there is a highly efficient higher order packaging mechanism that makes it possible to fit ~2 meters of DNA inside the human cell nucleus. The dynamics and regulation of this packaging system determine whether the genetic material is used for transcription, replication, recombination or repair. Structural features of DNA packaging have been experimentally studied at various scales, including nucleosome, chromatosome and chromatin. Since the determination of the first crystal structure of the nucleosome by Luger et al. in 1997 (6), various on- and off-dyad linker histone (LH) binding models have been proposed. However, recent structures of LH – nucleosome complexes determined by using NMR, cryo-EM and X-ray crystallography, as well as simulations, suggest that, instead of a single structure, an ensemble of configurations of chromatosomes exists (7–10, 17). In order to obtain a detailed understanding of DNA compaction mechanisms, complementary computational methods need to be applied. In this thesis, state-of-the-art MD, AMD and BD simulations are used to study the molecular details of chromatosome formation. Furthermore, I aimed to understand the effects of single point mutations and PTMs of the LH on nucleosome binding.

In Chapter 1, a detailed analysis of experimentally determined structures of LH – nucleosome complexes is given. Recent experiments showed that, even though researchers have been focusing on determining “the” structure of “the” chromatosome, LH – nucleosome complexes adopt range of structures suggesting that the chromatosome can exist as an ensemble of configurations. In Chapter 2, computational methods that can be used to investigate chromatosome formation, structure and dynamics are introduced.

In Chapter 3, MD, AMD and BD simulations are used to explore the dynamics of a LH – nucleosome complex, which is composed of *G. gallus* gH5 bound to a nucleosome with 167 bp DNA nucleosome with *X. laevis* core histones. MD simulations showed that free gH5 mostly stays in the closed form. Further, AMD analysis indicated that open-to-closed state transitions of LH are possible. Interestingly, MD analysis of the free nucleosome structure showed that the L-DNA arms have an asymmetric flexibility, which is probably determined by DNA sequence differences. By using BD simulations, I showed that it is possible to recapitulate the crystal structure of the LH – nucleosome complex determined by Zhou et al. (7). Additionally, I found that, depending on the L-DNA opening angle, various chromatosome configurations are possible, which is consistent with previous reports of diverse chromatosome structures (4). MD simulations of the gH5 – nucleosome complex indicated a closed to open conformational switch of the LH. This switch is mediated by the hydrophobic interactions of *G. gallus* gH5 V87 and A89 residues with thymine bases of DNA, which is also consistent with the previous experimental results of Cui et al. (71). Interestingly, MD simulations also indicated suppression and a change of directionality of the motion of both L-DNA arms upon LH binding to nucleosome. These findings revealed that both conformational selection and induced fit mechanisms are involved in LH – nucleosome binding. Remarkably, the analysis provides the first atomic detail level understanding of the dynamics of chromatosome formation, which is not possible to obtain by current experimental methods.

In *H. sapiens,* there are ten LH isoforms, whereas *D. melanogaster* has only one. My findings on the residue-specific contacts of LH – nucleosome binding in Chapter 3 raise the question “Are chromatosome configurations dependent on LH isoform sequences, single point mutations and PTMs of the LH globular domain?” In Chapter 4, by using *G. gallus* and *D. melanogaster* LH globular domains and specific residue mutations, the variability of LH GD – nucleosome complex configurations was investigated. I found that different LH isoforms can have different configurations of diffusional encounter complexes. Remarkably, similar to reproduction of the Zhou et al. (7) LH – nucleosome complex structure in Chapter 3, the recently determined Bednar et al. LH – nucleosome complex crystal structure (9) was also re-generated by using BD simulations. Furthermore, I found that specific single point mutations on *G. gallus* gH5 residues V80, K82, K85 and V87 and on *D. melanogaster* gH1 K102, I104, K107 and K109 significantly shift the LH – nucleosome binding mode. Additionally, S67 phosphorylation and K72 dimethylation PTMs on *D. melanogaster* gH1 cause significant shifts in LH – nucleosome complex configuration compared to WT LH. The analysis provides insights that may guide further experimental investigations of the structural and functional effects of LH mutations and PTMs.

In Chapters 3 and 4, I found that single residue mutations impact LH - nucleosome binding. Next, I wanted to explore whether the BD approach can be applied to predict the *in vivo* behavior of LH – chromatin binding. FRAP recovery half-time data for the binding of various LH mutants to

chromatin were previously reported by Brown et al. (16). Using BD simulations calibrated with WT LH – chromatin binding data of Lele et al. (192) and applying reaction dominant conditions, I showed that it is possible to reproduce the overall trend of experimental FRAP recovery half-times for various LH mutants binding to chromatin. In addition to computation of the FRAP recovery half-time data consistent with the experiments of Brown et al., the analysis provides binding free energies, association and dissociation constants for the formation of LH - nucleosome diffusional encounter complexes of various LH mutants. This project can be considered as proof of concept application of BD simulations to predict *in vivo* association and dissociation rate constants of macromolecular complexes and provide an interpretation of FRAP data with a physical model of protein – target binding.

There are various factors that should be considered for the interpretation of the results reported in my thesis. Two sources of uncertainty can be mentioned, the first can be attributed to experimental data and the second to the computational methods and LH – nucleosome complex systems used. One major issue for cross-analysis of different LH – nucleosome complex experimental data is the lack of standardized experimental conditions. In addition to variations in ionic strength and buffers used in LH – nucleosome complex studies, there is no consensus on the choice of DNA sequences and core histones used for nucleosome construction. Up to now, there is no systematic analysis of whether these variations could significantly impact chromatosome formation. Secondly, most of the LH – nucleosome complex structure determination studies are based on cryo-EM and X-ray crystallography, which can only give limited structural ensemble information. Alternatively, using in solution and in cell methods like hydroxyl radical footprinting, FRAP and NMR could give more detailed information on the dynamics of LH - chromatin binding. However, methods like FRAP and hydroxyl radical footprinting can only provide low-resolution structural information due to the low number of observables. Consequently, a major limitation to obtaining an understanding of DNA compaction mechanisms *in vivo* is methodological. However, recent developments of in-cell NMR approaches could in the future make it possible to understand proteins involved in DNA compaction inside of cell nucleus.

In addition to the experimental limitations mentioned above, there are also limitations that arise from the choice of the LH – nucleosome complex system. The first limitation, due to system choice, is the usage of artificial chromatosome systems composed of proteins and DNA sequences coming from various organisms. This may hinder identifying the “real” features of the system. Additionally, even though there are about 10 LH isoforms in *H. sapiens*, a systematic comparison among different isoforms requires further computational effort. Also, the LH – nucleosome complex systems that I used lack core histone and linker histone tails which may cause some deviations in behavior compared to full protein. Furthermore, the nucleosome structures used in my simulations have a specific DNA sequence whose variation could potentially impact chromatosome

configuration. Finally, I used a single nucleosome system which does not incorporate the *in vivo* conditions like specific ions, molecular crowders and linker DNA constraints (as nucleosomes are consecutively connected with linker arms).

The limitations of the computational methods used in my thesis should also be mentioned as a source of uncertainty. Standard MD simulations can usually reach up to microsecond time scales which may not be sufficient to cover all the conformational space of the chromatosome system. In order to overcome this issue in my thesis, AMD simulations were used. Even with the additional conformational space explored by AMD simulations, it may not be sufficient to overcome specific energy barriers. Secondly, the force fields used in MD simulations are still far from being perfect for simulating biomolecular structures. Additionally, BD simulations are conducted in implicit solvent, and the protocol used does not take into account the crowding effects which are present *in vivo*. To solve this issue, in Chapter 5, I used a scaling factor calibrated with WT LH experimental data. Also, in the BD simulations used in my thesis, molecular interactions are calculated for rigid structures which do not fully represent the dynamic nature of biomolecules in solution.

The work presented in this thesis corresponds to a significant step towards understanding isoform specific LH – nucleosome binding and its regulation by PTMs. With the advances of the experimental structural biology methods such as NMR, X-ray crystallography and cryo-EM, we will be able to obtain more information on DNA compacting mechanisms. These include but are not limited to impact of DNA sequence variations, their regulation via acetylation-methylation, LH isoforms and PTMs, higher order chromatin structure and additional proteins involved in chromatin assembly. Concurrent with this progress, it is also realistic to expect improvements in force fields, algorithms, software and hardware that would allow computational methods to provide more detailed information on DNA compaction mechanisms.

Many questions still exist on DNA compaction mechanisms. For example, very recently the Luger lab determined the crystal structure of an archaeal nucleosome which contains two histone types (201), whereas eukaryotic nucleosomes have four core histone proteins. Further studies are needed to understand nucleosome evolution. Another interesting area of progress can be observed in the detailed analysis of the CENP-A nucleosome which functions as an epigenetic marker to recruit kinetochore proteins (202). Parallel with the understanding of protein complexes involved in transcription, DNA regulation and repair, it is possible to apply molecular simulations on various systems to characterize the dynamics of these macromolecular machines in atomic detail.

In summary, in my thesis, by applying a combination of molecular simulation methods, our understanding of the dynamics and kinetics of chromatosome formation is expanded.

Bibliography

1. Whitesides, G.M. 2015. Reinventing Chemistry. Angew. Chemie Int. Ed. 54: 3196–3209.

2. Morange, M. 1998. A history of molecular biology. Harvard University Press.

3. Luger, K., and T.J. Richmond. 1998. DNA binding within the nucleosome core. Curr. Opin. Struct. Biol. 8: 33–40.

4. Fyodorov, D. V, B.-R. Zhou, A.I. Skoultchi, and Y. Bai. 2017. Emerging roles of linker histones in regulating chromatin structure and function. Nat. Rev. Mol. Cell Biol. 19 :192-206

5. Ramakrishnan, V., J.T. Finch, V. Graziano, P.L. Lee, and R.M. Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature. 362: 219–223.

6. Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 389: 251.

7. Zhou, B.-R.R., J. Jiang, H. Feng, R. Ghirlando, T.S.S. Xiao, and Y. Bai. 2015. Structural mechanisms of nucleosome recognition by linker histones. Mol. Cell. 59: 628–638.

8. Zhou, B., H. Feng, and H. Kato. 2013. Structural insights into the histone H1-nucleosome complex. Proc. Natl. Acad. Sci. 111: 1222.

9. Bednar, J., 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. 2017. Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. Mol. Cell. 66: 384–397.

10. Song, F., P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M.M. Xu, P. Zhu, and G. Li. 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science. 344: 376–380.

11. Allan, J., P.G. Hartman, C. Crane-Robinson, and F.X. Aviles. 1980. The structure of histone

H1 and its location in chromatin. Nature. 288: 675–679.

12. An, W., S.H. Leuba, K. van Holde, and J. Zlatanova. 1998. Linker histone protects linker DNA on only one side of the core particle and in a sequence-dependent manner. Proc. Natl. Acad. Sci. 95: 3396–3401.

13. Pruss, D., B. Bartholomew, J. Persinger, J. Hayes, G. Arents, E.N. Moudrianakis, and A.P. Wolffe. 1996. An asymmetric model for the nucleosome: A binding site for linker histones inside the DNA gyres. Science. 274: 614–617.

14. Zhou, Y.-B.B., S.E. Gerchman, V. Ramakrishnan, A. Travers, and S. Muyldermans. 1998. Position and orientation of the globular domain of linker histone H5 on the nucleosome. Nature. 395: 402–405.

15. Syed, S.H., D. Goutte-Gattat, N. Becker, S. Meyer, M.S. Shukla, J.J. Hayes, R. Everaers, D. Angelov, J. Bednar, and S. Dimitrov. 2010. Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. Proc. Natl. Acad. Sci. U. S. A. 107: 9620–5.

16. Brown, D.T., T. Izard, and T. Misteli. 2006. Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. Nat. Struct. Mol. Biol. 13: 250–255.

17. Zhou, B.-R., H. Feng, R. Ghirlando, S. Li, C.D. Schwieters, and Y. Bai. 2016. A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome. J. Mol. Biol. 428: 3948–3959.

18. Frueh, D.P., A.C. Goodrich, S.H. Mishra, and S.R. Nichols. 2013. NMR methods for structural studies of large monomeric and multimeric proteins. Curr. Opin. Struct. Biol. 23: 734–9.

19. Dyson, G. 2012. Turing centenary: The dawn of computing. Nature. 482:459-60

20. Campbell, M., A.J. Hoane, and F. Hsu. 2002. Deep Blue. Artif. Intell. 134: 57–83.

21. Schaller, R.R. 1997. Moore’s law: past, present and future. IEEE Spectr. 34: 52–59.

22. Levitt, M. 2001. The birth of computational structural biology. Nat. Struct. Mol. Biol. 8: 392-393.

23. Lee, E.H., J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten. 2009. Discovery through the computational microscope. Structure. 17: 1295–306.

24. Nowak, W. 2016. Applications of Computational Methods to Simulations of Protein Dynamics. : 1–43.

25. Pierce, L.C.T., R. Salomon-Ferrer, C. Augusto F. De Oliveira, J.A. McCammon, and R.C. Walker. 2012. Routine access to millisecond time scale events with accelerated molecular dynamics. J. Chem. Theory Comput. 8: 2997–3002.

26. Gabdoulline, R.R., and R.C. Wade. 1998. Brownian dynamics simulation of protein-protein diffusional encounter. Methods A Companion to Methods Enzymol. 14: 329–341.

27. Biswas, M., J. Langowski, and T.C. Bishop. 2013. Atomistic simulations of nucleosomes. Wiley Interdiscip. Rev. Comput. Mol. Sci. 3: 378–392.

28. White, C.L., R.K. Suto, and K. Luger. 2001. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J. 20: 5207–18.

29. Bernstein, E., and S.B. Hake. 2006. The nucleosome: a little variation goes a long way. Biochem. Cell Biol. 84: 505–507.

30. Kornberg, R.D. 1974. Chromatin structure: a repeating unit of histones and DNA. Science. 184: 868–871.

31. McAnena, P., J.A.L. Brown, and M.J. Kerin. 2017. Circulating nucleosomes and nucleosome modifications as biomarkers in cancer. Cancers (Basel). 9: 5.

32. Taube, J.H., and M.C. Barton. 2006. Chromatin and regulation of gene expression. In: Gene Expression and Regulation. Springer, New York, NY. pp. 95–109.

33. Grigoryev, S.A., and C.L. Woodcock. 2012. Chromatin organization — The 30nm fiber. Exp. Cell Res. 318: 1448–1455.

34. van Holde, K., and J. Zlatanova. 2007. Chromatin fiber structure: Where is the problem now? Semin. Cell Dev. Biol. 18: 651–658.

35. Finch, J.T., and A. Klug. 1976. Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. U. S. A. 73: 1897–1901.

36. Woodcock, C.L., L.L. Frado, and J.B. Rattner. 1984. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J. Cell Biol. 99: 42–52.

37. Ghirlando, R., and G. Felsenfeld. 2013. Chromatin structure outside and inside the nucleus. Biopolymers. 99: 225–232.

38. Routh, A., S. Sandin, and D. Rhodes. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc. Natl. Acad. Sci. U. S. A. 105: 8872–7.

39. Maze, I., K.-M. Noh, A.A. Soshnev, and C.D. Allis. 2014. Every amino acid matters:

essential contributions of histone variants to mammalian development and disease. Nat. Rev. Genet. 15: 259–271.

40. Roulland, Y., K. Ouararhni, M. Naidenov, L. Ramos, M. Shuaib, S.H. Syed, I.N. Lone, R. Boopathi, E. Fontaine, G. Papai, H. Tachiwana, T. Gautier, D. Skoufias, K. Padmanabhan, J. Bednar, H. Kurumizaka, P. Schultz, D. Angelov, A. Hamiche, and S. Dimitrov. 2016. The flexible ends of CENP-A nucleosome are required for mitotic fidelity. Mol. Cell. 63: 674–685.

41. Bade, D., and S. Erhardt. 2016. Regulation of centromeric chromatin. In: Chromatin Regulation and Dynamics. Elsevier. pp. 303–324.

42. Banks, D.D., and L.M. Gloss. 2004. Folding mechanism of the (H3-H4)2 histone tetramer of the core nucleosome. Protein Sci. A Publ. Protein Soc. 13: 1304–1316.

43. Davey, C.A., D.F. Sargent, K. Luger, A.W. Maeder, and T.J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. 319: 1097–1113.

44. Flanagan, T.W., J.K. Files, K.R. Casano, E.M. George, and D.T. Brown. 2016. Photobleaching studies reveal that a single amino acid polymorphism is responsible for the differential binding affinities of linker histone subtypes H1.1 and H1.5. Biol. Open. 5: 372-380

45. Zhou, B.-R., H. Feng, H. Kato, L. Dai, Y. Yang, Y. Zhou, and Y. Bai. 2013. Structural insights into the histone H1-nucleosome complex. Proc. Natl. Acad. Sci.110: 19390–5.

46. Roque, A., I. Ponte, and P. Suau. 2015. Interplay between histone H1 structure and function. Biochim. Biophys. Acta - Gene Regul. Mech. 1859: 444–454.

47. Puigdomènech, P., M. José, A. Ruiz-Carrillo, and Crane-Robinson. 1983. Isolation of a 167 basepair chromatosome containing a partially digested histone H5. FEBS Lett. 154: 151–155.

48. Kowalski, A., and J. Pałyga. 2016. Modulation of chromatin function through linker histone H1 variants. Biol. Cell. 108: 339–356.

49. Millán-Ariño, L., A. Izquierdo-Bouldstridge, and A. Jordan. 2016. Specificities and genomic distribution of somatic mammalian histone H1 subtypes. Biochim. Biophys. Acta - Gene Regul. Mech. 1859: 510–519.

50. Parseghian, M.H. 2015. What is the role of histone H1 heterogeneity ? A functional model emerges from a 50 year mystery. AIMS Biophys. 2: 724–772.

51. Parseghian, M.H., and B.A. Hamkalo. 2001. A compendium of the histone H1 family of

somatic subtypes: an elusive cast of characters and their characteristics. Biochem. Cell Biol. 79: 289–304.

52. Pan, C., and Y. Fan. 2016. Role of H1 linker histones in mammalian development and stem cell differentiation. Biochim. Biophys. Acta - Gene Regul. Mech. 1859: 496–509.

53. Drabent, B., P. Saftig, C. Bode, and D. Doenecke. 2000. Spermatogenesis proceeds normally in mice without linker histone H1t. Histochem. Cell Biol. 113: 433–442.

54. Fan, Y., A. Sirotkin, R.G. Russell, J. Ayala, and A.I. Skoultchi. 2001. Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. Mol. Cell. Biol. 21: 7933–7943.

55. Fan, Y., T. Nikitina, E.M. Morin-Kensicki, J. Zhao, T.R. Magnuson, C.L. Woodcock, and A.I. Skoultchi. 2003. H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. Mol. Cell. Biol. 23: 4559–4572.

56. Shen, X., and M.A. Gorovsky. 1996. Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell. 86: 475–483.

57. Ramón, A., M.I. Muro-Pastor, C. Scazzocchio, and R. Gonzalez. 2000. Deletion of the unique gene encoding a typical histone H1 has no apparent phenotype in Aspergillus nidulans. Mol. Microbiol. 35: 223–33.

58. Simpson, R.T. 1978. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. Biochemistry. 17: 5524–5531.

59. Widom, J. 1998. Chromatin structure: linking structure to function with histone H1. Curr. Biol. 8: 788–791.

60. Luger, K., T.J. Rechsteiner, and T.J. Richmond. 1999. Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol. Biol. 119: 1–16.

61. Luque, A., R. Collepardo-Guevara, S. Grigoryev, and T. Schlick. 2014. Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. Nucleic Acids Res. 42: 7553–7560.

62. Luque, A., G. Ozer, and T. Schlick. 2016. Correlation among DNA Linker Length, Linker Histone Concentration, and Histone Tails in Chromatin. Biophys. J. 110: 2309–2319.

63. Perišić, O., and T. Schlick. 2017. Dependence of the Linker Histone and Chromatin Condensation on the Nucleosome Environment. J. Phys. Chem. B. 121: 7823–7832.

64. Lowary, P.T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to

histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. 276: 19–42.

65. Tóth, K., V. Böhm, C. Sellmann, M. Danner, J. Hanne, M. Berg, I. Barz, A. Gansen, and J. Langowski. 2013. Histone- and DNA sequence-dependent stability of nucleosomes studied by single-pair FRET. Cytom. Part A. 83: 839–846.

66. Schalch, T., S. Duda, D.F. Sargent, and T.J. Richmond. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature. 436: 138–41.

67. Chua, E.Y.D., D. Vasudevan, G.E. Davey, B. Wu, and C.A. Davey. 2012. The mechanics behind DNA sequence-dependent properties of the nucleosome. Nucleic Acids Res. 40: 6338–6352.

68. Mackay, J.P., M.J. Landsberg, A.E. Whitten, and C.S. Bond. 2017. Whaddaya Know: A Guide to Uncertainty and Subjectivity in Structural Biology. Trends Biochem. Sci. 42: 155–167.

69. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin. 2003. HADDOCK: A protein−protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125: 1731–1737.

70. Bharath, M.M.S., N.R. Chandra, M.R.S. Rao, M.M. Srinivas Bharath, N.R. Chandra, and M.R.S. Rao. 2003. Molecular modeling of the chromatosome particle. Nucleic Acids Res. 31: 4264–4274.

71. Cui, F., and V.B. Zhurkin. 2009. Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. Nucleic Acids Res. 37: 2818–2829.

72. Fan, L., V.A. Roberts, and R.M. Stroud. 2006 Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. Proc. Natl. Acad. Sci. 103(22): 8384–8389

73. Pachov, G. V., R.R. Gabdoulline, and R.C. Wade. 2011. On the structure and dynamics of the complex of the nucleosome and the linker histone. Nucleic Acids Res. 39: 5255–5263.

74. Schübeler, D. 2015. Function and information content of DNA methylation. Nature. 517: 321–326.

75. Allfrey, V.G., R. Faulkner, and A.E. Mirsky. 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. 51: 786–794.

76. Wood, A., and A. Shilatifard. 2004. Posttranslational modifications of histones by

methylation. Adv. Protein Chem. 67: 201–222.

77. Bannister, A.J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. Cell Res. 21: 381–95.

78. Lawrence, M., S. Daujat, and R. Schneider. 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. 32: 42–56.

79. Zhang, T., S. Cooper, and N. Brockdorff. 2015. The interplay of histone modifications - writers that read. EMBO Rep. 16: 1467–81.

80. Bowman, G.D., and M.G. Poirier. 2015. Post-translational modifications of histones that influence nucleosome dynamics. Chem. Rev. 115: 2274–95.

81. Balhorn, R., R. Chalkley, and D. Granner. 1972. Lysine-rich histone phosphorylation. Positive correlation with cell replication. Biochemistry. 11: 1094–1098.

82. Wiśniewski, J.R., A. Zougman, S. Krüger, and M. Mann. 2007. Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. Mol. Cell. Proteomics. 6: 72–87.

83. Garcia, B.A., S.A. Busby, C.M. Barber, J. Shabanowitz, C.D. Allis, and D.F. Hunt. 2004. Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. J. Proteome Res. 3: 1219–1227.

84. Lu, A., A. Zougman, M. Pudełko, M. Bębenek, P. Ziółkowski, M. Mann, J.R. Wiśniewski, M. Bębenek, P. Ziółkowski, M. Mann, and J.R. Wiśniewski. 2009. Mapping of lysine monomethylation of linker histones in human breast and its cancer. J. Proteome Res. 8: 4207–4215.

85. Deterding, L.J., M.K. Bunger, G.C. Banks, K.B. Tomer, and T.K. Archer. 2008. Global changes in and characterization of specific sites of phosphorylation in mouse and human histone H1 isoforms upon CDK inhibitor treatment using mass spectrometry. J. Proteome Res. 7: 2368–2379.

86. Villar-Garea, A., and A. Imhof. 2008. Fine mapping of posttranslational modifications of the linker histone H1 from Drosophila melanogaster. PLoS One. 3: 1553.

87. Bonet-Costa, C., M. Vilaseca, C. Diema, O. Vujatovic, A. Vaquero, N. Omeñaca, L. Castejón, J. Bernués, E. Giralt, and F. Azorín. 2012. Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of Drosophila melanogaster linker histone H1. J. Proteomics. 75: 4124–4138.

88. Sarg, B., R. Lopez, H. Lindner, I. Ponte, P. Suau, and A. Roque. 2015. Identification of novel

post-translational modifications in linker histones from chicken erythrocytes. J. Proteomics. 113: 162–177.

89. Wiśniewski, J.R., A. Zougman, and M. Mann. 2008. N ε -Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res. 36: 570–577.

90. Snijders, A.P.L., S. Pongdam, S.J. Lambert, C.M. Wood, J.P. Baldwin, and M.J. Dickman. 2008. Characterization of post-translational modifications of the linker histones H1 and H5 from chicken erythrocytes using mass spectrometry. J. Proteome Res. 7: 4326–4335.

91. Kim, M.Y.M.Y., T. Zhang, and W.L. Kraus. 2005. Poly(ADP-ribosyl)ation by PARP-1: `PAR-laying' NAD+ into a nuclear signal. Genes Dev. 19: 1951–1967.

92. Poirier, G.G., C. Niedergang, M. Champagne, A. Mazen, and P. Mandel. 1982. Adenosine diphosphate ribosylation of chicken-erythrocyte histones H1, H5 and high-mobility-group proteins by purified calf-thymus poly (adenosinediphosphate-ribose) polymerase. Eur. J. Biochem. 127: 437–442.

93. Jiang, T., X. Zhou, K. Taghizadeh, M. Dong, and P.C. Dedon. 2007. N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc. Natl. Acad. Sci. 104: 60–65.

94. Izzo, A., and R. Schneider. 2016. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. Biochim. Biophys. Acta - Gene Regul. Mech. 1859: 486–495.

95. Horn, P.J., L.M. Carruthers, C. Logie, D.A. Hill, M.J. Solomon, P.A. Wade, A.N. Imbalzano, J.C. Hansen, and C.L. Peterson. 2002. Phosphorylation of linker histones regulates ATP-dependent chromatin remodeling enzymes. Nat. Struct. Biol. 9: 263–267.

96. Chubb, J.E., and S. Rea. 2010. Core and Linker Histone Modifications Involved in the DNA Damage Response. Subcell. Biochem. 50: 17–42.

97. Kim, K., K.W. Jeong, H. Kim, J. Choi, W. Lu, M.R. Stallcup, and W. An. 2012. Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription. Oncogene. 31: 4290–4301.

98. Happel, N., D. Doenecke, K.E. Sekeri-Pataryas, and T.G. Sourlingas. 2008. H1 histone subtype constitution and phosphorylation state of the ageing cell system of human peripheral blood lymphocytes. Exp. Gerontol. 43: 184–199.

99. Harshman, S.W., M.E. Hoover, C. Huang, O.E. Branson, S.B. Chaney, C.M. Cheney, T.J.

Rosol, C.L. Shapiro, V.H. Wysocki, K. Huebner, and M.A. Freitas. 2014. Histone H1 Phosphorylation in Breast Cancer. J. Proteome Res. 13: 2453–2467.

100. Fischle, W., Y. Wang, and C. David Allis. 2003. Binary switches and modification cassettes in histone biology and beyond. Nature. 425: 475–479.

101. Wood, C., A. Snijders, J. Williamson, C. Reynolds, J. Baldwin, and M. Dickman. 2009. Post-translational modifications of the linker histone variants and their association with cell mechanisms. FEBS J. 276: 3685–3697.

102. Boras, B.W., S.P. Hirakis, L.W. Votapka, R.D. Malmstrom, R.E. Amaro, and A.D. McCulloch. 2015. Bridging scales through multiscale modeling: a case study on protein kinase A. Front. Physiol. 6: 250.

103. Skjaerven, L., S.M. Hollup, and N. Reuter. 2009. Normal mode analysis for proteins. J. Mol. Struct. THEOCHEM. 898: 42–48.

104. Tama, F., and Y.H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. Protein Eng. 14: 1–6.

105. Lindahl, E., C. Azuara, P. Koehl, and M. Delarue. 2006. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res. 34: 52-6.

106. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case. 2004. Development and testing of a general amber force field. J. Comput. Chem. 25: 1157–1174.

107. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur. Biophys. J. 40: 843–856.

108. MacKerell, A.D., 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.K. 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. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102: 3586–3616.

109. Cornell, W.D., P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117: 5179–5197.

110. Cheatham, T.E., P. Cieplak, and P.A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J. Biomol. Struct. Dyn. 16: 845–62.

111. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins Struct. Funct. Genet. 65: 712–725.

112. Pérez, A., I. Marchán, D. Svozil, J. Sponer, T.E. Cheatham, C. Laughton, and M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. 92: 3817–29.

113. Allwright, J.C. 1976. Conjugate gradient versus steepest descent. J. Optim. Theory Appl. 20: 129–134.

114. Verlet, L. 1967. Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159: 98–103.

115. Ryckaert, J.-P., G. Ciccotti, and H.J.. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23: 327–341.

116. Case, D.A., T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods. 2005. The Amber biomolecular simulation programs. J. Comput. Chem. 26: 1668–1688.

117. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.

118. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen. 2005. GROMACS: Fast, flexible, and free. J. Comput. Chem. 26: 1701–1718.

119. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak. 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81: 3684–3690.

120. Feller, S.E., Y. Zhang, R.W. Pastor, and B.R. Brooks. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103: 4613–4621.

121. Leach, A.R. 2001. Molecular modelling : principles and applications. Prentice Hall.

122. Eastman, P., and V.S. Pande. 2010. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. J. Comput. Chem. 31: 1268–72.

123. Sagui, C., and T.A. Darden. 1999. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. Annu. Rev. Biophys. Biomol. Struct. 28: 155–179.

124. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98: 10089.

125. Anandakrishnan, R., A. Drozdetski, R.C. Walker, and A. V Onufriev. 2015. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. Biophys. J. 108: 1153–64.

126. Florová, P., P. Sklenovský, P. Banáš, and M. Otyepka. 2010. Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. J. Chem. Theory Comput. 6: 3569–3579.

127. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926.

128. Hamelberg, D., J. Mongan, and J.A. McCammon. 2004. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J. Chem. Phys. 120: 11919–11929.

129. Ermak, D.L., and J.A. McCammon. 1978. Brownian dynamics with hydrodynamic interactions. J Chem Phys. 69: 1352–1360.

130. Martinez, M., N.J. Bruce, J. Romanowska, D.B. Kokh, M. Ozboyaci, X. Yu, M.A. Öztürk, S. Richter, and R.C. Wade. 2015. SDA 7: A modular and parallel implementation of the simulation of diffusional association software. J. Comput. Chem. 36: 1631–1645.

131. Baker, N. a, D. Sept, S. Joseph, M.J. Holst, and J. a McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 98: 10037–41.

132. Madura, J.D., J.M. Briggs, R.C. Wade, M.E. Davis, B. a. Luty, A. Ilin, J. Antosiewicz, M.K. Gilson, B. Bagheri, L.R.R. Scott, and J.A.A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput. Phys. Commun. 91: 57–95.

133. Gabdoulline, R.R., and R.C. Wade. 1996. Effective Charges for Macromolecules in Solvent. J. Phys. Chem. 100: 3868–3878.

134. Elcock, A.H., R.R. Gabdoulline, R.C. Wade, and J.A. McCammon. 1999. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. J. Mol. Biol. 291: 149–162.

135. Gabdoulline, R.R., and R.C. Wade. 2009. On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer event. J. Am. Chem. Soc. 131: 9230–9238.

136. Smoluchowski, M. v. 1918. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. Zeitschrift für Phys. Chemie. 92: 129–168.

137. Öztürk, M.A., G. V Pachov, R.C. Wade, and V. Cojocaru. 2016. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. Nucleic Acids Res. 44: 6599–6613.

138. Crane-Robinson, C. 2015. Linker histones: History and current perspectives. Biochim. Biophys. Acta. 3 : 431-435

139. Catez, F., T. Ueda, and M. Bustin. 2006. Determinants of histone H1 mobility and chromatin binding in living cells. Nat. Struct. Mol. Biol. 13: 305–10.

140. Maeshima, K., R. Imai, S. Tamura, and T. Nozaki. 2014. Chromatin as dynamic 10-nm fibers. Chromosoma. 123: 225–237.

141. Robinson, P.J., and D. Rhodes. 2006. Structure of the “30 nm” chromatin fibre: A key role for the linker histone. Curr. Opin. Struct. Biol. 16: 336–343.

142. Robinson, P.J.J., L. Fairall, V.T. Huynh, and D. Rhodes. 2006. EM measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. Proc. Natl. Acad. Sci. 103: 6506–11.

143. Perišić, O., R. Collepardo-Guevara, and T. Schlick. 2010. Modeling studies of chromatin fiber structure as a function of DNA linker length. J. Mol. Biol. 403: 777–802.

144. Syed, S.H., D. Goutte-Gattat, N. Becker, S. Meyer, M.S. Shukla, J.J. Hayes, R. Everaers, D. Angelov, J. Bednar, and S. Dimitrov. 2010. Single-base resolution mapping of H1–nucleosome interactions and 3D organization of the nucleosome. Proc. Natl. Acad. Sci. 21:9620-5

145. Fan, L., and V.A. Roberts. 2006. Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. Proc. Natl. Acad. Sci. 103: 8384–8389.

146. George, E.M., T. Izard, S.D. Anderson, and D.T. Brown. 2010. Nucleosome interaction surface of linker histone H1c is distinct from that of H10. J. Biol. Chem. 285: 20891–20896.

147. Luque, A., R. Collepardo-Guevara, S. Grigoryev, and T. Schlick. 2014. Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. Nucleic

Acids Res. 42: 7553–7560.

148. Ngo, T.T.M., Q. Zhang, R. Zhou, J.G. Yodh, and T. Ha. 2015. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. Cell. 160: 1135–44.

149. von Hippel, P.H. 2007. From “simple” DNA-protein interactions to the macromolecular machines of gene expression. Annu. Rev. Biophys. Biomol. Struct. 36: 79–105.

150. Pearlman, D.A., D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput. Phys. Commun. 91: 1–41.

151. Joung, I.S., and T.E. Cheatham. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J. Phys. Chem. B. 112: 9020–9041.

152. Kräutler, V., W.F. Van Gunsteren, and P.H. Hünenberger. 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J. Comput. Chem. 22: 501–508.

153. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. J. Mol. Graph. 14: 33–38.

154. Roe, D.R., and T.E. Cheatham. 2013. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. 9: 3084–3095.

155. Dolinsky, T.J., P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, and N.A. Baker. 2007. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 35: W522-5.

156. Holst, M., N. Baker, and F. Wang. 2000. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples. J. Comput. Chem. 21: 1319–1342.

157. Ozer, G., A. Luque, and T. Schlick. 2015. The chromatin fiber: Multiscale problems and approaches. Curr. Opin. Struct. Biol. 31: 124–139.

158. Ponomarev, S.Y., V. Putkaradze, and T.C. Bishop. 2009. Relaxation dynamics of nucleosomal DNA. Phys. Chem. Chem. Phys. 11: 10633.

159. White, A.E., A.R. Hieb, and K. Luger. 2016. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. Sci. Rep. 6: 19122.

160. Hong, J., H. Feng, Z. Zhou, R. Ghirlando, and Y. Bai. 2013. Identification of functionally conserved regions in the structure of the chaperone/CenH3/H4 complex. J. Mol. Biol. 425: 536–545.

161. Vyas, P., and D.T. Brown. 2012. N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1(0) and H1c. J. Biol. Chem. 287: 11778–87.

162. Caterino, T.L., and J.J. Hayes. 2011. Structure of the H1 C-terminal domain and function in chromatin condensation. Biochem. Cell Biol. 89: 35–44.

163. Shaytan, A.K., G.A. Armeev, A. Goncearenco, V.B. Zhurkin, D. Landsman, and A.R. Panchenko. 2016. Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions. J. Mol. Biol. 428: 221–37.

164. Oudet, P., M. Gross-Bellard, and P. Chambon. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell. 4: 281–300.

165. Bayona-Feliu, A., A. Casas-Lamesa, A. Carbonell, P. Climent-Cantó, M. Tatarski, S. Pérez-Montero, F. Azorín, and J. Bernués. 2016. Histone H1: Lessons from Drosophila. Biochim. Biophys. Acta - Gene Regul. Mech. 1859: 526–532.

166. Orrego, M., I. Ponte, A. Roque, N. Buschati, X. Mora, and P. Suau. 2007. Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. BMC Biol. 5: 22.

167. Clausell, J., N. Happel, T.K. Hale, D. Doenecke, and M. Beato. 2009. Histone H1 Subtypes Differentially Modulate Chromatin Condensation without Preventing ATP-Dependent Remodeling by SWI/SNF or NURF. PLoS One. 4: e0007243.

168. Garg, M., L.R. Perumalsamy, G. V Shivashankar, and A. Sarin. 2014. The linker histone h1.2 is an intermediate in the apoptotic response to cytokine deprivation in T-effectors. Int. J. Cell Biol. 2014: 674753.

169. Terme, J.-M., B. Sesé, L. Millán-Ariño, R. Mayor, J.C.I. Belmonte, M.J. Barrero, and A. Jordan. 2011. Histone H1 Variants Are Differentially Expressed and Incorporated into Chromatin during Differentiation and Reprogramming to Pluripotency. J. Biol. Chem. 286: 35347–35357.

170. Happel, N., J. Warneboldt, K. Hänecke, F. Haller, and D. Doenecke. 2009. H1 subtype expression during cell proliferation and growth arrest. Cell Cycle. 8: 2226–2232.

171. Kowalski, A., and J. Pałyga. 2012. Linker histone subtypes and their allelic variants. Cell

Biol. Int. 36: 981–996.

172. Collepardo-Guevara, R., G. Portella, M. Vendruscolo, D. Frenkel, T. Schlick, and M. Orozco. 2015. Chromatin Unfolding by Epigenetic Modifications Explained by Dramatic Impairment of Internucleosome Interactions: A Multiscale Computational Study. J. Am. Chem. Soc. 137: 10205–10215.

173. Collepardo-Guevara, R., and T. Schlick. 2014. Chromatin fiber polymorphism triggered by variations of DNA linker lengths. Proc. Natl. Acad. Sci. U. S. A. 111: 8061–6.

174. Grigoryev, S. a, G. Arya, S. Correll, C.L. Woodcock, and T. Schlick. 2009. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. Proc. Natl. Acad. Sci. U. S. A. 106: 13317–13322.

175. Trott, O., and A.J. Olson. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31: 455–61.

176. Papamokos, G. V, G. Tziatzos, D.G. Papageorgiou, S.D. Georgatos, A.S. Politou, and E. Kaxiras. 2012. Structural role of RKS motifs in chromatin interactions: a molecular dynamics study of HP1 bound to a variably modified histone tail. Biophys. J. 102: 1926–1933.

177. Erler, J., R. Zhang, L. Petridis, X. Cheng, J.C.C. Smith, and J. Langowski. 2014. The role of histone tails in the nucleosome: a computational study. Biophys. J. 107: 2911–2922.

178. Zhang, R., J. Erler, and J. Langowski. 2017. Histone acetylation regulates chromatin accessibility: Role of H4K16 in inter-nucleosome interaction. Biophys. J. 112: 450–459.

179. Heo, L., H. Park, and C. Seok. 2013. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 41: W384–W388.

180. The PyMOL Molecular Graphics System, Version 1.7.2.1 Schrödinger, L. PyMOL. .

181. Warnecke, A., T. Sandalova, A. Achour, and R.A. Harris. 2014. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. BMC Bioinformatics. 15: 370.

182. Chung, H.H., S.K. Sze, A.R.E. Woo, Y. Sun, K.H. Sim, X.M. Dong, and V.C.-L. Lin. 2014. Lysine methylation of progesterone receptor at activation function 1 regulates both ligand-independent activity and ligand sensitivity of the receptor. J. Biol. Chem. 289: 5704–5722.

183. Homeyer, N., A.H.C. Horn, H. Lanig, and H. Sticht. 2006. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. J. Mol. Model. 12: 281–289.

184. Ponder, J.W., and D.A. Case. 2003. Force fields for protein simulations. Adv. Protein Chem. 66: 27–85.

185. Jin, Y., A.M. Rodriguez, J.D. Stanton, A.A. Kitazono, and J.J. Wyrick. 2007. Simultaneous mutation of methylated lysine residues in histone H3 causes enhanced gene silencing, cell cycle defects, and cell lethality in Saccharomyces cerevisiae. Mol. Cell. Biol. 27: 6832–6841.

186. Oueslati, A., K.E. Paleologou, B.L. Schneider, P. Aebischer, and H.A. Lashuel. 2012. Mimicking phosphorylation at serine 87 inhibits the aggregation of Human α-synuclein and protects against its toxicity in a rat model of Parkinson's disease. J. Neurosci. 32: 1536–1544.

187. Edidin, M., Y. Zagyansky, and T.J. Lardner. 1976. Measurement of membrane protein lateral diffusion in single cells. Science. 191: 466–468.

188. Axelrod, D., P. Ravdin, D.E. Koppel, J. Schlessinger, W.W. Webb, E.L. Elson, and T.R. Podleski. 1976. Lateral motion of fluorescently labeled acetylcholine receptors in membranes of developing muscle fibers. Proc. Natl. Acad. Sci. 73: 4594–4598.

189. Partikian, A., B. Olveczky, R. Swaminathan, Y. Li, and A.S. Verkman. 1998. Rapid diffusion of green fluorescent protein in the mitochondrial matrix. J. Cell Biol. 140: 821–829.

190. Reits, E.A., and J.J. Neefjes. 2001. From fixed to FRAP: measuring protein mobility and activity in living cells. Nat. Cell Biol. 3: 145-147.

191. Stasevich, T.J., F. Mueller, D.T. Brown, and J.G. McNally. 2010. Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. EMBO J. 29: 1225–1234.

192. Lele, T., S.R. Wagner, J.A. Nickerson, and D.E. Ingber. 2006. Methods for measuring rates of protein binding to insoluble scaffolds in living cells: histone H1-chromatin interactions. J. Cell. Biochem. 99: 1334–1342.

193. Beaudouin, J., F. Mora-Bermúdez, T. Klee, N. Daigle, and J. Ellenberg. 2006. Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins. Biophys. J. 90: 1878–1894.

194. Tabaka, M., T. Kalwarczyk, J. Szymanski, S. Hou, and R. Holyst. 2014. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. Front. Phys. 2.

195. Mourão, M.A., J.B. Hakim, and S. Schnell. 2014. Connecting the dots: The effects of macromolecular crowding on cell physiology. Biophys. J. 107: 2761–2766.

196. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.

197. Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.G. Cassarino, M. Bertoni, L. Bordoli, and T. Schwede. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42: 252-258.

198. Dolinsky, T.J., J.E. Nielsen, J.A. McCammon, and N.A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32: 665-667.

199. Minton, A.P. 2005. Influence of macromolecular crowding upon the stability and state of association of proteins: Predictions and observations. J. Pharm. Sci. 94: 1668–1675.

200. Spaar, A., C. Dammer, R.R. Gabdoulline, R.C. Wade, and V. Helms. 2006. Diffusional encounter of barnase and barstar. Biophys. J. 90: 1913–1924.

201. Mattiroli, F., S. Bhattacharyya, P.N. Dyer, A.E. White, K. Sandman, B.W. Burkhart, K.R. Byrne, T. Lee, N.G. Ahn, T.J. Santangelo, J.N. Reeve, and K. Luger. 2017. Structure of histone-based chromatin in Archaea. Science. 357: 609–612.

202. Müller, S., and G. Almouzni. 2017. Chromatin dynamics during the cell cycle at centromeres. Nat. Rev. Genet. 18: 192–208.