

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for
Mathematics
of the Ruperto-Carola University of Heidelberg,
Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M.Sc.(Physics): Georgi V. Pachov
Born in: Plovdiv, Bulgaria

Oral examination: 16.12.2009

Linker Histone-Nucleosome Interactions:
A Modelling and Simulation Study

Referees:

Prof. Dr. Jeremy C. Smith
Prof. Dr. Jörg Langowski

Im Zellkern wickelt sich die DNS um Histonproteine, bildet dabei Nukleosomenteilchen und packt diese in eine stark negativ geladene Struktur, die Chromatinfiber. Das Linker-Histon ist ein Protein, welches an das Nukleosom bindet und die Verknüpfung der Nukleosomen untereinander bestimmt. Um die Nukleosom-Linker-Histon-Wechselwirkung zu simulieren, wurden Simulationsmethoden der Brown'sche Dynamik (BD) und die Normalmodenanalyse (NMA) benutzt. Die NMA des Nukleosoms zeigte die wichtigsten Bewegungsmoden der beiden Linker-DNS Moleküle. Diese Ergebnisse wurden benutzt, um Konformationen der Linker-DNS zu generieren, welche in BD Simulationen zum Docking mit starren Körpern zwischen einer Linker-DNS und deren Mutanten an das Nukleosom eingesetzt wurden. Aus diesen Simulationen konnten zwei unterschiedliche Bindungsstellen und eine Nicht-Bindungsstelle auf dem Linker-Histon identifiziert werden. Die Aminosäuren, die sich in den Simulationen als wichtig für die Bindung herausstellten, sind konsistent mit experimentellen Daten. Darüberhinaus wurde für eine große Anzahl von Linker-DNS-Konformationen eine einzige vorherrschende Bindungsstelle des Linker-Histons an das Nukleosom gefunden. Basierend auf den Ergebnissen des BD Docking wurden die Assoziationsgeschwindigkeiten des Linker-Histons an das Nukleosom modelliert. Die erhaltenen hohen Assoziationsgeschwindigkeiten in der Nähe der Diffusionsgrenze ($10^9 \text{ M}^{-1} \text{ s}^{-1}$) wurden einer elektrostatischen Steuerung zwischen beiden Biomolekülen zugeordnet. Ein neues Verfahren zur Berücksichtigung von molekularer Flexibilität in den BD Simulationen wurde entwickelt. Hierbei wird ein Bindungspartner als ein diskreter Satz von strukturellen Konformationen behandelt, die die Flexibilität repräsentieren. Der Übergang zwischen den Konformationen wird durch einen Markow-Prozess beschrieben, dessen Wahrscheinlichkeiten vier verschiedenen Auswahlalgorithmen folgen: drei von ihnen basieren auf der Wechselwirkungsenergie zwischen den Molekülen und der vierte basiert auf einer Zufallsauswahl. Diese Methode wurde erfolgreich auf das Verknüpfungshiston-Nukleosomensystem angewandt, wobei das Nukleosom als flexibel behandelt wurde. Die Ergebnisse zeigen, dass das Verknüpfungshiston bevorzugt an eine offene Konformation des Nukleosoms bindet. Basierend auf den Ergebnissen der BD Simulationen wurde die Dynamik des Begegnungskomplexes in einer Molekulardynamiksimulation in atomaren Details auf einer Zeitskala von $\sim 10 \text{ ns}$ modelliert. Die Simulation ergab einen stabilen biomolekularen Komplex mit einer konformationellen Änderung einer Schleifenregion des Verknüpfungshistons, die einem Induced-Fit-Effekt zugeordnet werden konnte. Auf diese Art und Weise konnten Einsichten gewonnen werden in die maßgebenden Umstände der Verknüpfungshiston-Nukleosom Bindung. Zusätzlich werden diese Ergebnisse von Bedeutung sein für eine Modellierung der Chromatinfaser auf höhere Ebene.

In the cell nucleus, DNA wraps around histone proteins, forming nucleosome particles, and packs into a highly negatively charged structure, the chromatin fiber. The linker histone is a protein that binds to the nucleosome and determines how the nucleosomes are linked to each other. To simulate the nucleosome-linker histone interactions, a Brownian Dynamics (BD) technique together with normal mode analysis (NMA) was applied. NMA of the nucleosome revealed the most prominent modes of motion of its two linker DNAs. The results were used to generate conformations of the linker DNAs which were used in BD simulations of the rigid-body docking of a linker histone and its mutants to the nucleosome. From the simulations, two distinct binding sites and one non-binding site on the linker histone were identified. The amino acids found to be most important for binding in the simulations with the linker histone mutants are consistent with experimental data. Moreover, a dominant binding mode of the linker histone to the nucleosome was found for a wide range of conformations of the linker DNAs. The association kinetics of the linker histone to the nucleosome was modeled using the results obtained by the BD docking. The obtained high association rates close to the diffusional limit ($10^9 \text{ M}^{-1} \text{ s}^{-1}$) were attributed to electrostatic steering between both biomolecules. A new method accounting for molecular flexibility during BD simulation was developed. One of the binding partners is treated as a discrete set of structural conformations representing its flexibility. The transition between the conformations is described by a Markovian process with probability following four different selection algorithms: three are based on the interaction energy between the molecules and the fourth is based on random selection. The method was successfully applied to the linker histone-nucleosome system, where the nucleosome was modeled as a flexible molecule. The result suggests that the linker histone preferentially binds to more open nucleosome conformations. Following the BD results obtained, the encounter complex dynamics was modeled by atomic detail Molecular Dynamics (MD), taking into account all degrees of freedom in classical mechanics on $\sim 10 \text{ ns}$ time scale. The simulation resulted in a stable biomolecular complex with a conformational change on the loop region of the linker histone, which could be attributed to an induced fit effect. As well as providing insights into the determinants of linker histone-nucleosome binding, the results are expected to be valuable for higher order modelling of the chromatin fiber.

The work in this thesis was carried out under the supervision of
Dr. Rebecca Wade, Head of the Molecular and Cellular
Modeling group at the EML Research, Heidelberg, Germany

Acknowledgements

I would like to acknowledge the following individuals and institutions for their generous assistance during my PhD study.

- First I would like to acknowledge Dr. Rebecca Wade for her excellent guidance during these 3.5 years.
- I thank Prof. Dr. Jeremy C. Smith and Prof. Dr. Jörg Langowski for their valuable discussions throughout this research project. I am also thankful to our collaborators.
- Special thank deserve all the members of the MCM group at EML Research for their support and encouragement during my doctoral study. Particularly, I am deeply grateful to Razif Gabdoulline for introducing me to the SDA software package and to Vlad Cojocaru for his help and assistance during the MD simulations.
- My thanks go to the administration at the EML Research for their assistance throughout my work.
- I gratefully acknowledge Deutsche Forschungsgemeinschaft (German Research Foundation) and Klaus Tschira Foundation for their financial support.
- I would like also to thank my family and friends for their support and encouragement during the PhD study.

Contents

Abstract	v
Acknowledgements	ix
1 Introduction	1
2 Theoretical Overview	5
2.1 Statistical physics in biology	5
2.1.1 Equilibrium statistical mechanics	5
2.2 Brownian motion	7
2.2.1 Smoluchowski theory	7
2.2.2 Einstein theory	8
2.2.3 Langevin and Smoluchowski equations	9
2.3 Reaction kinetics	10
2.4 Small (thermal) fluctuations	12
2.5 Dynamics of Biomolecular Association	14
2.5.1 Bimolecular Association	14
2.5.2 Molecular interaction	15
3 Chromatin Fiber	19
3.1 Structure and biological function	19
3.1.1 Chromatin	19
3.1.2 Nucleosome	21
3.1.3 Linker histone and its role in chromatin	22
4 Computational Methods for Studying Biomolecular Dynamics and Interactions	25
4.1 Normal Mode Analysis	26
4.1.1 Elastic Network Model	26
4.2 Brownian Dynamics	27
4.2.1 Ermak-McCammon algorithm	27
4.2.2 Simulation of diffusional association	28
4.3 Molecular Dynamics	34
4.3.1 Force fields	34
4.3.2 Velocity <i>Verlet</i> algorithm	36
4.3.3 Thermostats and Barostats	37
5 Brownian Dynamics docking of the linker histone to the nucleosome	39
5.1 System preparation	40
5.2 NMA	40

CONTENTS

5.3	Electrostatic potential calculation	45
5.4	Effective charges calculation	46
5.5	Brownian Dynamics setup	46
5.6	Structure of the nucleosome-linker histone GH5	48
5.7	Summary and implications to chromatin fiber	54
6	Brownian Dynamics docking of the linker histone mutants to the nucleosome	57
6.1	Mutants	57
6.2	BD docking and analysis of the results	58
6.2.1	Nucleosome-linker histone complexes	59
6.2.2	Dynamics	63
6.2.3	Energetics	67
6.2.4	Statistics	68
6.3	Summary	73
7	Association rates of GH5 to the nucleosome by Brownian Dynamics	75
7.1	Brownian Dynamics set up	76
7.2	Association rates results	78
7.2.1	Rates to the reference structure 7_0	78
7.2.2	Rates to the extreme conformations 7_6 and 8_3	82
7.2.3	Summary	84
8	Flexibility in Brownian Dynamics simulation	87
8.1	Metropolis Monte Carlo algorithm	87
8.2	Methodology	89
8.2.1	Implementation	90
8.2.2	Discussion and limitations	92
8.3	Application to the nucleosome and the linker histone	93
8.4	Summary	99
9	Molecular Dynamics of the chromatosome particle	101
9.1	MD setup	101
9.2	Results	102
9.3	Summary	108
10	Conclusion and Outlook	111
A	Computation of the Brownian Dynamics forces and torques	117
	Bibliography	120

1

Introduction

The formation of biological complexes between proteins, proteins and small molecules, and proteins and nucleic acids plays a part in many biological processes, including gene transcription, cell signalling, enzyme catalysis and the immune response. Molecular association is governed by both the kinetic and the thermodynamic properties of the molecules and the medium involved. Inside cells, the medium is crowded with a variety of different macromolecules. In the cell nucleus, the DNA molecules compact to highly ordered chromatin structures assembling a biological network. Within this network, the DNA combines with other proteins and they together form complexes called nucleosomes, which in turn interact with each other. The conformation and compaction of the chromatin depend on the interactions between the nucleosomes as well as on the presence of other factors influencing chromatin structure and dynamics. One of these factors is the so-called *linker histone* protein, a family of DNA binding proteins [1, 2], which interacts with the nucleosome and contributes to the DNA compaction into 30 nm chromatin fiber [3]. Although the conformational properties of DNA have been the subject of intensive investigation, the physical structure of chromatin as well as of the nucleosome-linker histone complex remains unknown. Not only the structural properties, but also the dynamic features of chromatin and its components are not fully understood.

Chromatin and nucleosome structures have been studied intensively *in vitro* using different experimental techniques: x-ray crystallography [4, 5], force probe experiments [6–8] and neutron scattering [9, 10]. Other experimental techniques focus more on the particular interactions and dynamics within chromatin. Some of the most relevant to this study are those using fluorescent proteins to track and mark molecules of interest: Fluorescence Recovery After Photobleaching (FRAP) [11, 12], Fluorescence (Förster) Resonance Energy Transfer (FRET) [13] and Fluorescence Correlation Spectroscopy (FCS) [14–16]. For example, systematic mutagenesis combined with FRAP *in vivo* revealed the binding geometry of the linker histone H1⁰ within

the nucleosome in chromatin fiber [17]. At a nucleosomal level FRET experiments have shown the importance of different factors governing the interactions and dynamics within chromatin such as linker histones [18], DNA unwinding [19, 20] and salt effects [21]. Quantitative models of chromatin structure and dynamics are essential for interpreting these experiments.

Since chromatin has features on different time and length scales, a considerable number of theoretical models exist that aim to elucidate the driving forces for chromatin compaction [22–25]. However, the time scale in chromatin fiber varies from picoseconds for a small number of atoms to minutes for large biomolecules and considerable efforts are needed to describe all biomolecular interactions present. A way to tackle this problem is to use physical principles combined with computational methods in order to transfer information between different spatial and temporal scales. The macroscopic properties of a particular system are averaged over the microscopic events. Computational modelling of these events is not always reasonable when pronounced biological phenomena such as DNA-protein binding, need to be described. For example, Molecular Dynamics (MD), a computational technique for simulating biomolecules in atomic detail, is restricted by the number of atoms in the system. The reason is that in MD the system is sampled in phase space taking into account all classical degrees of freedom such as bond stretching and twisting, etc. Elimination of the non-essential degrees of freedom can lead to faster sampling and, thus, high energy barriers on complex biomolecular energy landscapes can be overcome. One computational method neglecting the explicit solvent degrees of freedom is called Brownian Dynamics. It models the diffusional motion in the configurational space and can reach time scales of seconds depending on details of the model applied. An advantageous step towards improving the BD method would account for flexible units modelled in atomic detail. In case of highly dynamic biological systems as chromatin this can be an important issue for reproducing experimental conditions.

An understanding of the linker histone protein dynamics in the formation of a linker histone-nucleosome complex is essential for the development of more robust chromatin models as well as for the linker histone biological function. The binding dynamics of the linker histone to the nucleosome was investigated experimentally [17, 26, 27] and theoretically [28–30]. However, these models and experimental data are not consistent with each other proposing different three-dimensional arrangements and interactions of the linker histone on the nucleosome. The binding of the linker histone and its stoichiometry as well as the nucleosome repeat lengths¹ influence chromatin

¹Nucleosome repeat length (NRL) is the total length of DNA wrapped around a nucle-

compaction leading to topologically distinguishable fibers [31]. The basic unit of chromatin, the nucleosome, consists of 147 base pairs (bp) of DNA [5] in the absence of linker histone. When the linker histone is present the nucleosome-linker histone complex contains 167 bp of DNA [32] and a *chromatosome* particle is formed. Due to the negatively and positively charged nucleosome and linker histone, respectively, the latter facilitates partially charge neutralization contributing to the folding of 30 nm chromatin fiber [3]. Therefore, precise knowledge of the chromatosome structure, which has not been resolved yet, is of major interest for the scientific community. In addition, in many biological processes, like DNA transcription, replication and repair, the proteins involved must quickly find their target site. The kinetics of such a process are directly influenced by the level of DNA exposure and histone tail acetylation on the nucleosome [33]. All these considerations are topics of the present study and they are tackled by physics based computational simulations and methods assessing different time scales.

This thesis aims at answering several important questions about nucleosome linker histone interactions:

- what is the structure of the nucleosome-linker histone complex?
- what are the binding sites and position of the linker histone with respect to the nucleosome?
- how is the binding affected by a replacement of amino acids with others in the structure of the linker histone?
- how fast does the linker histone bind to the nucleosome?
- does nucleosome flexibility affect the binding?
- is the bound state stable?

Thesis overview

The present thesis is organized in 10 chapters. In Chap. 2 an overview of basic physical theories is given. Statistical physics concepts as well as characteristics and determinants of a diffusional process are outlined. The kinetics and interactions involved in the formation of bimolecular² complexes are described in detail.

Chapter 3 introduces the system under investigation starting from the higher-order chromatin fiber through the nucleosome and ending with the

osome plus the DNA length connecting successive nucleosomes.

²Over the thesis the term 'bimolecular' will refer to two particles, i.e. 'biparticle'.

linker histone structure. Available chromatin models in the literature are briefly mentioned.

The computational methods and tools used in the thesis are described in Chap. 4. First, the Normal Mode Analysis (NMA) is presented by a model called Elastic Network Model (ENM) [34]. Then, the Brownian Dynamics algorithm is described in detail emphasizing specific simulation issues. Also, a brief introduction on the principles of Molecular Dynamics is given.

Chapter 5 is the first chapter presenting results. Normal Mode Analysis and Brownian Dynamics docking of linker histone H5 to 13 nucleosome conformations were carried out and the results are presented and discussed.

Chapter 6, describes the BD docking of wild type linker histone and 24 linker histone mutants, i.e. amino acids with specific residues replaced by others. The effect of these mutations on linker histone binding is discussed in detail.

Results on the binding kinetics of the linker histone to the nucleosome are presented in Chapter 7 and possible models for linker histone-nucleosome binding are proposed.

In Chapter 8 a new method accounting for molecular flexibility in Brownian Dynamics simulation is introduced. Results of its application to the nucleosome-linker histone system are shown as well.

In Chapter 9, the nucleosome-linker histone encounter complex is further refined to higher resolution with Molecular Dynamics simulation and preliminary results are shown and discussed.

A summary of the results is given in Chap. 10 and the relation of this work to the literature is briefly discussed.

2

Theoretical Overview

2.1 Statistical physics in biology

Biological science is dealing with living matter and its properties can be described by the methods of statistical physics. Nowadays they are widely applied to all kind of biological problems, including small and large biomolecules, membranes and colloids. All these entities, within and outside the cell, are not static objects, but also dynamic, performing their biological functions in accordance with the laws of physics. Below, the basic concepts of statistical mechanics will be given.

2.1.1 Equilibrium statistical mechanics

Biological systems consist of many atoms and molecules, which interact with each other and, thus, build a very complicated *macroscopic* system. In statistical physics, this *macroscopic* system is not described by its time evolution, but rather by the instantaneous state of copies of the system having different *microscopic* parameters. Such states are called *microstates* and this collection of copies is known as a *statistical ensemble* [35]. Each ensemble gives the probability of realization of the possible microstates. Each *macrostate* is characterized by macroscopic physical observables, e.g. temperature T , pressure P , after a suitable averaging procedure over the microstates. This averaging is called *ensemble averaging*. If the *time average* of a physical observable \mathcal{A} , $\langle \mathcal{A} \rangle_{\text{time}}$ is equal to its *ensemble average* $\langle \mathcal{A} \rangle_{\text{ensemble}}$ then a system fulfilling this condition is called *ergodic* [36].

Let a system be ergodic and described classically by canonical coordinates \mathbf{q} and conjugate momenta \mathbf{p} in phase space. Then the probability of a certain microstate being in a volume $d\Gamma = d\mathbf{q}d\mathbf{p}$ of the phase space Γ is $\rho(\mathbf{q}, \mathbf{p})d\Gamma$, where $\rho(\mathbf{q}, \mathbf{p})$ is the density distribution in the phase space. Then for any

physical observable $\mathcal{A}(\mathbf{q}, \mathbf{p})$ is valid

$$\langle \mathcal{A} \rangle_{\text{ensemble}} = \int_{\Gamma} \mathcal{A}(\mathbf{q}, \mathbf{p}) \rho(\mathbf{q}, \mathbf{p}) d\Gamma \equiv \langle \mathcal{A} \rangle_{\text{time}} = \lim_{t_{\text{obs}} \rightarrow \infty} \frac{1}{t_{\text{obs}}} \int_0^{t_{\text{obs}}} \mathcal{A}(\mathbf{q}, \mathbf{p}, t) dt. \quad (2.1)$$

The time t_{obs} describes the observable time interval. When the density distribution depends on the time, $\rho(\mathbf{q}, \mathbf{p}, t)$, but there is no creation and annihilation of states in the phase space, then *Liouville's theorem* holds [35]. It states that the density distribution in the phase space is conserved along any trajectory and mathematically is expressed by

$$\frac{d\rho(\mathbf{q}, \mathbf{p}, t)}{dt} = \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial t} + \dot{\mathbf{q}} \cdot \nabla_{\mathbf{q}} \rho(\mathbf{q}, \mathbf{p}, t) + \dot{\mathbf{p}} \cdot \nabla_{\mathbf{p}} \rho(\mathbf{q}, \mathbf{p}, t) = 0. \quad (2.2)$$

In Eq. 2.2, $\dot{\mathbf{q}}$ means differentiation with respect to time and $\nabla_{\mathbf{q}}$ and $\nabla_{\mathbf{p}}$ are the gradient operators acting over the coordinates \mathbf{q} and momenta \mathbf{p} , respectively.

Physical observables are macroscopic quantities, e.g. pressure P and volume V , defining the macroscopic system. Depending on the contact of the system with the surroundings, different statistical ensembles can be constructed. For example, a *canonical ensemble* is coupled to a thermal bath ($T = \text{const}$) in a fixed volume ($V = \text{const}$) having constant number of particles ($N = \text{const}$). Then a microstate of the system can be characterized by a Hamiltonian $H(\mathbf{q}, \mathbf{p})$ with probability

$$\rho(\mathbf{q}, \mathbf{p}) = \frac{1}{Z(T, V, N)} \exp\left(-\frac{H(\mathbf{q}, \mathbf{p})}{k_B T}\right), \quad (2.3)$$

where k_B is the Boltzmann constant and $Z(T, V, N)$ is the *partition function* given by

$$Z(T, V, N) = \int_{\Gamma} \exp\left(-\frac{H(\mathbf{q}, \mathbf{p})}{k_B T}\right) \frac{d\mathbf{q}d\mathbf{p}}{h^{3N} N!}. \quad (2.4)$$

The factor $1/N!$ takes into account the indistinguishability of the particles in the system, while $1/h^{3N}$ is a correction for the case of ideal gas. The partition function $Z(T, V, N)$ gives the total number of microscopic states with $T, V, N = \text{const}$. The free energy of the system reads

$$F(T, V, N) = -k_B T \ln Z(T, V, N). \quad (2.5)$$

In such a way, the macroscopic properties of a physical system in thermal equilibrium can be characterized by the statistical distributions for the different ensembles. Important relation between the microscopic and the macroscopic state is the so-called *equipartition theorem* [35]

$$\left\langle q_i \frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial q_i} \right\rangle_{\text{ensemble}} = \left\langle p_i \frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial p_i} \right\rangle_{\text{ensemble}} = k_B T, \quad (2.6)$$

stating that for every microscopic degree of freedom, the corresponding energy equals $k_B T/2$.

2.2 Brownian motion

A particle immersed in a fluid experiences inherent and incessant motion called *Brownian motion*. The name is in honour of Robert Brown, who investigated the way in which pollen acted during impregnation in 1827 and was the first trying to give a scientific explanation to his observations [37]. Later the irregular nature of the motion was examined in other cases and it was concluded that the smaller the size of the suspended objects, the greater their motions. This motion is caused by numerous collisions between the solute and the medium particles. In most cases, the solute particle is much heavier than the solvent particles and upon an elastic collision its velocity change will be very small. On average it will not even be observable, but the velocity fluctuations close to that average will have strong observable effect. In order to describe quantitatively these fluctuations; Smoluchowski developed a kinetic model for collisions of hard spheres on a microscopic scale (Sec. 2.2.1), whereas Einstein based his model on statistical assumptions on a mesoscopic level of detail (Sec. 2.2.2) [38].

2.2.1 Smoluchowski theory

In his theory, Smoluchowski assumed that the heavy particles do not collide with each other and the interactions with the solvent particles are described by elastic collisions. In liquid, a spherical particle with radius R will have a mean free path $l \ll R$ meaning that the collisions of the light particles with the heavy one are not independent events. Nevertheless, Smoluchowski made an important assumption stating that the heavy particle moves like on a polymer chain¹ with a mean free path $\langle l \rangle = \langle v \rangle \tau$ resembling a *gas* particle. The time $\tau = m/\zeta$ expresses the velocity relaxation time down to the mean velocity of the Brownian particle $\langle v \rangle$. The particle mass is denoted with m , while the solvent friction constant is ζ . Then Smoluchowski inserted $\langle l \rangle$ in the formula $\langle r^2 \rangle = \langle l^2 \rangle 2/\tau$ [39], where $\langle r^2 \rangle$ is the mean square displacement per unit time, and obtained

$$\langle r^2 \rangle = \frac{2m\langle v^2 \rangle}{\zeta}. \quad (2.7)$$

¹Smoluchowski did not use the definition of "a polymer chain"

Considering his calculations for a dilute system [38], Smoluchowski multiplied the above equation by a factor of $8/3\sqrt{2/3}$ and applying the *Stoke's formula* for the friction constant, $\zeta = 6\pi\eta R$, the mean square displacement for time t yielded

$$\langle r^2 \rangle_t = \frac{64}{81} \frac{2m\langle v^2 \rangle}{\pi\eta R} t = \frac{64}{81} \frac{6k_B T}{\pi\eta R} t. \quad (2.8)$$

The equipartition theorem was used for substituting the mean square velocity $\langle v^2 \rangle$ in Eq.(2.8), where T is the temperature of the system.

2.2.2 Einstein theory

Einstein described the probability density $p(\mathbf{r}, t)$ of a solute molecule in a fluid before and after a given time step δt . After such a time step the suspended particle will be displaced by a distance $\delta\mathbf{r}$. During the displacement, the solute will encounter a huge number of collisions with the solvent molecules leading to loss of memory for its initial position, i.e. the displacement will depend only on its previous position. Thus, the probability distribution yields

$$p(\mathbf{r}, t + \delta t) = \int p(\mathbf{r} - \delta\mathbf{r}, t) p'(\delta\mathbf{r}, \delta t) d\delta\mathbf{r}. \quad (2.9)$$

The probability $p'(\delta\mathbf{r}, \delta t)$ gives the position of the particle after time δt and using Fokker-Planck equation [38], the *diffusion equation* with a diffusion constant D for the Brownian particle reads

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = \lim_{t \rightarrow 0} \frac{1}{2\delta t} \int \Delta^2 p'(\delta\mathbf{r}, \delta t) d\delta\mathbf{r} \nabla^2 p(\mathbf{r}, t) \quad (2.10)$$

$$D = \lim_{t \rightarrow 0} \frac{1}{2\delta t} \int \Delta^2 p'(\delta\mathbf{r}, \delta t) d\delta\mathbf{r}. \quad (2.11)$$

Equation (2.10), if multiplied by r^2 and integrated over the space, gives the well known Einstein equation

$$\langle r^2 \rangle = 6Dt. \quad (2.12)$$

In similar manner the root-mean-square angle $\langle \Delta\theta^2 \rangle$ for a spherical particle can be expressed by the rotational diffusion constant D_R . The diffusion constants D and D_R can be related to the solvent viscosity η by the Stokes-Einstein relation

$$D = \frac{k_B T}{\zeta} = \frac{k_B T}{6\pi\eta R}, \quad (2.13)$$

$$D_R = \frac{k_B T}{\zeta_R} = \frac{k_B T}{8\pi\eta R^3}, \quad (2.14)$$

where ζ_R is the rotational friction constant. Using this relation and the equipartition theorem one can compare the Smoluchowski (Eq. 2.8) and Einstein (Eq. 2.12) mean square displacement $\langle r^2 \rangle$.

2.2.3 Langevin and Smoluchowski equations

In 1908 Langevin [40] combined Einstein and Smoluchowski theories by taking into account the force acting on the solute particle due to the presence of solvent molecules. He obtained an equation, which is known as *free Langevin equation*

$$m \frac{d^2 \mathbf{r}(t)}{dt^2} = -\zeta \frac{d\mathbf{r}(t)}{dt} + \mathbf{f}^R(t), \quad (2.15)$$

where m is the mass of the suspended particle, $-\zeta \dot{\mathbf{r}}$ is a mean or frictional force and $\mathbf{f}^R(t)$ is a fluctuating or random force. The former force describes the solute motion dynamically while the latter describes it in a probabilistic way. Assuming that the random force \mathbf{f}^R has the properties

$$\begin{aligned} \langle \mathbf{f}^R(t) \rangle &= 0 \\ \langle \mathbf{f}^R(t) \cdot \mathbf{f}^R(t + \delta t) \rangle &= \mathcal{C} \delta(\delta t) \\ \langle \mathbf{f}^R(t) \cdot \mathbf{r}(t - \delta t) \rangle &= 0 \\ \langle \mathbf{f}^R(t) \cdot \dot{\mathbf{r}}(t - \delta t) \rangle &= 0, \end{aligned} \quad (2.16)$$

i.e. the random collisions are instantaneous and uncorrelated, the mean square solution of Eq. 2.15 yields

$$\langle \dot{\mathbf{r}}^2 \rangle = \dot{\mathbf{r}}_0^2 \exp\left(\frac{-2\zeta t}{m}\right) + \frac{\mathcal{C}}{2m\zeta} \left[1 - \exp\left(\frac{-2\zeta t}{m}\right)\right]. \quad (2.17)$$

If $t \rightarrow \infty$, i.e. the particle velocity relaxes towards its equilibrium value, then with the usage of the equipartition theorem the *fluctuation-dissipation theorem* is obeyed

$$\frac{\mathcal{C}}{2m\zeta} = k_B T. \quad (2.18)$$

Under this condition the Einstein result for the mean square displacement is deduced (see Eq. (2.12)). The fluctuation-dissipation theorem links nonequilibrium to equilibrium processes, but only when the system is not far from thermal equilibrium. This means that the amplitudes \mathcal{C} and ζ of the fluctuating and dissipative forces, respectively, should satisfy a temperature-dependent relation. For a Brownian motion defined by Einstein's theory, the fluctuations are caused by the thermal motions of the solvent molecules and the inertia of the suspended particle is not taken into account. Consequently,

Einstein describes his theory in configuration space, i.e. only the positional distribution of the Brownian particle is followed. This is correct when the velocity relaxation time is much faster than the time needed to reach positional equilibrium, i.e. $t \gg m/\zeta$. Usually, experiments investigating diffusional motion probe the process in configuration space. The reason is that in dense systems, diffusion is quite slow, while velocity relaxation is rapid. Therefore, the ordinary diffusion equation (Eq. 2.10) should be an adequate description except at extremely short times.

If there is an external force $\mathbf{F}(\mathbf{r})$ acting on the Brownian particle, the Langevin equation reads

$$m \frac{d^2 \mathbf{r}(t)}{dt^2} = -\zeta \frac{d\mathbf{r}(t)}{dt} + \mathbf{F}(\mathbf{r}) + \mathbf{f}^R(t). \quad (2.19)$$

Assuming that there is no acceleration Eq. (2.19) yields

$$\begin{aligned} 0 &= -\zeta \frac{d\mathbf{r}(t)}{dt} + \mathbf{F}(\mathbf{r}) + \mathbf{f}^R(t) \\ \dot{\mathbf{r}} &= \frac{\mathbf{F}(\mathbf{r}) + \mathbf{f}^R(t)}{\zeta}. \end{aligned} \quad (2.20)$$

Using Fokker-Planck equation [38] one obtains for the distribution function an expression known as *Smoluchowski equation*

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = -\frac{1}{\zeta} \frac{\partial [\mathbf{F}(\mathbf{r})p(\mathbf{r}, t)]}{\partial \mathbf{r}} + D \nabla^2 p(\mathbf{r}, t). \quad (2.21)$$

This equation leads to the diffusion equation (Eq. 2.10) if no external force $\mathbf{F}(\mathbf{r})$ is present. The current derivation of the Smoluchowski equation is not really justified, because only on average there is no acceleration. However, since the final result is correct and this is what we are interested in, we will not discuss the rigorous derivation [38].

2.3 Reaction kinetics

Reaction association can be considered to consist of two steps: first an intermediate is formed by diffusion: this is called a 'diffusional encounter complex'. Then this intermediate evolves to form a tightly bound complex. Association is diffusion-controlled when the first step is rate-limiting, and it is reaction-controlled when the second step determines the rate of the association process.

If a molecule of type X forms a complex Z with a molecule of type Y then the rate of formation of species Z and the rate of depletion of X are given by

$$\nu_Z = \frac{dC_Z(\mathbf{r}, t)}{dt}, \quad \nu_X = -\frac{dC_X(\mathbf{r}, t)}{dt}, \quad (2.22)$$

where $C_Z(\mathbf{r}, t)$ and $C_X(\mathbf{r}, t)$ are the concentrations of the Z and X species, respectively. This reaction process depends on the association and dissociation rate constants k_{on} and k_{off} , respectively



The reaction flux $J(\mathbf{r}, t)$ can be expressed via the rate constant k by the equation

$$J(\mathbf{r}, t) = kC^m(\mathbf{r}, t), \quad (2.24)$$

where m denotes the order of a reaction. For our case we have

$$J_Z(\mathbf{r}, t) = k_{\text{off}}C_Z(\mathbf{r}, t) \quad (2.25)$$

$$J_X(\mathbf{r}, t) = J_Y(\mathbf{r}, t) = k_{\text{on}}C_X(\mathbf{r}, t)C_Y(\mathbf{r}, t), \quad (2.26)$$

where Eq. 2.25 describes a first-order reaction whereas binding is a second-order process (see Eq. 2.26). The rate constants can be related to an equilibrium association constant

$$K_a = \frac{k_{\text{on}}}{k_{\text{off}}}. \quad (2.27)$$

The reciprocal of K_a is the equilibrium dissociation constant K_d .

Let molecules X and Y be uniform spheres having radii R_X and R_Y , respectively. If Y diffuses towards X then the process is described by the diffusion equation (Eq. 2.10)

$$\frac{\partial C_Y(\mathbf{r}, t)}{\partial t} = -D_Y \nabla_{\mathbf{r}}^2 C_Y(\mathbf{r}, t), \quad (2.28)$$

where D_Y is the diffusion constant of the molecule Y . We assume that D_Y is spatially independent. If there is a steady state, i.e. the particle flux is conserved, then

$$C_Y(\mathbf{r}, t) = C_Y(\mathbf{r}) \quad (2.29)$$

$$\nabla_{\mathbf{r}}^2 C_Y(\mathbf{r}) = 0. \quad (2.30)$$

The solution of the above equation in spherical coordinates is given by

$$C_Y(\mathbf{r}) = -\frac{a}{r} + b, \quad (2.31)$$

where a and b are integration parameters. They can be determined applying boundary conditions to the problem

- $C_Y(r \rightarrow R_X + R_Y) = 0$
- $C_Y(r \rightarrow \infty) = C_Y(\infty)$.

For flux through a spherical surface using Eq. (2.26) an analytical solution of the diffusion-controlled association constant k_{on} can be obtained for uniform spheres reacting at a distance $r = R_X + R_Y$ [41]

$$k_{\text{on}} = 4\pi r(D_X + D_Y), \quad (2.32)$$

where D_X is the diffusion constant for molecules of type X. Eq. 2.32 is valid when there are no forces between the spheres. In the case of interacting spheres using the Smoluchowski equation (Eq. 2.21), k_{on} is given by [42]

$$k_{\text{on}} = \frac{4\pi(D_X + D_Y)}{\int_r^\infty \frac{e^{U(\mathbf{r})/kT}}{r^2} d\mathbf{r}}, \quad (2.33)$$

where $U(\mathbf{r})$ is a centrosymmetric interaction potential between the spheres. For more complicated geometries and interaction forces, numerical approaches are necessary to compute association rates (Sec. 4.2.2) [43, 44].

2.4 Small (thermal) fluctuations

Let a system of N atoms be slightly perturbed from its equilibrium position \mathbf{q}^0 as²

$$\mathbf{q} = \mathbf{q}^0 + \boldsymbol{\xi}, \quad (2.34)$$

where $\boldsymbol{\xi}$ is the displacement from \mathbf{q}^0 . The potential energy \mathcal{V} can be expanded around the equilibrium position \mathbf{q}^0

$$\begin{aligned} \mathcal{V}(\mathbf{q}) = & \mathcal{V}(\mathbf{q}^0) + \sum_{i=1}^N \left(\frac{\partial \mathcal{V}}{\partial q_i} \right)_{\mathbf{q}^0} \xi_i + \frac{1}{2} \sum_{i,j=1}^N \left(\frac{\partial^2 \mathcal{V}}{\partial q_i \partial q_j} \right)_{\mathbf{q}^0} \xi_i \xi_j + \\ & + \frac{1}{6} \sum_{i,j,k=1}^N \left(\frac{\partial^3 \mathcal{V}}{\partial q_i \partial q_j \partial q_k} \right)_{\mathbf{q}^0} \xi_i \xi_j \xi_k + \dots, \end{aligned} \quad (2.35)$$

where we can denote

$$F_0 = \left(\frac{\partial \mathcal{V}}{\partial q_i} \right)_{\mathbf{q}^0} = 0 \quad (2.36)$$

$$K_{ij} = \left(\frac{\partial^2 \mathcal{V}}{\partial q_i \partial q_j} \right)_{\mathbf{q}^0} \quad (2.37)$$

$$\Upsilon_{ijk} = \left(\frac{\partial^3 \mathcal{V}}{\partial q_i \partial q_j \partial q_k} \right)_{\mathbf{q}^0}. \quad (2.38)$$

²For clarity the index i in \mathbf{q}_i^0 for atom i is omitted.

2.4. SMALL (THERMAL) FLUCTUATIONS

The total force acting on the system can be expressed by

$$\mathbf{F}(\mathbf{q}) = -\nabla_{\mathbf{q}}\mathcal{V}(\mathbf{q}) = -\mathbf{K}\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\Upsilon}\boldsymbol{\xi}^2. \quad (2.39)$$

When the displacement $\boldsymbol{\xi}$ is very small, then the potential energy \mathcal{V} can be approximated to its second term and the higher-order terms can be neglected. In this case the force $\mathbf{F}(\mathbf{q}) = -\mathbf{K}\boldsymbol{\xi}$ is called quasi elastic or *elastic*. The potential energy is described by harmonic potential and the kinetic energy is given by

$$\mathcal{T} = \frac{1}{2} \sum_{i,j=1}^N M_{ij} \dot{\xi}_i \dot{\xi}_j. \quad (2.40)$$

Both matrices, K_{ij} and M_{ij} , are real and symmetric, i.e. $K_{ij} = K_{ji}$. The system can be described by N equations of motion around the equilibrium position

$$\sum_j M_{ij} \ddot{\xi}_j + \sum_j K_{ij} \xi_j = 0, \quad (2.41)$$

where \mathbf{M} and \mathbf{K} are the mass and force matrices of the system, respectively. The solution of Eq. (2.41) can be given by $\xi_j = a_j \exp(-i\omega t)$, where a_j and ω are the amplitude and angular frequency of the motion, respectively. In a matrix-vector form, Eq. (2.41) can be represented by an eigenvalue problem

$$\mathbf{K}\mathbf{A} = \boldsymbol{\Lambda}\mathbf{M}\mathbf{A} \quad (2.42)$$

$$\mathbf{A}^\dagger \mathbf{M}\mathbf{A} = \mathbf{I}, \quad (2.43)$$

where \mathbf{I} is the identity matrix, $\boldsymbol{\Lambda}$ gives the eigenfrequencies $\lambda = \omega^2$ and \mathbf{A} is the eigenvector orthogonal matrix, which ensures that \mathbf{K} is diagonalized. These eigenmodes are often called *normal modes* of the system and are related to the coordinates $\boldsymbol{\xi}$ by the relation $\mathbf{Q} = \mathbf{A}^\dagger \boldsymbol{\xi}$. In the new normal coordinates, \mathbf{Q} , the potential and kinetic energies are expressed by

$$\mathcal{V}_k(\mathbf{Q}) = \frac{1}{2} \sum_{k=1}^N \omega_k^2 Q_k^2, \quad (2.44)$$

$$\mathcal{T}_k(\mathbf{Q}) = \frac{1}{2} \sum_{k=1}^N \dot{Q}_k^2. \quad (2.45)$$

2.5 Dynamics of Biomolecular Association

2.5.1 Bimolecular Association

In the cellular environment, two biomolecules can bind by diffusing towards each other. Active transport processes may also contribute to binding but these will not be discussed here. Considering association as a two-step process, which entails an intermediate called 'diffusional encounter complex' and a final state of the bound complex, two ways of describing the reaction can be adopted. When the association rate is determined by the formation of an intermediate complex, this reaction is called diffusion-controlled, otherwise - reaction-controlled. In Section (2.3) the diffusion-controlled association rate was derived.

Diffusional Encounter complex

The characterization of the diffusional encounter complex is of high importance for protein and nucleic acid design studies aimed at altering the association kinetics. In the case of diffusion-controlled processes, formation of the encounter complex determines the bimolecular diffusional rate constant. The rate of diffusional association has an upper limit: a reaction between two molecules cannot be faster than their rate of collision. This limit is around $10^9 \text{ M}^{-1}\text{s}^{-1}$ for uniformly reactive spheres of typical macromolecular size [41] in aqueous solvent with no forces between them.

As a rule, a random collision of two molecules does not result in binding - a freely diffusing molecule X must come close to its binding patch on a target molecule Y in order to form a diffusional encounter complex. Geometrically, this complex can be viewed as an ensemble of configurations able to evolve to the bound state.

During a single encounter, the two molecules have time to undergo substantial rotational reorientation while remaining trapped in the vicinity of each other and undergoing multiple collisions. This effect is known as a "diffusive entrapment". A computational technique, called Brownian Dynamics (BD) (see Sec. 4.2), was used by Northrup *et al.* [45] to model two non-interacting spheres of the size of small proteins, which showed about 400 times larger association rate ($2 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$), attributed to the diffusive entrapment effect, than the rate calculated by a simple geometric correction of the Smoluchowski rate considering two contacts as the criterion for binding ($1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$). An association rate constant of about $10^6 \text{ M}^{-1}\text{s}^{-1}$ is typical of protein-protein pairs that bind without strong electrostatic interactions. Attractive electrostatic forces can lead to enhancement of the rates to values

very close to the Smoluchowski rate.

Bound Complex

Once the encounter complex has formed, the biomolecules must reorient with respect to each other to form a fully bound complex. They may also undergo changes in conformation and induced fit in order to achieve a bound complex. Within the complex, the biomolecules are held together by close-range noncovalent interactions such as salt bridges, hydrogen bonds and van der Waals interactions. These interactions depend on the chemical nature of the interacting groups of both molecules as well as on their spatial arrangement and can be mediated by individual water molecules. A biomolecule can have one or several binding sites stabilizing the complex. A subtle change in the binding site can change the binding mode significantly. Therefore, biological associations are dependent on the structure of both molecules and can be highly specific.

2.5.2 Molecular interaction

The interactions between biomolecules vary in strength, type and source. Therefore, a wide spectrum of different forces contribute to complex formation [46]. Here, we will discuss only the electrostatic, hydrodynamic and hydrophobic interactions since their contribution to the kinetics of bimolecular association is shown to be considerable. We describe well-established methods for modelling these interactions.

Electrostatics

Electrostatic interactions are important for bimolecular association because they are relatively long range interactions and may therefore guide the association process by means of attractive and repulsive interactions. Their importance is shown by the dependence of association rate on ionic strength and the generally much greater influence on the association rate of mutations of charged than of neutral amino acids.

The biological entities in the cell are surrounded by ions, which screen the electrostatic interactions between the species. One way to account for the ions is to compute the molecular electrostatic potential $\varphi(\mathbf{r})$ using the nonlinear Poisson-Boltzmann equation

$$-\nabla \cdot \epsilon(\mathbf{r})\nabla\varphi(\mathbf{r}) = \rho(\mathbf{r}) + \sum_i q_i n_i \exp\left(-\frac{q_i\varphi(\mathbf{r})}{k_B T}\right), \quad (2.46)$$

where $\epsilon(\mathbf{r})$ is the position dependent dielectric permittivity, $\rho(\mathbf{r})$ is the molecular charge density, q_i and n_i are the charge and the concentration of the i -th ionic species in the bulk, respectively. The above equation can be approximated by the linear Poisson-Boltzmann equation if the exponential is expanded as a Taylor series

$$-\nabla \cdot \epsilon(\mathbf{r})\nabla\varphi(\mathbf{r}) + \epsilon\kappa^2\varphi = \rho(\mathbf{r}), \quad (2.47)$$

where κ is the Debye-Hückel screening length related to the ionic strength I by

$$\kappa^2 = \frac{2Ie^2}{k_B T \epsilon} \quad I = \sum_i \frac{1}{2} \frac{q_i^2}{e^2} n_i. \quad (2.48)$$

Equation (2.46) and Eq. (2.47) are used in studies of interactions between macromolecules in continuous media, i.e. media for which water molecules and ions are not modelled explicitly.

When two molecules approach each other in an aqueous solvent, an electrostatic desolvation effect arises due to the lower dielectric constant of the solute compared to that of the solvent. Charges located on the bimolecular complex interface become desolvated upon complex formation and this results in unfavourable electrostatic energy changes. This desolvation effect becomes significant at short distances and is mainly dependent on the location and magnitude of the charged groups.

Hydrodynamics

Hydrodynamic interactions between molecules are caused by the flow of solvent due to their mutual motions. Depending on the structure and shape of the molecules, these interactions can be either attractive or repulsive. To model them, one usually represents the diffusion coefficient as a tensor describing the properties of the solute and the media [47]. Another representation uses a mean field approach for hydrodynamic interactions with the diffusion coefficient depending on a local volume fraction of a system mimicking macromolecular crowding [48].

Hydrophobicity

The hydrophobic effect results in favourable interactions between two macromolecules in aqueous solvent. The reason for this is that the nonpolar groups on the surface of the molecules can avoid interaction with polar groups or molecules (water) by forming a complex. This is caused by the fact that the water molecules are orientationally restricted by the presence of the nonpolar species, and this leads to an entropy decrease. Furthermore, the presence

2.5. DYNAMICS OF BIOMOLECULAR ASSOCIATION

of the nonpolar species affects the ability of the water molecules to make hydrogen bonds. The hydrophobic force is temperature dependent and it is entropy driven only at low to room temperatures. At higher temperatures, the enthalpic contributions become significant [49, 50]. Although, a generalized theory describing the hydrophobic forces does not exist, hydrophobic interactions can be modeled by incorporating a solvent accessible surface area term in the Gibb's free energy [51] or using a hydrophobic potential of mean force [52].

3

Chromatin Fiber

3.1 Structure and biological function

3.1.1 Chromatin

The DNA, the information code carrier, is situated within the cell nucleus in eukaryotic organisms¹. The organization of the genome in the cell is governed to a large extent by the interplay of mechanical deformability of DNA with its structure at different length scales (see Fig. 3.1) [3, 53]. The double helix DNA molecule wraps around four dimeric proteins (H2A, H2B, H3, H4) called *histones* and forms a structure named *nucleosome* (Fig. 3.1). The histones and nucleosome are also referred in the literature as *histone core* and *nucleosome core particle* (NCP), respectively. The "bead-chain" structure of nucleosomes spaced regularly on DNA will compact into a higher order structure under physiological conditions leading to a fiber-like structure with a diameter of approximately 30 nm, the so-called '30-nm chromatin fiber' [54], which plays an important role in the biomolecular machinery [55]. However, its structure and dynamics are still obscure and matter of continuous experimental and theoretical investigations. There exist two structural models of chromatin fiber, the one-start (solenoid) [56] and the two-start (zig-zag) [4] chromatin structures. In the one-start model the successive nucleosomes are connected via bent DNA and, thus, form a solenoid-like structure. The two-start helix model corresponds to the crystal structure of the tetranucleosome (four nucleosomes) (Fig. 3.2), where the neighbouring nucleosome are linked by a straight DNA in a zig-zag fashion. In this way, for the nucleosome numbered n , the closest neighbour is with a number $n \pm 2$. In addition to the histone core, it is also known that there exists another protein belonging to the histone family [1, 2, 57] called *linker histone* (coloured green in Fig. 3.1). It is usually denoted as H5 or H1 and it contributes to the compaction of

¹The eukaryotes have a nucleus in the cell.

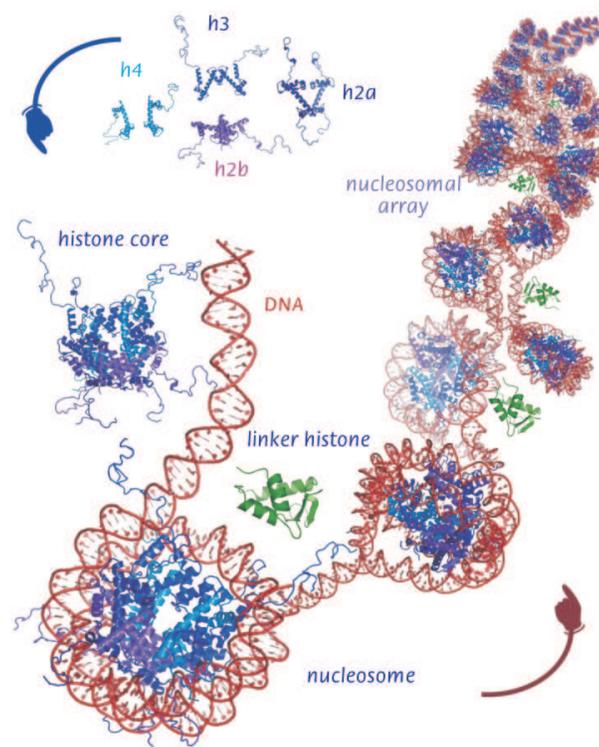


Figure 3.1: Schematic structure of chromatin fiber.

DNA in chromatin fiber. Since DNA is highly negatively charged molecule due to the phosphate groups on the backbone, it is hard to achieve such compaction without the presence of positively charged species. All the histones carry net positive charge, which might contribute to the higher-order structure of chromatin. Monovalent, divalent ions and histone tails have also been considered to contribute to chromatin compaction [58]. The mechanism of chromatin remodelling, opening and closing of the structure during transcription, and many other biological processes related to the higher order structure of the genome cannot be understood without fundamental knowledge of the arrangement of nucleosomes and DNA in the chromatin fiber and its variations during different physiological states of the cell.

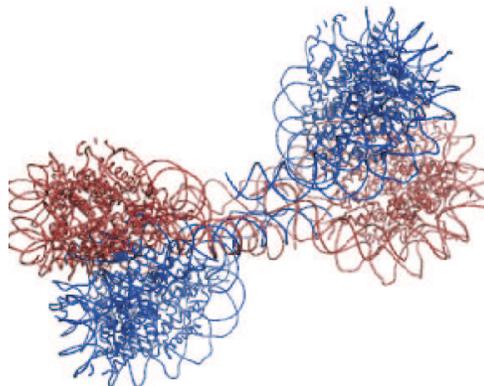


Figure 3.2: X-ray crystal structure of the tetranucleosome resolved at 9 Å resolution [4].

3.1.2 Nucleosome

The determination of the nucleosome structure to a resolution of 1.9 Å [5] has greatly contributed to our understanding of possible mechanisms of DNA-histone binding, nucleosome repositioning and the formation of higher-order chromatin structures. The DNA wrapped around the histone core consists of 147 base pairs (bp)² and varies from 150 to 250 bp [59] depending how far away the next nucleosome is. The existence of different chromatin fiber models implies that the DNA variation is not only in length, but also in conformation. Each histone in the histone core has a long, mobile and basic N-terminal tail extending out of the main body of the nucleosome. Particularly, the H3 and H4 tails are located at the ends of DNA (linker DNAs) on the nucleosome. The nucleosome has been a subject of extensive theoretical and computational modelling trying to shed light on its structure [60–62], dynamics [62–65] and energetics [66–68] and relate them to the formation of chromatin fiber. It has been modelled as a cylinder, sphere, ellipsoid, etc. with a radius of around 50-60 Å and with different DNA lengths. However, only recently, it became evident that the linker histone is one of the major determinants for folding of chromatin fiber [69].

²Base pair is the term referring to two connected nucleotides (A, G, T, C) in the double helix DNA (1 bp = 3.4 Å along the DNA strand).

3.1.3 Linker histone and its role in chromatin

The importance of the linker histone-nucleosome interactions in the biological function of DNA such as its transcription and replication, has already been shown [70]. However, the lack of knowledge how the linker histone binds to the nucleosome and where exactly makes it difficult to understand the proper structure and function of chromatin fiber. Experimental studies [32, 71] have shown that the linker histone forms a stem structure with the nucleosome and binds between the linker DNAs close to a point on the nucleosomal DNA called *dyad point*. Since the crystal structure of the chromatosome particle is still not resolved, a variety of speculative and controversial ideas about the complex are proposed in the literature. Some of them will be discussed in detail in the next chapters.

Globular domain of linker histone H5 (GH5)

Each protein is built of amino acids (residues) which are 20 in nature. Depending on their sequence, different protein structures can be formed. The X-ray crystal structure of the globular domain of the linker histone H5 (GH5)³ has been obtained [72] (Fig. 3.3). It has three α helices, two β strands and three loop regions consisting of 74 residues (see Fig. 3.3). On the surface of the GH5 there are 12 residues (Lys and Arg), which are positively charged and contribute to the overall positive potential of the linker histone. In addition, there are 10 hydrophobic residues (Val, Ala, Leu) positioned on the surface as well. Hydrophobic residues are electrostatically neutral and they contribute to protein stability mainly by forming contacts with each other in the protein interior. The structure of the linker histones, in addition to the globular domain, is also composed of a long and basic C-terminal domain (more than 100 residues) and a short N-terminal tail [73]. It has been shown that the C-terminal domain is crucial for H1-linker DNA binding and for stabilizing chromatin fiber [74, 75]. However, this domain is unstructured at physiological ionic strengths and therefore difficult to crystallize. The C-terminal domain will not be used in this study, but possible implications on the GH5-nucleosome interactions will be discussed.

Linker histone models

The contribution of the linker histone to the nucleosome-nucleosome interactions as well as to the higher order folding of chromatin fiber has been investigated theoretically [22, 24, 28–30, 58, 76–80] and experimentally [6,

³The resolved X-ray crystal structure is a dimer, here only chain B is used

3.1. STRUCTURE AND BIOLOGICAL FUNCTION

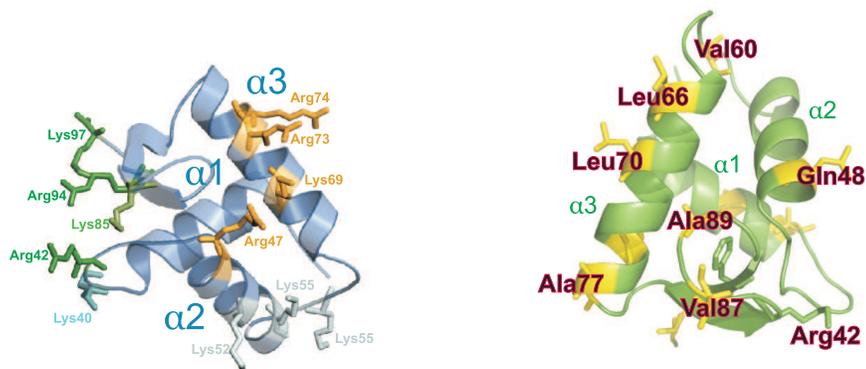


Figure 3.3: X-ray crystal structure of the GH5 at 2.3 Å resolution [72]. The positive (left) and hydrophobic (right) residues on the surface of the GH5 are depicted. The labels α_1 , α_2 and α_3 denote the alpha helices.

8, 17, 18, 21, 27, 31, 32, 56, 58, 71, 81–96]. Most of the experimental work points to a formation of a linker histone-nucleosome complex with a linker histone located between the linker DNAs of the nucleosome. It is believed that the globular domain of the linker histone contacts directly the nucleosomal DNA, while the C-terminal domain acts as a bridge between the linker DNAs [70]. The exact position of the GH5 on the nucleosome has not been determined experimentally, but atomic detail theoretical models have attempted to resolve this issue [28, 29]. However, the studies are not consistent with each other, proposing from two to three binding sites on the GH5, different orientations of the protein with respect to the nucleosome, different residues participating in the binding process as well as a unclear role of the linker histone on the DNA accessibility. Next chapters will address these issues and the results obtained will be related and compared to the already existing models.

4

Computational Methods for Studying Biomolecular Dynamics and Interactions

Biological systems consist of a huge number of interacting atoms and in order to describe their dynamics accurately one has to solve the equations of motion analytically taking into account the individual nature of each atom. Such a calculation is not an easy task and in certain cases it is impossible to be carried out. However, with the development of the computer power it became evident that the equations of motion for many particle system can be solved instantaneously on a computer. A method called *Molecular Dynamics* (Sec. 4.3) integrates numerically the equations of motion for all atoms and, thus, the phase space trajectory of a system can be followed for a certain time. This time is highly dependent on the system size and sometimes it is not sufficient to describe interesting phenomena occurring on longer time scales. On the other hand, another method called *Brownian Dynamics* (Sec. 4.2) can reach much longer time scales, because it does not take into account the motions of the solvent particles, but only the motions of the solute object under interest. In addition, *Normal Mode Analysis* is a method assessing the molecular dynamics by representing the molecule as a set of coupled harmonic oscillators, each of them describing an independent normal mode.

Consequently, the computational methods serve as a link between physical theory and experiment and their efficiency is dependent on the available computer power and on the model employed. In a perfect situation one would just describe quantum mechanically even extremely large systems and then perform computational simulations to follow their dynamics. Even within the recent computer advances such simulations would need hundreds of years computational time. Therefore, most of the computational methods are approximate to a certain extent.

4.1 Normal Mode Analysis

Normal Mode Analysis is a technique which is used for studying the slowest motions in dynamical systems. It is based on the theory of small fluctuations like the lattice vibrations in solid state physics and was already introduced in the Sec. 2.4.

4.1.1 Elastic Network Model

The normal modes (eigenvectors) and their associated frequencies (eigenvalues) are obtained by the second derivative matrix (Hessian) (Eq. (2.37)) of the potential energy of a given structure around a local minimum. A model called Elastic Network Model (ENM) represents the energy landscape by a harmonic potential in a way that the atoms in the structure are connected via springs within a certain cutoff C [34]. Then, the potential energy is given by

$$\mathcal{V}(\mathbf{r}_{ij}) = \sum_{\text{all pairs } i,j} k(\mathbf{r}_{ij}^0) (|\mathbf{r}_{ij}| - \mathbf{r}_{ij}^0)^2, \quad (4.1)$$

where \mathbf{r}_{ij}^0 is the pair distance between atom i and j in the equilibrium conformation. The force constant $k(\mathbf{r}_{ij}^0)$ is modelled to decrease exponentially with the distance [97]

$$k(\mathbf{r}_{ij}^0) = k_{\text{in}} \exp\left(-\frac{|\mathbf{r}_{ij}^0|}{d_0}\right), \quad (4.2)$$

where d_0 is a fixed parameter in \AA and k_{in} is the input force constant.

Usually the atomic Cartesian coordinates and the Hessian matrix are mass-weighted before the eigenvalue problem is applied (see Eq. 2.42). The normal mode vectors \mathbf{Q} are related to the temperature by their amplitude ϑ as

$$\vartheta^2 \omega^2 = 2k_B T. \quad (4.3)$$

In case of a biological macromolecule with N atoms, the Hessian matrix will be $3N \times 3N$ matrix, i.e. there will be $3N$ normal modes (vectors). They represent only a direction of movement rather than absolute value for the displacement. The low normal modes with small ω describe large-amplitude motions and they are of particular interest in protein and nucleic acid dynamics. The advantage of using the ENM is that for short computational time it can give a reasonable sampling of biomolecular dynamics on the configuration space. However, the ENM does not take into account solvent damping and anharmonic effects, which might be crucial for the system's dynamics.

4.2 Brownian Dynamics

In the Brownian Dynamics (BD) method discussed in this study, the diffusional motion is treated with a particle-based approach, where the solute molecules are placed in a continuum medium describing aqueous environment. The solutes are represented by all-atom structures keeping only six degrees of freedom - translation and rotation.

4.2.1 Ermak-McCammon algorithm

A method for modelling the interaction and dynamics between Brownian particles was proposed by Ermak and McCammon [47]. Instead of integrating Newton's or Hamilton's equations for the system, Ermak and McCammon used the Langevin equation (Eq. 2.19) for the Brownian particles assuming that the time scales for the momentum and position relaxations are well separated [47]. The diffusive displacement \mathbf{r} after a certain time step Δt^1 was derived² and the translational component reads

$$\mathbf{r} = \mathbf{r}_0 + \frac{D}{k_B T} \mathbf{F}(\mathbf{r}_0) \Delta t + \mathbf{R}(\Delta t), \quad (4.4)$$

where \mathbf{r}_0 is the position before the time step Δt , $\mathbf{F}(\mathbf{r}_0)$ is the systematic force acting on the molecule and $\mathbf{R}(\Delta t)$ is a random displacement with a Gaussian distribution such that $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R}^2 \rangle = 6D\Delta t$. The rotational displacement $\boldsymbol{\theta}$ is computed as

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{D_R}{k_B T} \mathbf{T}(\boldsymbol{\theta}_0) \Delta t + \boldsymbol{\Theta}(\Delta t), \quad (4.5)$$

where $\boldsymbol{\theta}_0$ is the rotation angle before Δt , $\mathbf{T}(\boldsymbol{\theta}_0)$ is the torque and $\boldsymbol{\Theta}(\Delta t)$ is a random rotational angle satisfying $\langle \boldsymbol{\Theta} \rangle = 0$ and $\langle \boldsymbol{\Theta}^2 \rangle = 6D_R\Delta t$. Equation 4.4 and Eq. 4.5 imply that the Brownian particles have to be described as rigid bodies rather than objects with flexible units. In case of biomolecules, this reduction of degrees of freedom 'freezes' global structural changes, as side chains movements, accompanying the diffusional motion. All-atom flexibility is not taken into account in the Ermak-McCammon BD algorithm.

¹ Δt should fulfil $\Delta t \gg mD/k_B T$

²In the original derivation the diffusion D and the friction ζ have been treated in a tensor form due to the hydrodynamic interactions. Here, it is assumed that both are constants. For a complete derivation see Ref. [47].

4.2.2 Simulation of diffusional association

When two biomolecules diffuse freely in an ambient medium they can attract each other and form an encounter complex (Sec. 2.5.1). The stability of this favourable state will depend on the interaction energy between both molecules. In addition, the formation of the diffusional encounter complex will be also governed by the kinetics of both approaching molecules. Brownian dynamics simulations can reveal these aspects of biomolecular diffusion. A program called *Simulation of Diffusional Association* (SDA) [98] has been developed in order to address these issues.

Encounter complexes

Structural and energetic knowledge of the diffusional encounter complexes can be very important for finding out how the biomolecules function and where their binding patches are. A version of SDA, called *SDAC*, has been developed for these purposes. It is based on a docking procedure between the biomolecules under interest. Using the Ermak-McCammon algorithm, one of the molecules (molecule 2) is placed on a surface b (see Fig. 4.1) around the other molecule (molecule 1) and a trajectory is initiated. At this distance b the forces between them are negligible or centrosymmetric. Each trajectory is started from a randomly chosen position and orientation on the ' b surface'. The trajectory is truncated when the molecule 2 reaches a separation c (at the ' c surface'). The translational diffusion constant D in Eq. (4.4) is the relative diffusion constant $D = D_1 + D_2$ for both particles (denoted 1 and 2). It is assumed to be spatially independent and D and D_R are given by the Stokes-Einstein relation (see Eq. 2.13). An encounter complex is recorded if the center-to-center distance between both biomolecules satisfies predefined constraints. These constraints can be based on any experimental or biochemical data. In addition to the structural information, the interaction energy is recorded as well. From the ensemble of complexes many features about the mutual interactions and binding modes can be extracted as well as possible bound complexes predicted. The two binding partners can be modeled in atomic detail using, for example, X-ray resolved crystal structures (from Protein Data Bank³ (PDB)).

Association rates

The biomacromolecular complexes have different properties depending on the molecules involved. Some can bind quickly while others bind slowly and

³URL: <http://www.rcsb.org>

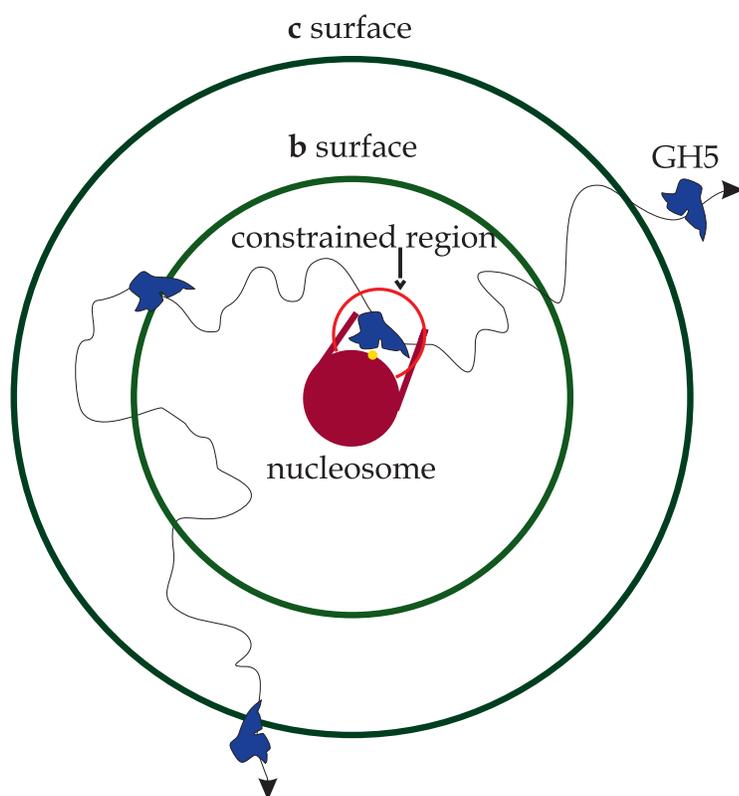


Figure 4.1: Schematic picture of the system simulated by Brownian Dynamics. The trajectory starts at the *b* surface and a complex is recorded if the second molecule enters the constrained region. The trajectory is truncated when molecule 2 (GH5) passes the *c* surface.

weakly. The most commonly used formalism for calculating bimolecular association rate constants is known as the Northrup-Alison-McCammon (NAM) method [99]. It is assumed that the rate constant for the approach of two molecules to a separation b at which the forces between them are negligible or centrosymmetric is given by the analytical Smoluchowski expression [41] (Eq. 2.32), with $r = b$

$$k_{\text{on}}(b) = 4\pi bD, \quad (4.6)$$

$$k_{\text{on}}(b) = \frac{4\pi D}{\int_b^\infty \frac{e^{U(r)/kT}}{r^2} dr}. \quad (4.7)$$

Equation 4.6 and Eq. 4.7 yield the association rate constant for noninteracting and interacting molecules, respectively. By generating thousands of trajectories and monitoring those that fulfil criteria for forming an encounter complex, the probability of reaction, β , can be obtained, and thus, the association rate constant $k_{\text{on}}^{\text{BD}}$ calculated as

$$k_{\text{on}}^{\text{BD}} = k_{\text{on}}(b) \frac{\beta}{1 - (1 - \beta) \frac{k_{\text{on}}(b)}{k_{\text{on}}(c)}}. \quad (4.8)$$

In the equation above the $k_{\text{on}}(b)$ is given by Eq. 4.6 when the interactions at distance b are negligible. The trajectories are handled as described in Sec. *Encounter Complexes*. The quantification of the encounter complex, which is crucial for determining the association rates, is of major importance. In SDA, three different models (labelled 1, 2 and 3) defining a reaction criterion are implemented, but here only the one used in the simulations (3) will be discussed. The models are based on monitoring the polar atom-atom distances between the biomolecules during simulation within initially specified distance windows. These donor-acceptor contacts are any pair of atoms formed in the experimentally determined bound complex of both molecules. The pairs taken into account are only those separated by a specific distance d^{pairs} (between an atom on molecule 1 to an atom on molecule 2). Usually this distance is $d^{\text{pairs}} \sim 5 - 6 \text{ \AA}$. A list of the potential pairs is created and in model 3 the number of possible contacts is restricted to the pairs separated by a distance d^{ind} . This means that if any two atoms on the same molecule form two pairs with other two atoms on the other molecule, then these contacts are counted as two if the distance between the former atoms is greater than d^{ind} , i.e. then the pairs are independent. Otherwise, only one contact is assumed. In such a way only the independent contacts are monitored during simulation. Normally, the independent distance is chosen to be around 6 Å depending on the size and structure of both molecules. In the output the

association rates for forming 1, 2, 3 and 4 contacts are recorded for each window distance. However, a trajectory can be counted as 'reactive' if at least two of the native polar contacts of the fully bound complex are formed. These contacts are looser than in the bound complex since the complex is partially desolvated and this is considered by the distance d^{pairs} . The procedure described has been implemented in the program (SDA) [98], which is constantly under development.

Forces

Currently, the SDA program uses electrostatic, electrostatic desolvation, hydrophobic desolvation and exclusion forces (torques) for modelling bimolecular interactions. Among them only the electrostatic forces are long-ranged and therefore they are of high importance for the formation of diffusional encounter complex. The force is given by

$$\mathbf{F}(\mathbf{r}) = -\frac{\partial(\Delta G^{\text{total}})}{\partial \mathbf{r}}, \quad (4.9)$$

where ΔG^{total} is the total interaction energy between the molecules and is expressed by

$$\Delta G^{\text{total}} = \Delta G^{\text{elec}} + \Delta G^{\text{eldesolv}} + \Delta G^{\text{hyddesolv}}. \quad (4.10)$$

The first term ΔG^{elec} in Eq. 4.10 is the electrostatic interaction energy, the second term $\Delta G^{\text{eldesolv}}$ gives the electrostatic desolvation interaction energy and the third term $\Delta G^{\text{hyddesolv}}$ describes the hydrophobic interaction between the biomolecules. The exact form of the electrostatic terms, which play an important role in the nucleosome-linker histone interactions, will be given below. All the forces are modelled on a 3-dimensional grid, which encompasses both biomolecules, and at each Brownian step the force between the molecules is computed as a finite-difference derivative of the interaction energy given in Eq. 4.10. The size of the grids depend on the size of the molecules as well as on the goal of the simulation.

Electrostatic forces The long-range electrostatic forces are computed by solving numerically the Poisson-Boltzmann equation (see Eq.2.46) on a grid. There exist several numerical methods for solving this equation on a cubic 3D grid like the finite difference and multigrid methods [100]. The PB equation is solved after defining a dielectric boundary between the solute and the solvent and choosing a suitable dielectric constants of the biomolecule and solvent. The temperature is set and the ionic strength of the solvent is

modelled implicitly assuming a Boltzmann distribution of ions. In the calculation the partial atomic charges of the biomolecule are used. Two software packages were used for performing the calculations, the Adaptive Poisson Boltzmann Solver (APBS) [101] and the University of Houston Brownian Dynamics (UHBD) [102]. At each grid point on the 3D cubic lattice mapped on the biomolecule a value of the Poisson-Boltzmann electrostatic potential $\varphi(\mathbf{r})$ is assigned (see Fig. A.1). In this thesis the electrostatic potential was computed using the nonlinear PB equation (Eq. 2.46).

Effective charges for macromolecules In the BD algorithm, the forces are computed at each step, but evaluation of the electrostatic intermolecular potentials required for the forces, is computationally expensive. The reason is the huge number of atomic charges on the biomolecules, which might restrict the dynamics to several steps. In simulating long diffusional process this is not beneficial. This problem has been overcome by representing the second protein by a set of effective charges in a uniform medium [103]. The idea is as follows: point atomic charges inside a molecule are imaged through its dielectric boundary and, thus, image charges are created. These image charges are responsible for the electrostatic potential created inside the molecule and, in opposite, the external electrostatic potential can be determined by the image charges placed inside the molecule. In this way the effective charges are these image charges creating the external electrostatic potential. The electrostatic potentials $\varphi_1(\mathbf{r})$ and $\varphi_2(\mathbf{r})$ of molecule 1 and 2, respectively, can be solved by the Poisson-Boltzmann equation (Eq. 2.46) and the electrostatic interaction energy reads

$$\Delta G^{\text{elec}} = \frac{1}{2} \sum_i q_{i1}^{\text{eff}}(\mathbf{r}) \varphi_2(\mathbf{r}) + \frac{1}{2} \sum_j q_{j2}^{\text{eff}}(\mathbf{r}) \varphi_1(\mathbf{r}), \quad (4.11)$$

where $q_1^{\text{eff}}(\mathbf{r})$ and $q_2^{\text{eff}}(\mathbf{r})$ are the effective charges fitted to the electrostatic potentials $\varphi_1(\mathbf{r})$ and $\varphi_2(\mathbf{r})$ of molecule 1 and 2, respectively. The summation in Eq. 4.11 is carried out over all effective charges in the molecules. The fitting procedure is performed in a shell within a specified distance around the molecule [103]. In SDA4C only the second term multiplied by two in Eq. 4.11 is considered, i.e. the effective charges of molecule 2, $q_2^{\text{eff}}(\mathbf{r})$, move on a three dimensional grid of the electrostatic potential of molecule 1, $\varphi_1(\mathbf{r})$. This consideration is accurate as long as cavity penalties at short distances for both biomolecules are not taken into account. Then two additional terms come into the interaction energy and an approximation of such interaction will be discussed in the next paragraph.

The effective charges are placed on the carboxylate oxygen atoms of Asp, Glu and C-terminus, on the amino nitrogen atoms of Lys, Arg and

N-terminus in the histone proteins as well as on the P atoms of DNA backbone in the nucleosome.

Electrostatic desolvation forces. Unfavourable electrostatic desolvation effects occur at short distances due to the displacement of a water layer in close proximity to the surfaces of the biomolecules. The electrostatic desolvation interaction energy is computed using an approximation given by [104]

$$\Delta G^{\text{eldesolv}} = \sum_i [q_{i1}^{\text{eff}}(\mathbf{r})]^2 \varphi_2^{\text{eldesolv}}(\mathbf{r}) + \sum_j [q_{j2}^{\text{eff}}(\mathbf{r})]^2 \varphi_1^{\text{eldesolv}}(\mathbf{r}), \quad (4.12)$$

$$\begin{aligned} \Delta G_1^{\text{eldesolv}} = \sum_i [q_{i1}^{\text{eff}}(\mathbf{r})]^2 \varphi_2^{\text{eldesolv}}(\mathbf{r}) &= \alpha \frac{\varepsilon_s - \varepsilon_m}{\varepsilon_s(2\varepsilon_s + \varepsilon_m)} \sum_{ij} [q_{i1}^{\text{eff}}(\mathbf{r})]^2 \times \\ &\times (1 + \kappa r_{ij})^2 \exp(-2\kappa r_{ij}) \frac{a_j^3}{r_{ij}^4}, \end{aligned} \quad (4.13)$$

where in Eq. 4.13 ε_s and ε_m are the solvent's and solute molecule's dielectric constants, respectively, $q_{i1}^{\text{eff}}(\mathbf{r})$ is the i -th effective charge on molecule 1, a_j is the dielectric cavity radius of j -th atom on molecule 2 and r_{ij} is the distance between them. The Debye-Hückel parameter is κ , α is a scaling factor and the second term in Eq. 4.12 is computed as $\Delta G_1^{\text{eldesolv}}$ (Eq. 4.13) for all pairs of effective charges on molecule 2, $q_{j2}^{\text{eff}}(\mathbf{r})$, and the corresponding electrostatic desolvation potential grid points for molecule 1, $\varphi_1^{\text{eldesolv}}(\mathbf{r})$. The electrostatic desolvation potentials were also computed on a 3D cubic lattice encompassing the crystal structures of the biomolecules.

In this thesis, electrostatic desolvation was not used in the BD docking simulations, because we aimed at obtaining a tight diffusional encounter complex between the nucleosome and the linker histone. However, for the calculation of the association rates the usage of $\Delta G^{\text{eldesolv}}$ was necessary. Details on the computation of the electrostatic forces are given in Appendix A.

Hydrophobic desolvation forces. These forces are short range favourable forces caused by the attractive interactions between nonpolar surfaces. Simulations with these forces were not carried out due to the much stronger electrostatic attractive interaction in the linker histone-nucleosome recognition process. Details on the potential used have been already reported [105].

Exclusion forces. To avoid overlaps between both molecules an exclusion grid is assigned on the first molecule based on the van der Waals atom radii.

If a step results in an overlap then the step is repeated with a different random number.

The software package SDA has two versions, 4 and 5. In version 4, named SDA4, hydrophobic desolvation interactions are not included as well as a slightly different treatment of the rotation-translational procedure between the SDA4 and SDA5 exists. In addition, in the newer version SDA5 a better random number generator is implemented and the electrostatic interactions are computed using the grids and charges of both biomolecules. The simulations in the next two chapters were carried out with the SDA4 while the SDA5 was used for the others.

4.3 Molecular Dynamics

4.3.1 Force fields

Molecular Dynamics (MD) is a computational technique based on classical physics, where the force \mathbf{F}_i acting on each atom i due to the presence of all atoms in the system is computed and the atom i is put into motion via the Newton's equation of motion

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i = -\nabla_{\mathbf{r}_i} \mathcal{V}(\mathbf{r}), \quad (4.14)$$

where $\mathcal{V}(\mathbf{r})$ is the total potential energy of the system. Here, the Born-Oppenheimer approximation is assumed allowing to express the energy as a function of the atomic (or nuclear) coordinates only. Often such a potential is based upon a simple model of interactions within the system and is named a *force field*. Force fields usually include intra- and inter-molecular contributions and a common functional form is

$$\begin{aligned} \mathcal{V}(\mathbf{r}) &= \mathcal{V}^{\text{bond}}(\mathbf{r}) + \mathcal{V}^{\text{angle}}(\mathbf{r}) + \mathcal{V}^{\text{dihedral}}(\mathbf{r}) + \mathcal{V}^{\text{vdW}}(\mathbf{r}) + \mathcal{V}^{\text{Coulomb}}(\mathbf{r}) = \\ &= \sum_{\text{bonds } i} k_i^{\text{bond}} (r_i - r_i^0)^2 + \sum_{\text{angles } i} k_i^{\text{angle}} (\theta_i - \theta_i^0)^2 + \\ &+ \sum_{\text{dihedrals } i} k_i^{\text{dihedral}} [1 + \cos(n_i \phi - \gamma_i)] + \\ &+ \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \\ &+ \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}}, \end{aligned} \quad (4.15)$$

where the first three terms describe the stretching, bending and torsional bonded interactions, the fourth term is the Lennard-Jones potential and the

last term describes the electrostatic interactions. The internal coordinate parameters (e.g. k^{bond} , r^0) as well as those for the nonbonded parts of the system (see Eq. (4.15)) are usually developed from data on small molecules and tested on a small number of systems. Therefore, force fields are *empirical* and their key feature lies in their transferability, that is, once parametrized, a force field can be applied to a wide range of problems and systems.

In this study the *Amber99SB* force field [106] was used within the MD software package NAMD [107] (version *2.7b1*). NAMD is particularly designed for running MD simulations of large biomolecular systems, like the nucleosome particle, since it is based on an efficient parallel scaling implementation. Here the methodology of performing our MD simulation in the context of NAMD will be described, although the overall procedure and protocol are similar to many biomolecular MD software packages.

For accurate treatment of the non-bonded electrostatic interactions as well as to avoid boundary surface effects, periodic boundary conditions are employed. These are also required for constant pressure simulations since pressure is controlled by adjusting the volume of the system at different timesteps (description in Sec. 4.3.3). The solute system is initially solvated in water in a box of a chosen shape (cubic, octahedral, rhombic dodecahedral, etc.) and a certain number of ions are added to satisfy both concentration and zero net charge requirements. The latter are needed for accurate treatment of the long-range electrostatic interactions through the Ewald summation scheme [108], which overcomes the slow convergence of the electrostatic potential energy at great distances and assumes each charge is surrounded by a neutralizing charge distribution of Gaussians. This allows the potential to be splitted into different contributions each of them converging more quickly:

$$\mathcal{V}^{\text{Ewald}} = \mathcal{V}^{\text{dir}} + \mathcal{V}^{\text{rec}} + \mathcal{V}^{\text{self}} + \mathcal{V}^{\text{surface}}, \quad (4.16)$$

where \mathcal{V}^{dir} and \mathcal{V}^{rec} are the real-space (or direct) and reciprocal⁴-space summations, respectively. The other two terms, $\mathcal{V}^{\text{self}}$ and $\mathcal{V}^{\text{surface}}$, represent the self-energy and surface energy, respectively. The former is usually a constant, whereas the latter is very small due to the large difference between the dielectric constants of the molecule interior and the solvent medium [107].

Written in the above form, the electrostatic potential is still computationally demanding and it scales as N^2 (or $N^{3/2}$ in the best scenario) for a system with N particles in the primary simulation cell. In order to reduce the scaling factor, the fast Fourier transform (FFT) algorithm is implemented to compute the most expensive reciprocal-space term \mathcal{V}^{rec} . The speed up is considerable - $N \log N$. The disadvantage, however, is that the FFT

⁴Term used mainly in solid state physics.

method requires discrete data points, which condition is achieved by replacing the continuous charges by a 3D grid-based charge distribution. Different so called "particle-mesh" Ewald (PME) approaches exist to map the charges of the system on a grid that correctly interpolates to the desired potential at each atom, and NAMD specifically uses the *smooth PME* method [109].

Van der Waals interactions are treated within a pre-defined non-bonded cutoff since they fall off very quickly with distance. When periodic boundary conditions are employed, the cutoff value should be chosen as to not allow an atom to interact with its own image. Together with the minimum image convention (stating that every atom interacts with a maximum one image of every other atom in the system), the cutoff should not be greater than half the dimension of the unit cell. To let the Van der Waals potential smoothly (discontinuously) go to zero beyond the cutoff, a smaller switching distance is usually appropriate to specify, where a switching function is applied between it and the cutoff distance.

4.3.2 Velocity *Verlet* algorithm

There exist several ways of carrying out a numerical integration on Eq. 4.14, but for all of them the differentials are replaced by the corresponding finite differences. Then the continuous phase trajectory $\Gamma(t)$ is replaced by a set of points $\Gamma(0)$, $\Gamma(\delta t)$, $\Gamma(2\delta t)$, \dots , $\Gamma(n\delta t)$, \dots , where δt is the time step of the simulation and n expresses the number of step, i.e. the number of times numerical integration is performed. First a Taylor expansion of the coordinates (Cartesian), velocities, accelerations, etc. around the current time is applied and second a certain algorithm (integrator) is chosen. One of the most common integrators is the *velocity-Verlet* method [110]

$$\begin{aligned}\mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{\delta t^2}{2} \mathbf{a}(t), \\ \mathbf{v}(t + \delta t) &= \mathbf{v}(t) + \frac{\delta t}{2} [\mathbf{a}(t) + \mathbf{a}(t + \delta t)],\end{aligned}\tag{4.17}$$

where $\mathbf{v}(t)$ are the velocities and the accelerations $\mathbf{a}(t)$ are given by

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i} \nabla_{\mathbf{r}_i} \mathcal{V}(\mathbf{r}(t)).\tag{4.18}$$

At the initial condition of the system, i.e. given Cartesian coordinates and force field parameters for all atoms, the accelerations are computed according

to Eq. 4.18. Then, the procedure follows as

$$\begin{aligned}
 \mathbf{r}_{n+1} &= \mathbf{r}_n + \delta t \mathbf{v}_n + \frac{\delta t^2}{2} \mathbf{a}_n, \\
 \mathbf{v}_{n+\frac{1}{2}} &= \mathbf{v}_n + \frac{\delta t}{2} \mathbf{a}_n, \\
 \mathbf{a}_{n+1} &= \mathbf{a}(\mathbf{r}_{n+1}), \\
 \mathbf{v}_{n+1} &= \mathbf{v}_{n+\frac{1}{2}} + \frac{\delta t}{2} \mathbf{a}_{n+1}, \\
 \dots &= \dots
 \end{aligned}
 \tag{4.19}$$

The choice of the time step δt in the MD has to be carefully considered. On one hand, it should be small enough to capture the fastest movements in the system like the highest-frequency vibrations of hydrogen bonds, which are on a timescale of 1 fs (femtosecond)⁵. On the other hand, the larger the time step, the less computational time will be required. One can restrict the usually of less interest high frequency motions like bond vibrations by imposing constraints to the bonds of the atoms linked to hydrogens. This allows increase in the time step δt by a factor of two (to 2 fs). The procedure for applying such constraints is called either SHAKE [111] or RATTLE [112], and we used the former.

4.3.3 Thermostats and Barostats

The Newton's equations of motion for an isolated system propagate a trajectory in the phase space along which $d - 1$ integral of motions are defined (d is the dimensionality of the phase space). Time homogeneity implies one of these integral of motion is the total energy of the system. Over the course of a very long (infinite) time, the dynamic trajectory will cover the whole phase space, i.e. each visited state will have the same energy. Such states form an ensemble of configurations (see Sec. 2.1.1) called the microcanonical ensemble described by constant number of particles, constant volume and constant energy (NVE). According to the ergodic hypothesis, the time average of a property of the system over the phase path will equate to the ensemble average over all configurations in the ensemble as well as to an experimentally measured value of the property. Therefore, a MD simulation evolving by the Newton's equations of motion for a sufficiently long time will produce a trajectory of configurations with time averages equivalent to the experimentally obtained averages. However, rarely do real (biomolecular) systems and reactions take place under experimental conditions of constant

⁵1 fs=10⁻¹⁵ s

volume and energy. In order to be able to computationally mimic experiments, modified equations of motions should be developed with integrals of motion matching experimentally invariant macroscopic variables such as temperature and pressure. The time evolution of such equations will sample configurations belonging to ensembles with constant temperature (canonical NVT) and/or pressure (Gibbs NPT). To keep either the temperature of the system or its pressure or both constant during an MD simulation run, different algorithms have been developed - with or without accurate sampling of the desired statistical ensemble.

NAMD controls the temperature of the system in several ways. Langevin equations of motions (Eq. 2.19) integrated by Brünger-Brooks-Karplus (BBK) integrator [107, 113] introduce a random force which acts as a heat bath. An alternative way is to couple the system to an external heat bath that is fixed at the desired temperature [114]. Other not so rigorous methods include the simple rescaling or reassignment of the velocities in the system, which are usually used during the initial heating and equilibration phases of the simulation. Out of these four temperature control schemes, only the Langevin equations of motion (if properly implemented) generate rigorous canonical averages. We used the Langevin heat bath during equilibration, whereas relied on the Berendsen coupling mechanism in the production phase of the simulation.

In case of the isothermal-isobaric ensemble (NPT), pressure can be maintained by Nosé-Hoover Langevin piston method [107, 115, 116] (also called the extended system method) or the Berendsen barostat [114] (also called the weak-coupling method). Since pressure is controlled by change in volume, the two algorithms introduce volume as an additional degree of freedom in the Newton's equations. In the former case, volume is being adjusted by a fictitious piston with a specified mass, where the motion of the piston is partially damped by the Langevin equation. The system is in contact with a heat bath, but since this contact is through a single degree of freedom (the piston) only, additional suitable temperature control method should be used in conjunction in order to sample the NPT ensemble. The weak-coupling method for pressure control is analogous to the thermostat coupling and indeed both have to be used together because the piston overdamped coupling leads to a dissipation of energy, which should be put back into the system through the temperature control. The Berendsen thermo-/barostat algorithms do not produce trajectories in a known ensemble. We used the Nosé-Hoover Langevin piston method for pressure control during equilibration and the Berendsen coupling schemes during the production.

5

Brownian Dynamics docking of the linker histone to the nucleosome

The key role of the linker histone (LH) proteins (H1/H5 families) in chromatin fiber formation is well established [69]. H5/H1 contributes not only to the compaction of chromatin into a 30 nm fiber, but also participates in the regulation of processes such as replication and transcription [117]. The existence of two proposed structures of chromatin, of the one-start (solenoid) [56] and the two-start (zig-zag) [4] (Fig. 3.2) helices, implies that the linker DNA (lDNA) connecting successive nucleosomes varies not only in length but also in conformation. It is known that the LH binds to the nucleosome, but exactly how the two interact, and how this interaction is affected by and itself affects the conformation and dynamics of the lDNA, is not yet understood.

From *in vivo* FRAP experiments, Brown et al. identified two binding sites and one nonbinding site on the globular domain of the H1 LH, GH1⁰, and modeled a complex of GH1⁰ with the nucleosome to fit this data [17]. Two binding sites were also suggested on the basis of *in vitro* photocrosslinking data for the H5 LH, GH5 [27], and molecular modelling of GH1 [29], but, in these studies, different binding modes to the nucleosome were deduced. On the other hand, two computational docking studies [28, 30] of the H5 LH globular domain, GH5, (which is 97% homologous to GH1⁰) to the nucleosome showed three binding sites on the GH5 and different docking positions with respect to the nucleosome. In light of these inconsistencies, the aim of this study is to determine the position and orientation of GH5 with respect to the nucleosome, and how binding of GH5 is influenced by and influences the DNA conformation and dynamics. These factors are crucial for the structure and function of the higher-order chromatin fiber. To achieve this, the globular domain of linker histone H5 was docked to the nucleosome using the BD software package SDA4 (Sec. 4.2.2).

5.1 System preparation

The crystal structure of the nucleosome core particle [5] (NCP, Protein Data Bank - PDB code 1kx5, 1.9 Å resolution) was used as a reference structure. The histone tails were removed from the structure because they are much more mobile than other parts of the nucleosome and a recent experimental study [31] showed that histone tail removal does not affect binding of the H5 linker histone significantly. In addition, another experimental study [95] pointed out that the position of the linker histone H1⁰ on the nucleosome is not affected by the histone tails. To include the linker DNAs, 20 bp of DNA from the tetranucleosome structure [4] (see Fig. 3.2), 10 bp at each entry/exit, were added to the reference structure and the nucleosome structure obtained (tNCP) was used as an equilibrium conformation in the Normal Mode Analysis (NMA). The conformation of the initial nucleosome structure (tNCP) is depicted in Fig. 5.1. The globular domain of the H5 linker histone

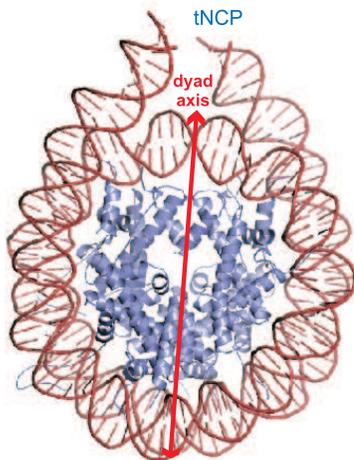


Figure 5.1: Nucleosome structure (tNCP) used in the NMA. The red arrow designates the dyad axis.

(GH5) was obtained from its crystal structure [72] (GH5, PDB code 1hst, 2.5 Å resolution) (see Fig. 3.3). Chain B of GH5 was used for the simulations.

5.2 NMA

The obtained nucleosome structure (tNCP) was subjected to Normal Mode Analysis based on the Elastic Network Model (ENM) [34] (Sec. 4.1.1). The harmonic approximation was built around the crystal structure itself, which was assumed to be in a global minimum. Such an approximation avoids the usage of computationally expensive minimization procedure, but has the disadvantage of having the initial structure trapped either in a local minimum

or far away from any minimum. In such a way the NMA might describe in detail the dynamics around a state, which has a low density distribution in the phase space $\Gamma(t)$. Moreover, the NMA dynamics does not take into account solvent damping and anharmonic effects, which might be crucial for the biomolecular function. Surprisingly, the large amplitude fluctuations in lysozyme obtained by MD yielded similar directions with the low-frequency normal modes [118]. Other studies [97, 119, 120] also showed qualitatively that the low-frequency normal modes characterize well the collective domain motions in proteins. In chromatin, the highly packed DNA molecule should partially or completely unwrap from the nucleosome in order to perform its function in the cell nucleus [121, 122]. In addition, the way it is packed tunes its accessibility to other proteins and, thus, alters the DNA-protein recognition dynamics [70]. For the purpose of this study the usage of NMA on the crystal structure is justified by the extremely stable nucleosome core structure in the nanosecond time scales [64]¹. In addition, our goal is to sample more conformations in the configurational space regardless of their quantitative character. As it was already pointed out in Sec. 4.1 the absolute values of the eigenfrequencies of each mode obtained by the ENM do not have physical meaning and they are rather arbitrary due to the neglected anharmonic effects and solvent damping. However, the relative eigenfrequencies are physically meaningful and, thus, they can give insights into the relative motions of the obtained normal modes.

We used the Nomad-Ref program [123] (<http://lorentz.immstr.pasteur.fr/nomad-ref.php>) to calculate the first 20 normal modes of the nucleosome (tNCP). The calculation was done with all nonhydrogen atoms present in the nucleosome. The default parameters of a cutoff of 10 Å, distance weight parameter d_0 of 5 Å [97] and average output RMSD of 3 Å were used [123]. For calculating the frequencies a force constant of 100 kcal/Mol/Å² was applied. The output nucleosome conformations were 15 for each mode, 7 on each side of the starting equilibrium conformation. The conformations were determined by the direction of the normal mode vectors. The potential energy of each conformation can be given by

$$\mathcal{V}_i(|\mathbf{r}_{\text{RMSD}}^i|) \propto B\omega^2|\mathbf{r}_{\text{RMSD}}^i|^2, \quad (5.1)$$

where ω is the calculated mode eigenfrequency (in 1/cm), $\mathbf{r}_{\text{RMSD}}^i$ is the all atom RMSD (in Å) between conformation i and the equilibrium one (tNCP) and B is a constant in kJ/mol. This nucleosome energy dependence on

¹Roccatano *et al.* [64] showed that the main body of the nucleosome stays relatively close ($\text{RMSD}^{\text{histones}} < 3 \text{ \AA}$, $\text{RMSD}^{\text{DNA}} < 7 \text{ \AA}$) to the crystal structure in a 20 ns MD simulation of a nucleosome with 147 bp of DNA with and without histone tails.

CHAPTER 5. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE TO THE NUCLEOSOME

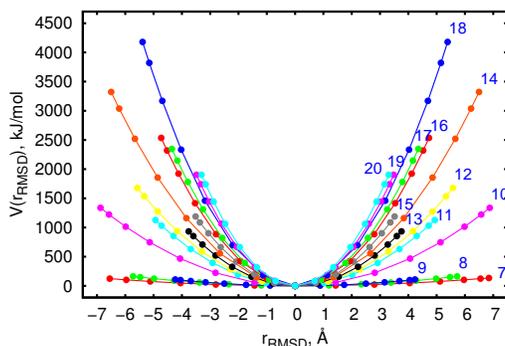


Figure 5.2: Energy dependence of the all-atom RMSD distance from equilibrium for each mode. The mode numbers are labelled on the right.

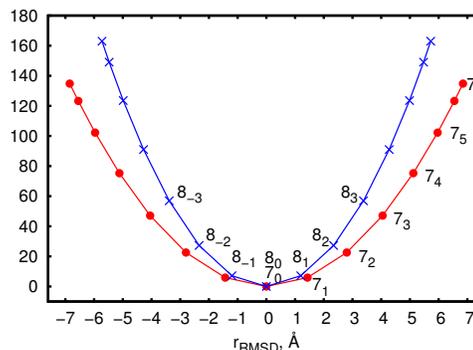


Figure 5.3: Energy profile of modes 7 and 8. Labelled are the nucleosome conformations used in the simulations.

r_{RMSD} is shown in Fig. 5.2. Since the first 6 modes are translation and rotation the "first" mode is numbered 7. Each point on the parabolas represents a different nucleosome conformation. It can be seen that the largest deviations from equilibrium experience modes 10, 7, 14, 8 and 12. However, modes 7 and 8 have much lower relative energies than the others, i.e. they can be attributed to the low frequency collective motions. Therefore, these two modes were chosen for our further analysis (See Fig. 5.3).

It has been shown that the density of the slow vibrational modes $g(\omega)$ of globular proteins does not follow Debye's theory ($g(\omega) \sim \omega^2$), but rather exhibits anomalous behaviour with $g(\omega) \sim \omega$ [124]. Based on these observations Tirion [34] proposed the ENM (Sec. 4.1.1) for the slow vibrational modes. The function $G(\omega)$ giving the scaled total number of modes up to frequency ω reads

$$G(\omega) = \frac{1}{n_m} \int_0^\omega g(\omega') d\omega', \quad (5.2)$$

where n_m is the total number of modes. Plot of $G(\omega)$ for the nucleosome is displayed in Fig. 5.4. Clearly, our data follow the ω^2 law, rather than ω^3 , and fall into a universal curve like other globular proteins [124]. Actually, this behaviour is expected since the ENM was used in determining the eigenvectors. It should be noted that the picture in Fig. 5.4 is plotted against only the first 20 modes and also, our structure contains DNA, which obviously does not affect the shape of the density of states. Interestingly, we see two regimes, one for the lowest three frequencies, and another for the higher frequencies. This observation has no obvious explanation. Avraham [124] concluded that the proteins behave as two-dimensional objects as far as slow vibrations are

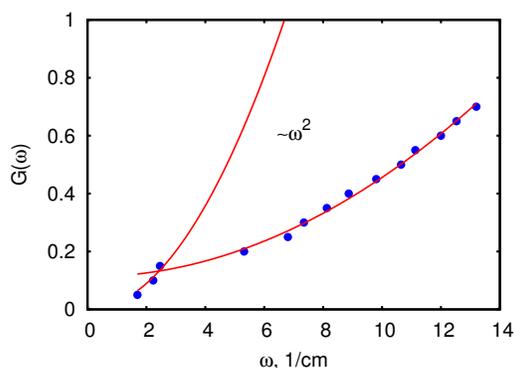


Figure 5.4: Scaled density of states $G(\omega)$ obtained by the NMA. The line is fitted to the data points as ω^2 (red).

concerned.

The largest motions obtained by the NMA were exhibited by the two 10 bp long linker DNAs (lDNAs) while the main core of the nucleosome remained stable. This DNA flexibility can play an important role in protein-DNA interactions [89] as well as in transcription and replication [70]. To represent the principal structural variations of the nucleosome, 13 different conformations were chosen from the two lowest frequency modes: modes 7 and 8 (modes 1-6 describe rigid body translation and rotation) (see Fig. 5.5).

For the 7th mode, only open conformations with respect to the equilibrium

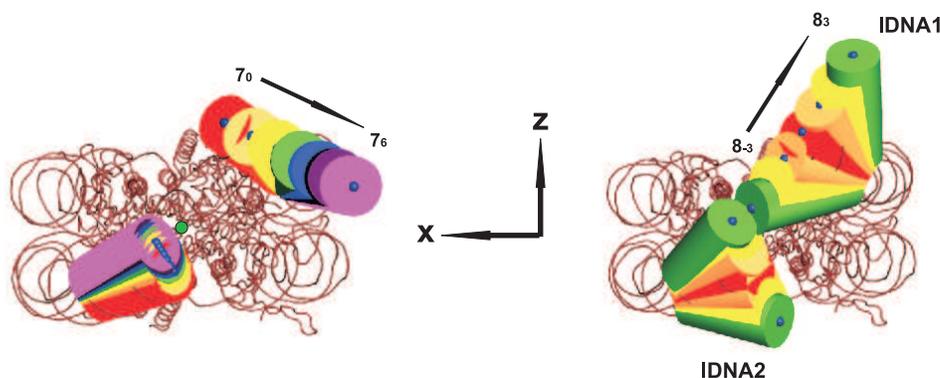


Figure 5.5: Nucleosome conformations generated by NMA. The lDNA conformations chosen from the 7th (left) and the 8th modes (right) are shown by cylinders with the blue points marking the cylinder axis. The equilibrium structure, (7_0) (corresponding to the crystal structure), is colored red.

structure (7_0) were chosen whereas the 8th mode was represented by three conformations on each side of the equilibrium structure (8_0) (Fig. 5.3). One of the lDNAs (lDNA1) showed larger fluctuations than the other (lDNA2)

CHAPTER 5. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE TO THE NUCLEOSOME

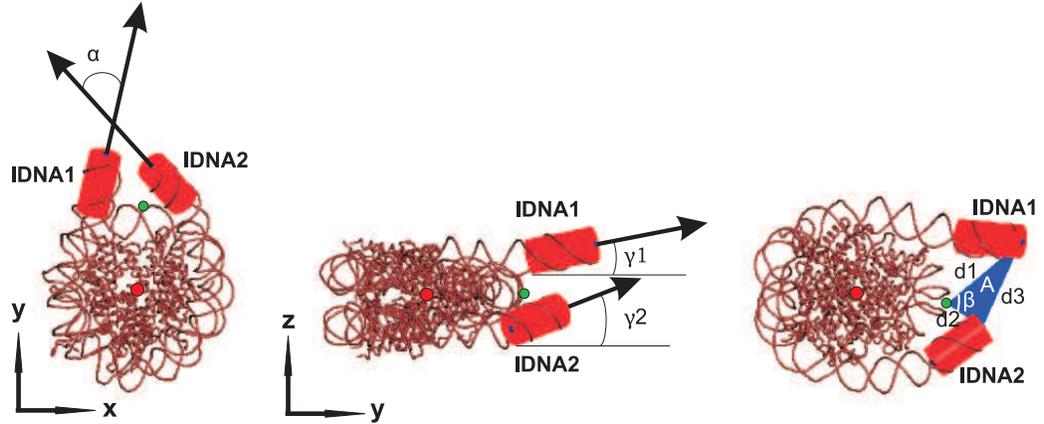


Figure 5.6: Geometric parameters for specifying each nucleosome conformation. The green (dyad) point lies on the dyad axis and the red point is located at the center of mass of the nucleosome.

(Fig. 5.5). These nucleosome conformations were quantified by 9 geometrical parameters displayed in Fig. 5.6. They are: α and ζ are the angles between IDNA1 and IDNA2 (α in the xy plane); γ_1 (γ_2) is the angle between IDNA1 (IDNA2) and the y -axis in the yz plane; A is the area of the triangle formed by the ends of IDNA1 and IDNA2 and the dyad point with angle β and sides d_1 , d_2 and d_3 . The parameters are given in Table 5.1.

Table 5.1: Values of the geometric parameters for each nucleosome (tNCP) conformation generated by NMA used in the BD simulations.

Conformation	α , deg	ζ , deg	A , \AA^2	β , deg	d_1 , \AA	d_2 , \AA	d_3 , \AA	γ_1 , deg	γ_2 , deg
7_0	71.44	70.76	468.40	92.31	40.34	23.24	47.36	14.57	2.65
7_1	56.34	55.50	499.61	80.32	44.73	22.66	46.62	7.21	4.94
7_2	43.81	43.72	539.35	72.19	50.83	22.29	48.86	0.73	7.12
7_3	34.25	36.31	587.34	67.47	57.49	22.12	53.11	4.14	9.09
7_4	27.33	32.53	639.66	65.14	63.81	22.10	58.09	7.48	10.75
7_5	22.56	31.05	689.21	64.19	69.08	22.16	62.69	9.64	12.06
7_6	18.14	30.65	749.37	63.89	74.80	22.31	68.00	11.53	13.40
8_{-3}	78.21	86.86	49.06	9.18	22.67	27.14	5.98	17.49	25.40
8_{-2}	77.27	80.80	197.11	39.48	25.58	24.24	16.87	9.59	19.16
8_{-1}	75.11	74.51	336.42	68.46	31.88	22.69	31.63	1.43	11.39
8_0	71.44	70.76	468.40	92.31	40.34	23.24	47.36	14.57	2.65
8_1	66.42	71.78	607.85	109.21	49.76	25.87	63.19	27.29	6.15
8_2	60.84	76.25	769.41	119.46	59.25	29.83	78.35	37.07	13.96
8_3	55.68	81.40	954.95	125.07	68.13	34.25	92.17	43.45	20.26

5.3 Electrostatic potential calculation

Partial charges, atomic radii and hydrogens were assigned to all atoms in the structures with the PDB2PQR program [125]. The nucleosome particle (tNCP) and the globular domain of the linker histone (GH5) have 13496 and 565 nonhydrogen atoms, respectively. The net formal charges are $\sim -237e$ for the nucleosome conformations and $+11e$ for the linker histone. Electrostatic potentials were computed by solving the nonlinear Poisson-Boltzmann equation on grids (1\AA spacing) with 257^3 and 200^3 points for each chosen nucleosome conformation and the linker histone, respectively (Sec. 4.2.2). The calculations were performed with the programs APBS [101] and UHBD [102] using asymptotic boundary conditions. The temperature was set to 300 K, the solvent dielectric constant to 78, the solute dielectric constant to 2 and the ionic strength to 100 mM. The surface of the biomolecules was represented by

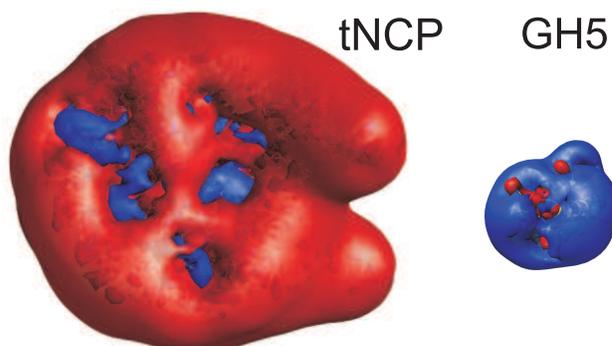


Figure 5.7: Electrostatic equipotential at $\pm 1 k_B T/e$ for the tNCP (left) and GH5 (right). Blue represents the positive potential, while red is the negative.

Van der Waals radii, which has been shown to stabilize protein-nucleic acid complex formation [126]. The parameters were chosen to resemble available experimental data [18, 21, 127, 128]. The obtained negative electrostatic potential of the nucleosome (Fig. 5.7) is due to the highly negatively charged DNA molecule, which screens the positive potential of the histone protein. Nevertheless, in Fig. 5.7 (left) a positive portion of potential on the nucleosome surface can be seen. On the other hand, the residues on the surface of the GH5 contribute to the overall positive potential of the linker histone.

5.4 Effective charges calculation

An important step for reducing the number of partial charges on the molecules, and thus, increasing the simulation time is the introduction of a small number of effective charges, which represent accurately the electrostatic potential around the molecule in a homogenous medium [103]. Effective charges were assigned to both molecules with the ECM program (Sec. 4.2.2). The num-

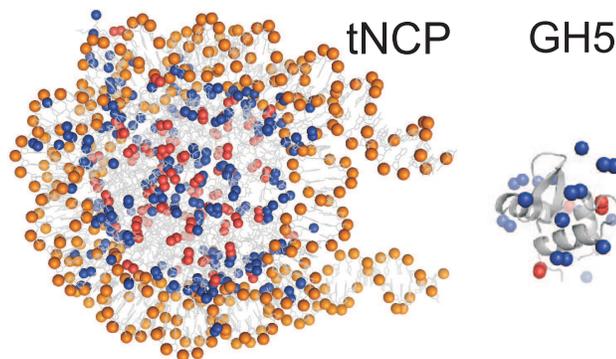


Figure 5.8: Effective charges represented as spheres for the tNCP (left) and GH5 (right). Blue spheres are the effective charges on positive residues, red on negative and orange on P atoms (negative) of the nucleic acid backbone.

ber of effective charges was 730 and 26 for the tNCP and GH5, respectively (Fig. 5.8). They were calculated using the previously obtained solution of the nonlinear Poisson-Boltzmann equation. The computation was done in a probe layer of 4 up to 7 Å around the molecular surface. The ionic strength was set to 100 mM and the solvent dielectric constant to 78.54. In the BD simulation procedure, the effective charges of GH5 move on the electrostatic potential grid of tNCP.

5.5 Brownian Dynamics setup

The BD docking simulations were carried out with the SDA4C package [98] (Sec. 4.2.2) modified in such a way that diffusional encounter complexes were recorded only if they satisfy predefined constraints. For our case, we used two constraints: (i) a center-to-center distance between both particles (< 74 Å) and (ii) dyad point-center distance between the dyad at the nucleosomal DNA (nDNA) and the center of the GH5 (< 40 Å). This criterion is based on experimental studies indicating that GH5 binds between and protects 20 bp of the lDNAs from nuclease digestion [32].

5.5. BROWNIAN DYNAMICS SETUP

In the simulations, the molecules are modeled as rigid bodies with the short-range attractive interactions neglected. An exclusion volume grid with 0.5 Å spacing was assigned to the nucleosome conformations to avoid overlaps. The trajectories start at a center-to-center distance $b = 300$ Å and finish at a distance $c = 640$ Å (See Fig. 4.1). The time step was set to 0.25 ps for center-to-center distances up to 130 Å and it increased linearly for larger distances. The translational diffusion constant was calculated according to the Einstein-Stokes relation (Eq. 2.13) and its value was $D = 0.0185$ Å²/ps. The rotational diffusion constant of the GH5 was set to 5.04×10^{-5} rad²/ps. If GH5 spent more time than $t^{\text{hit}}(r^{\text{hit}}) = 0.2$ ms within $r^{\text{hit}} = r_1^{\text{max}} + r_2^{\text{max}} + r^{\text{probe}} + r_{\text{atom}}^{\text{max}}$ for all sampled trajectories, the BD run was truncated (a run typically contained 10-100 trajectories). This is done to prevent very long sampling of bound configurations. The GH5 probe radius r^{probe} was assigned to 1.6 Å, the maximum radius of an atom $r_{\text{atom}}^{\text{max}}$ was 1.9 Å, the maximum distance from the center of mass to the furthest atom r_2^{max} of the GH5 was 19.9 Å while r_1^{max} had values depending on the nucleosome conformations ($r_1^{\text{max}} \in [77.4, 117.4]$ Å). The interaction energies as well as the coordinates of the complexes satisfying the constraints were recorded.

GH5 was docked to the 13 nucleosome conformations generated by NMA (Fig. 5.5) separately and for each system five different runs with different random generators, i.e. different starting positions and orientations, were performed. This assures different sampling paths for the GH5 in the configuration space. For each nucleosome structure, 25 000 complexes were recorded and the 2500 lowest energy docked complexes were clustered according to the backbone RMSD between them using the PDPIPE software [129]. After clustering, the representative of the top ranking cluster (with greatest number of docked complexes) was designated as the 'docked position'. The distance of each residue to either the nDNA or lDNA in the configurations generated in the BD simulations was monitored. The atoms N_ζ and C_β on Lys and Arg, respectively, were chosen for the distance calculation, whereas C_β (Ala, Val) and C_γ (Leu) were used for the hydrophobic residues. For each residue, its binding strength to nDNA and/or lDNA was described by weight factors defined by

$$\omega_n = \frac{N^{\text{nDNA}}}{\langle d \rangle_{\text{nDNA}}}, \quad \omega_l = \frac{N^{\text{lDNA}}}{\langle d \rangle_{\text{lDNA}}} \quad (5.3)$$

where N^{nDNA} and N^{lDNA} give the number of complexes (out of 2500) for which a certain residue is close (less than 15 Å) to the nDNA or lDNA, respectively, while $\langle d \rangle$ is the average distance of the residue to either nDNA or lDNA. The maximum possible value of ω was 1093.5 (1/Å) for $d = 2.29$ Å.

5.6 Structure of the nucleosome-linker histone GH5

In 8 out of the 13 conformations, the simulations revealed a single binding mode in which GH5 binds approximately one helical turn away from the dyad point, close to IDNA1 (Fig. 5.9). For these conformations, with geometry defined by $\alpha \in [55.7, 71.4]$, $\beta \in [63.9, 92.2]$ and $A \in [468, 749] \text{ \AA}^2$ (Fig. 5.6 and Table 5.1), the RMSD of GH5 from its docked position on the 7_0 conformation is within 6 \AA (Fig. 5.10). GH5 approaches the nDNA with helix 3, containing residues R47, K69, R73 and R74 (the K69site), and IDNA1 with the residues R42, R94 and K97 (the R42site). The third charged site, containing K52, K55 and K59 (the K59site), does not contact the nucleosome (Fig. 5.9). Although, residues K85 and K40 are situated on loops on both sides of the R42site they cannot be assigned to any of the proposed sites. The curves N^{K69site} and N^{R42site} in Fig. 5.10 give the number (in percentage) of docked complexes, for which the corresponding sites K69site and R42site are within 6 \AA to nDNA and IDNA, respectively. The distance is averaged over the distances of the residues comprising the relevant site. Hence, N^{K69site} and N^{R42site} are measures for the orientation of the GH5 with respect to the nucleosome. For the extreme conformation in the 7th mode (7_6), the same

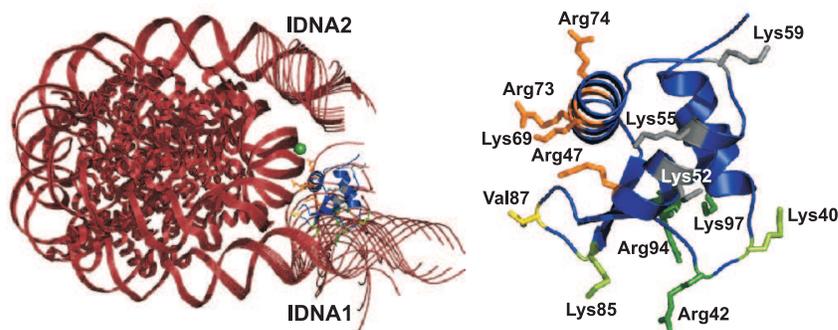


Figure 5.9: The docked position of GH5 (blue) on the nucleosome (red) shown with 13 superimposed conformations generated by NMA (left). The charged sites on GH5: K69site (orange), R42site (green) and K59site (grey) and the hydrophobic residue Val87 (yellow).

two sites on GH5 contact the DNA, but the contacts are different, i.e. the R42site contacts the nDNA and the K69site contacts IDNA2 (Fig. 5.10),

similar to the docking mode proposed by Zhou *et al.* [27] and Bharath *et al.* [29]. For conformations 8_2 and 8_3 , the GH5 also binds closer to IDNA2 than to IDNA1, but contacts IDNA2 with the R42site. Since 8_{-2} and 8_{-3} are closed nucleosome conformations (area $A < 200 \text{ \AA}^2$) in which the dyad point is not freely accessible to the GH5, the docked positions are outside rather than in between the IDNAs.

To identify the residues most important for binding and to compare them with experimental data, scaled weight factors (averaged over the configurations with $\text{RMSD} < 6 \text{ \AA}$, see Fig. 5.10) were plotted against the inverse FRAP half-time $t_{1/2}$ for recovery [17] (Fig. 5.11). Both, the weight factor and the FRAP half-time are indicators of binding strength, but there is no direct relation between them. One can, however, observe a qualitative agreement between computation and experiment. Not only do the two predicted binding sites (K69site and R42site) agree with experimental data, but both the simulation and the experiment show that Lys69 contributes most to binding whereas Lys59 contributes least. The simulation slightly overestimates the binding strength for Arg94 and Lys85.

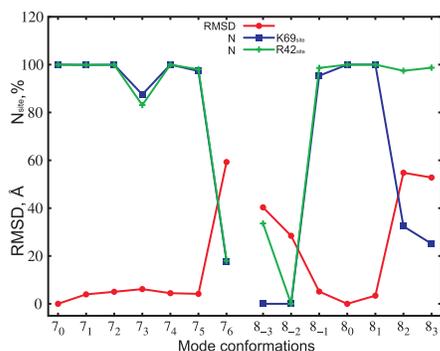


Figure 5.10: Comparison of the docking modes to the 13 nucleosome conformations showing the RMSD of the docked positions of GH5 from that to the 7_0 conformation of the nucleosome (red), the percentage of docked configurations in which the K69site (blue) or the R42site (green) is within 6 \AA of the nucleosome either to nDNA or IDNA, respectively.

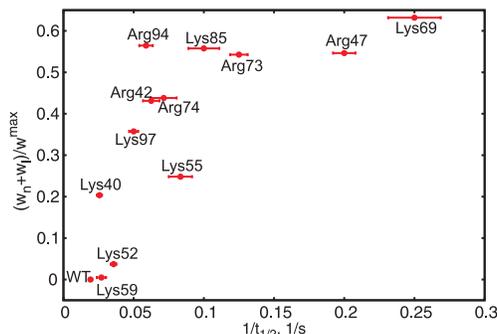


Figure 5.11: Weight factor (computed from simulations) versus the experimentally measured FRAP inverse half-time for recovery. The positions of the 12 positive and 10 hydrophobic residues (not plotted) on the surface of GH5 with respect to the nucleosome were quantified by weight factors ω_n and ω_l , which are indicators of binding strength to the nucleosomal DNA (nDNA) and IDNAs, respectively

Hydrophobic interaction energy

The hydrophobic loop containing Val87 lies between the nDNA and IDNA1 (Fig. 5.9). This loop could interact with the AT-rich IDNA [30] and, by means of short-range hydrophobic interactions, facilitate small readjustments of the GH5-nucleosome encounter complex to form the fully bound complex. The hydrophobic residues Leu66, Leu70, Ala77 and Ala78 belong to helix 3, which has been identified as binding to the nucleosomal DNA [17]. Val87 and Ala89 together with Lys85 belong to the loop between the beta sheets and structurally this loop appears between the two binding sites, K69site and R42site (Fig. 5.12). It is seen that the hydrophobic residues have less

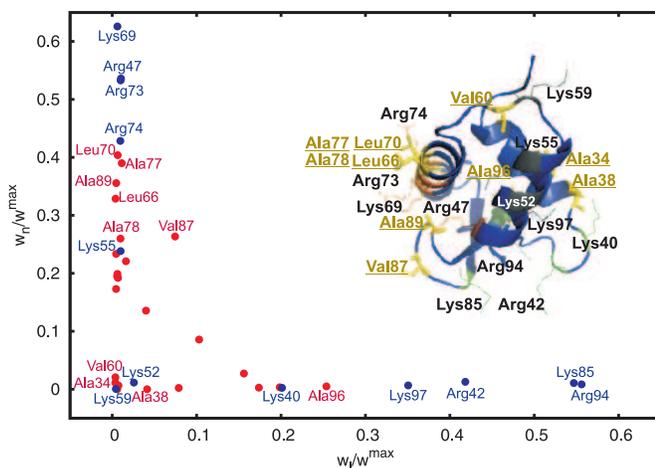


Figure 5.12: Scaled weight factors for the positive (blue) residues and for all 25 hydrophobic residues (red) of GH5 are plotted for binding to nDNA (ordinate) and IDNA (abscissa). The inset shows the location of the residues on the surface of GH5.

binding strength than the positive residues on K69site and R42site. In order to identify the contribution of the hydrophobic residues to the docked position of GH5 obtained we computed the hydrophobic interaction energy between the GH5 and a hydrophobic probe ('DRY' probe) using the program GRID [130–132]. Since it is thought that the hydrophobic effect arises due to water rearrangements in the hydration shell, the water entropy is taken into account for computing the hydrophobic interaction energy. Molecule GH5 is represented by grid points and for each point the hydrophobic interaction energy is computed. It is given by three terms

$$E^{\text{hydro}} = E^{\text{WENT}} + E^{\text{LJ}} - E^{\text{HB}}, \quad (5.4)$$

where E^{WENT} is a favourable entropic contribution to the free energy of water, which is assumed to have the same magnitude for the hydration shell not participating in the interactions, whereas E^{LJ} and E^{HB} describe the favourable Lennard-Jones and unfavourable hydrogen bonding (due to the polar atoms

5.6. STRUCTURE OF THE NUCLEOSOME-LINKER HISTONE GH5

on the surface) interactions, respectively. The most favourable value of the hydrophobic interaction energy was -1.93 kcal/mol around residues Leu66 and Leu70. A favourable hydrophobic region was also clearly seen around Val87 at an interaction energy of -1.45 kcal/mol (Fig. 5.13). It lies adjacent

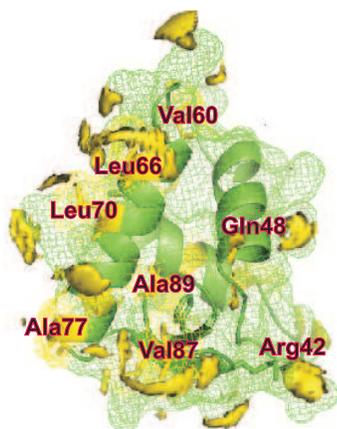


Figure 5.13: Hydrophobic binding sites on GH5 identified using the program GRID [130–132]. GH5 (green cartoon) is shown with its van der Waals surface (green mesh) and hydrophobic regions (yellow) with interaction energy more favorable than -0.5 kcal/mol for a hydrophobic probe, the 'DRY' probe (yellow). The residues near the hydrophobic regions are labelled. The figure was made with Pymol (www.pymol.org).

to a loop which was considered as a separate binding site interacting with AT nucleotides in the major groove of IDNA by Cui and Zhurkin [30]. These authors claimed firstly, that an AT-rich IDNA facilitates LH binding and, secondly, that the hydrophobic interactions between the Val87 loop and the AT nucleotides lead to bending of the IDNA. Our results are in agreement with this statement. The GH5 preferably contacts the IDNA1 (10 bp), which has 6 A or T nucleotides in contrast to only 2 on the IDNA2, and NMA showed that the IDNA1 is more flexible than IDNA2 and the reason could be the presence of more AT nucleotides. However, our BD docking procedure identified the encounter complex based only on electrostatic interactions between molecules and this primary, initial interaction involved the charged binding sites, K69site and R42site, on the GH5. Due to short-range hydrophobic interactions, a subsequent second interaction of GH5 might involve the Val87 loop, which could turn into the major groove of IDNA1 as suggested by Cui and Zhurkin [30]. Such an interaction with its accompanying conformational relaxation requires higher resolution modeling with treatment of conformational flexibility, e.g. by atomic detail Molecular Dynamics (MD). This will be presented in Chapter. 9.

The dominant BD docking mode is thus consistent with the models proposed by Brown *et al.* [17] and Cui *et al.* [30] using FRAP and sequence analysis, respectively, but differs from other proposed models [27–29].

CHAPTER 5. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE TO THE NUCLEOSOME

NMA of the nucleosome-linker histone complex

We next addressed the question of how H5 tunes nucleosomal DNA accessibility by performing a NMA of the nucleosome with GH5 docked to it in the position obtained from the BD simulations (Table 5.2) and comparing the motions of the IDNAs in the NMA with and without GH5 (Fig. 5.14). GH5 influences the way both IDNAs move. For IDNA2, the diagonal pattern indicates similar motion² between neighboring modes, whereas the modes of IDNA1 have different directions in the presence and absence of GH5 (Fig. 5.14). The motion of IDNA1 is more suppressed by GH5 than IDNA2 as indicated by the distance d_1 between the dyad point and the end of IDNA1 in Fig. 5.15. Zlatanova et al. [70] proposed that GH5 binds close to one of

Table 5.2: Values of the geometric parameters for each nucleosome conformation generated by NMA for the (tNCP+GH5) complex.

Conformation	α , deg	ζ , deg	A, Å ²	β , deg	d_1 , Å	d_2 , Å	d_3 , Å	γ_1 , deg	γ_2 , deg
7 ₀	71.44	70.76	468.40	92.31	40.34	23.24	47.36	14.57	2.65
7 ₁	58.67	55.97	462.61	71.62	40.01	24.37	39.75	13.39	17.25
7 ₂	47.00	46.18	522.15	59.63	39.75	30.45	35.82	12.25	28.36
7 ₃	37.37	41.36	622.35	55.11	39.56	38.36	36.06	11.22	35.52
7 ₄	30.03	39.58	733.22	53.86	39.43	46.05	39.16	10.33	39.83
7 ₅	24.85	39.20	832.88	53.76	39.36	52.47	43.14	9.63	42.36
7 ₆	19.99	39.43	946.05	54.14	39.30	59.40	48.35	8.90	44.38
8 ₋₃	97.56	94.27	433.76	105.83	36.13	24.96	49.19	19.76	6.77
8 ₋₂	90.50	88.12	426.43	102.23	35.80	24.38	47.39	18.89	5.52
8 ₋₁	81.51	80.06	438.43	97.46	37.17	23.79	46.66	17.20	4.12
8 ₀	71.44	70.76	468.40	92.31	40.34	23.24	47.36	14.57	2.65
8 ₁	61.76	61.55	510.60	87.61	44.93	22.75	49.51	11.35	1.20
8 ₂	53.71	53.71	558.32	83.71	50.28	22.34	52.74	8.16	0.42
8 ₃	47.70	47.77	605.86	80.67	55.74	22.03	56.52	5.46	1.57

the IDNAs while the C-terminal domain acts as a bridge between the IDNAs and thus locks the nucleosomal gate and shuts down DNA transcription and replication. Our models suggest that, due to the spontaneous accessibility of the DNA [133], H5 could first bind strongly to the nucleosome in the dominant position identified here (see Fig. 5.9) and secondly its globular and C-terminal domains could bring the IDNAs together and H5 would remain bound for a long time because the IDNA motion is suppressed. In our model, the binding of H5 depends not only on the interactions with the nucleosome, but also on the geometry and sequence of the IDNAs. The nucleosome is asymmetric with respect to the IDNAs and so is the binding of

²The motion is similar for an average angle between the modes of 0-45 and 135-180 degrees, while it is dissimilar for an angle between 45 and 90 degrees.

5.6. STRUCTURE OF THE NUCLEOSOME-LINKER HISTONE GH5

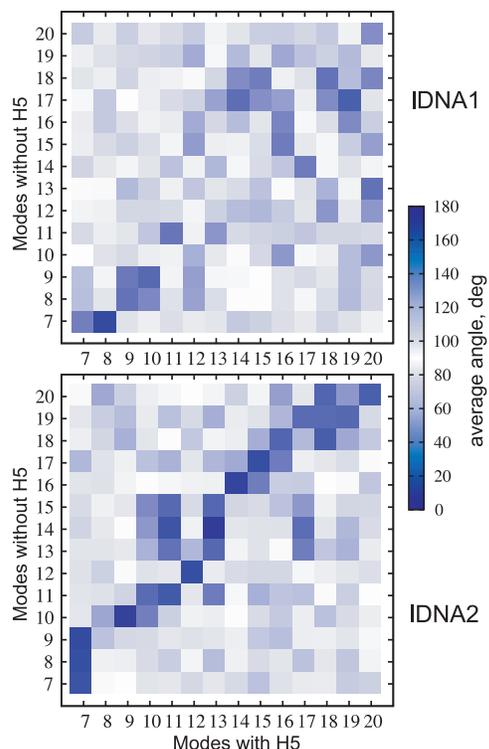


Figure 5.14: Comparison of the eigenvectors of modes 7-20 of IDNA1 (up) and IDNA2 (down) computed for the nucleosome with (abscissa) and without (ordinate) GH5 docked.

GH5, even though LH binding has been modeled as symmetric in several studies [22, 76, 76, 77, 80]. In the dominant binding mode, GH5 binds to IDNA1, but, in the open 7_6 , 8_2 and 8_3 conformations, GH5 binds to IDNA2, even though IDNA1 is accessible. This suggests that GH5 binding to IDNA1 favors the more compact form of chromatin whereas binding to IDNA2 tends to prevent chromatin fiber compaction. The present models provide the basis for future studies/MD simulations (see Chapter. 9) with a more detailed treatment of conformational flexibility to investigate induced fit upon LH-nucleosome binding.

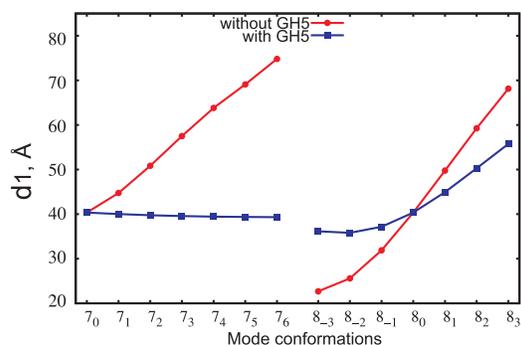


Figure 5.15: The distance d_1 between the dyad point and IDNA1 with (blue) and without (red) GH5 docked for modes 7 and 8.

5.7 Summary and implications to chromatin fiber

In this chapter, we identified the nucleosome-globular domain of linker histone H5 structure (chromatosome particle) by extensive Brownian Dynamics (BD) simulations combined with Normal Mode Analysis (NMA). The linker DNAs (lDNAs), 10 bp each, exhibited the most pronounced conformational changes obtained by the NMA and, based on these conformations, the GH5 was docked to the nucleosome by BD in order to find out the diffusional encounter complex. A dominant binding mode of GH5 with respect to the nucleosome, placed asymmetrically one helical turn away from the dyad point close to lDNA1, was found for 8 nucleosome conformations. Two binding sites of GH5, the K69site (R47, K69, R73, R74) and R42site (R42, K97, R94), binding to the nucleosomal DNA (nDNA) and lDNA1, respectively, were revealed. Residue K69 was found to contribute most to binding, while K59 least. All these findings are in agreement with an experimental FRAP study [94]. The other 5 extreme conformations of the nucleosome showed different GH5 binding modes with the most open structures (7₆, 8₂ and 8₃) having GH5 bound to lDNA2.

The obtained chromatosome structures can contribute differently to the formation of higher-order chromatin fiber. Recently, experimental electron microscopy (EM) and theoretical Monte Carlo study [58] reported that chromatin fiber can exist in a heteromorphic state, i.e. simultaneously having zig-zag and solenoidally connected nucleosomes. It was argued that the transition between these states is tuned by the presence of linker histones and divalent ions; the former forming a tight zig-zag structure [58]. In addition, the Monte Carlo simulations showed that the linker DNA flexibility is an important issue for the formation of structurally different chromatin fibers and the linker DNAs conformations vary significantly in the presence and absence of linker histone. This is in accord with our NMA data showing large linker DNAs fluctuations. For example, changing the angle α (Fig. 5.6) can lead to different intersection points between lDNA1 and lDNA2 and, hence if extended, to different linker lengths. Another EM study [85] showed that the chromatin compaction for 167 bp and 197 bp (the most common in nature) of nucleosome repeat length (NRL) has small and high dependence on H5 (0.5 and 1 H5 per nucleosome), respectively. The authors claimed that the '30 nm' fiber is an ordered interdigitated solenoid structure with NRL > 177 bp [56], whereas the 167 bp NRL fiber is a zig-zag structure in the presence of H5 [85]. The compaction of chromatin fiber by the presence of linker histone has been also confirmed by FRET data [18] as well as by theoretical models [22, 76, 77, 79]. The dynamic nature of chromatin [55] implies that DNA

5.7. SUMMARY AND IMPLICATIONS TO CHROMATIN FIBER

takes different conformations and interacts in dynamic fashion with DNA binding proteins [89]. The wrapping and unwrapping of DNA [6, 122, 134] from the nucleosome is a way by which other proteins can gain access to its genetic information.

In regard to all these studies, our modelling of different nucleosome conformations with a linker histone bound can shed more light into the compaction of a '30 nm' chromatin fiber. The dominant binding mode found might contribute to a more compact chromatin fiber, while the modelled complexes with the most open nucleosome conformations might be relevant for a loose chromatin fiber.

6

Brownian Dynamics docking of the linker histone mutants to the nucleosome

Proteins can be mutated, i.e. one or several residues can be substituted by other residue(s) or chemical group(s), and useful information can be gathered about their function, possible reactions with other biomolecular species and structural changes. In living cells, the mutational process appears naturally, for example, somatic hypermutation is a process related to the immune system function [135].

In this chapter, the effect of linker histone GH5 mutants on the binding mode of GH5 as well as on the tNCP-GH5 structure is investigated by Brownian Dynamics simulation carried out with the SDA4C program. A linker histone with replaced single or multiple residues will be referred to as a *mutant*.

6.1 Mutants

Mutants were modelled to GH5 by replacing one or several residues with Pymol [136]. In total 25 different structures of GH5 were obtained (see Table 6.1) including the wild type (WT) linker histone GH5. The structures with several residues mutated at once (S1S2, SITE1, SITE2) follow the Brown *et al.* [17] notation, i.e. all are mutated to alanine (Ala) as S1S2 (K69A, K73A, K85A, K40A, R42A, K52A, R94A), SITE1 (K69A, K73A, K85A) and SITE2 (K40A, R42A, K52A, R94A). The mutant 6ALA represents simultaneous mutations to Ala of residues K40, R42, K52, K69, R74 and K85. Most of the mutations replace positively charged residues with either neutral or negative residues. The linker histone mutants were prepared for a BD simulation using the procedure described in Chapter 5. The net

CHAPTER 6. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE MUTANTS TO THE NUCLEOSOME

Table 6.1: Mutated structures of the GH5. The mutants S1S2, SITE1, SITE2 and 6ALA include multiple residue mutations to Ala.

R94A	D65K	H25E	K40A	K52A
K55A	K55DD65K	K55E	K59A	K69A
K82V	K85A	K97A	R42A	R42E
R47A	R47E	R73A	R73E	R74A
Multiple mutations to Ala	S1S2 K40, R42, K52, K69, R73, K85, R94	SITE1 K69, R73, K85	SITE2 K40, R42, K52, R94	6ALA K40, R42, K52, K69, R74, K85

formal charge for the mutants varied between +5e and +13e. Brownian Dynamics docking was performed with the same criteria as for the wild type (WT) GH5. Each of the linker histone mutants was docked to the 13 nucleosome conformations generated by NMA separately and, for each system, five different runs with different random generators, i.e. different starting positions and orientations, were performed. In total, 1625 simulations were carried out.

6.2 BD docking and analysis of the results

Two main binding sites, K69site (on helix 3) and R42site (on loop 1 and beta sheet loop) on the WT linker histone have already been identified in Chapter 5, but the contribution of each residue to binding depends on the nucleosome conformation and whether such contributions are additive remained unclear. Brown *et al.* [17] combined systematic mutagenesis on murine H1⁰ linker histone variant *in vivo* using FRAP, and based on the experimental results, proposed a model for the chromosome particle. Our data on the WT linker histone agree with the proposed two binding sites and their orientation with respect to the nucleosome by Brown *et al.* [17]. We have found that for 8 nucleosome conformations ($7_0 \rightarrow 7_6, 8_{-1} \rightarrow 8_1$) WT binds approximately one helical turn away from the dyad axis close to 1DNA1 (see Fig. 5.9), while for the other open conformations ($7_6, 8_2, 8_3$) WT binds close to 1DNA2 with a different orientation. Here, by introducing mutations to the structure of GH5, we try to shed more light onto the linker histone nucleosome interactions.

6.2.1 Nucleosome-linker histone complexes

The dominant docking position of WT GH5 to the nucleosome was used as a reference structure and the docked representative of all mutants was compared to it (Fig. 6.1 and Fig 6.2). The plots provide information about i) the position of each docked mutant to the dominant WT binding mode expressed by the RMSD (red line), and, ii) the accuracy of this position expressed by the percentage of docked mutants $N\%$ (out of 2500) (blue line) for which the RMSD of their representative was calculated. We use the following notations (schemes) for describing how much the docked mutants differ from the WT data

- If $\text{RMSD} \in [0,15] \text{ \AA}$

$$\begin{cases} N\% \in [80, 100] & \rightarrow \text{alike} \\ N\% \in [60, 80] & \rightarrow \text{matching} \\ N\% \in [40, 60] & \rightarrow \text{comparable} \end{cases}$$
- If $\text{RMSD} \in [15,30] \text{ \AA}$

$$\begin{cases} N\% \in [60, 100] & \rightarrow \text{proximate} \\ N\% \in [40, 60] & \rightarrow \text{intermediate} \end{cases}$$
- If $\text{RMSD} \in [30,100] \text{ \AA}$

$$\begin{cases} N\% \in [40, 100] & \rightarrow \text{distinct} \end{cases}$$
- If $\text{RMSD} \in [0,100] \text{ \AA}$

$$\begin{cases} N\% \in [0, 40] & \rightarrow \text{undefined.} \end{cases}$$

The '*undefined*' notation means that the solution plotted does not represent statistically all the complexes obtained and more detailed analysis of the other occupancy clusters is needed. The two notations *intermediate* and *distinct* are of interest as well, because they differ significantly from the established dominant WT binding mode.

In Fig. 6.1 results are given for all nucleosome conformations of mode 7 except for conformation 7_3 , which has similar features to 7_2 . General inspection of the plot (Fig. 6.1) shows that most of the single mutations do not contribute to a significant displacement from the WT dominant binding mode apart from the extreme 7_6 . The mutants displaying a *distinct* binding mode are K40A and R73E in conformation 7_0 , SITE1 in 7_1 , R73A in 7_4 , R73A, SITE1 in 7_5 and K55DD65K, K82V, R73A and S1S2 in 7_6 . Apart from the mode 7_6 , which was considered as a metastable conformation in regard to chromatin fiber compaction, the other conformations lead to a completely different diffusional encounter complex only for one (7_1 , 7_4), two (7_0 , 7_5) or zero (7_2 , 7_3) mutants. Mutant SITE1 consists of multiple mutations and its deviation in 7_1 and 7_5 is understandable. Since the deviation for K40A appears only in 7_0 and K40 was considered as nonbinding we will not discuss

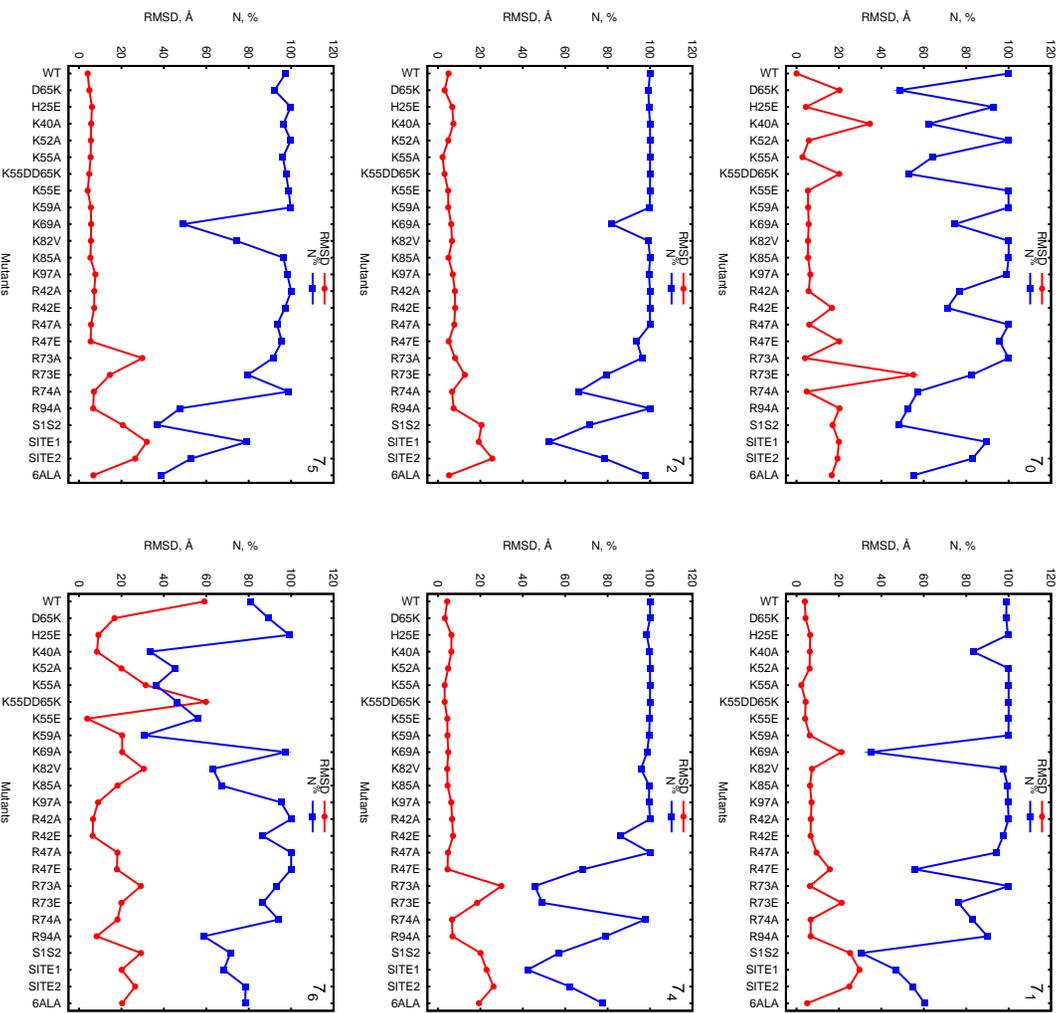


Figure 6.1: RMSD of the docked solution of each mutant to nucleosome conformations (T_0 , T_1 , T_2 , T_4 , T_5 , T_6) to the docked WT GH5 in conformation T_0 (red). The blue line shows the percentage $N\%$ of the solutions which are represented by this docked solution.

its contribution in detail. The curves of the single mutants indicate that neutral or negative mutations of R73 (belonging to helix 3) affect the final position of GH5. This means that residue R73 might contribute significantly to binding of GH5 to the nucleosome for 7_4 and 7_5 . Worth mentioning is also that neutral mutation of R73A is implemented in the SITE1 and S1S2, while SITE2 and 6ALA do not include it.

Next level of complexes, which deviate from the WT are the *intermediate* ones. The mutants falling into this group are D65K, K55DD65K, R94A, S1S2, 6ALA in 7_0 , R47E, SITE2 in 7_1 , SITE1 in 7_2 , R73E, S1S2, SITE1 in 7_4 , SITE2 in 7_5 and K52A in 7_6 . Again, for every nucleosome conformation, apart from 7_6 , there are multiple mutations exhibiting different binding modes. This is not as surprising as the stability of the single mutants to the WT dominant binding mode. We do not observe even a single mutant which has a clear deviation from the WT in 7_0 position for all nucleosome conformations of mode 7. This may imply that single mutations are not sufficient to alter the position of the linker histone with respect to the nucleosome. This could be related to the experimental study of An *et al.* [95], who showed a unaffected binding position of the linker histone in presence and absence of N-terminal histone tails. In both cases, experiment and simulation, there is a reduction of the electrostatic interaction energy between the nucleosome and the linker histone. However, in the experiment, this reduction is caused by charged sites on the nucleosome, while in the simulation the charged sites are on the linker histone. Although the magnitude of reduction is different, the experimental observation could have an equivalent effect for the binding position of single mutants, i.e. they remain close to the WT dominant position. On the other hand, the multiple mutations display a ruffle pattern on the plots for all the conformations suggesting their important role in linker histone binding. The third important group consists of residues belonging to the *undefined* class. These are K69A in 7_1 , S1S2 in 7_5 , K40A, K55A and K59A in 7_6 . Inspection of the other cluster representatives of K69A in 7_1 show that K69A binds in the mirror position of the WT, i.e. close to IDNA2. Similar behaviour is observed for S1S2 in 7_5 as well. Overall, the plots of mode 7 mainly give insights into the positional distribution of the BD docked mutants with respect to the dominant position of the WT. Apart from the multiple mutants, single mutations of residues R73, K40, D65, R94, K52, K69, K55 and K59 give slight change in the GH5 position only for some conformations. In general, the single mutations do not change significantly the position of the diffusional encounter complex, which remains *alike* only for the intermediate conformations 7_2 , 7_3 , 7_4 and 7_5 . Upon a more detailed look one can distinguish conformations 7_2 and 7_3 having the 'smoothest' pattern among the others. This observation would imply that the conformations 7_2

and 7_3 do not disturb the binding mode of the single mutants and, hence, they might be the most favourable for binding of the linker histone. Since the crystal structure of the equilibrium conformation 7_0 was resolved without linker histone such a statement might be reasonable.

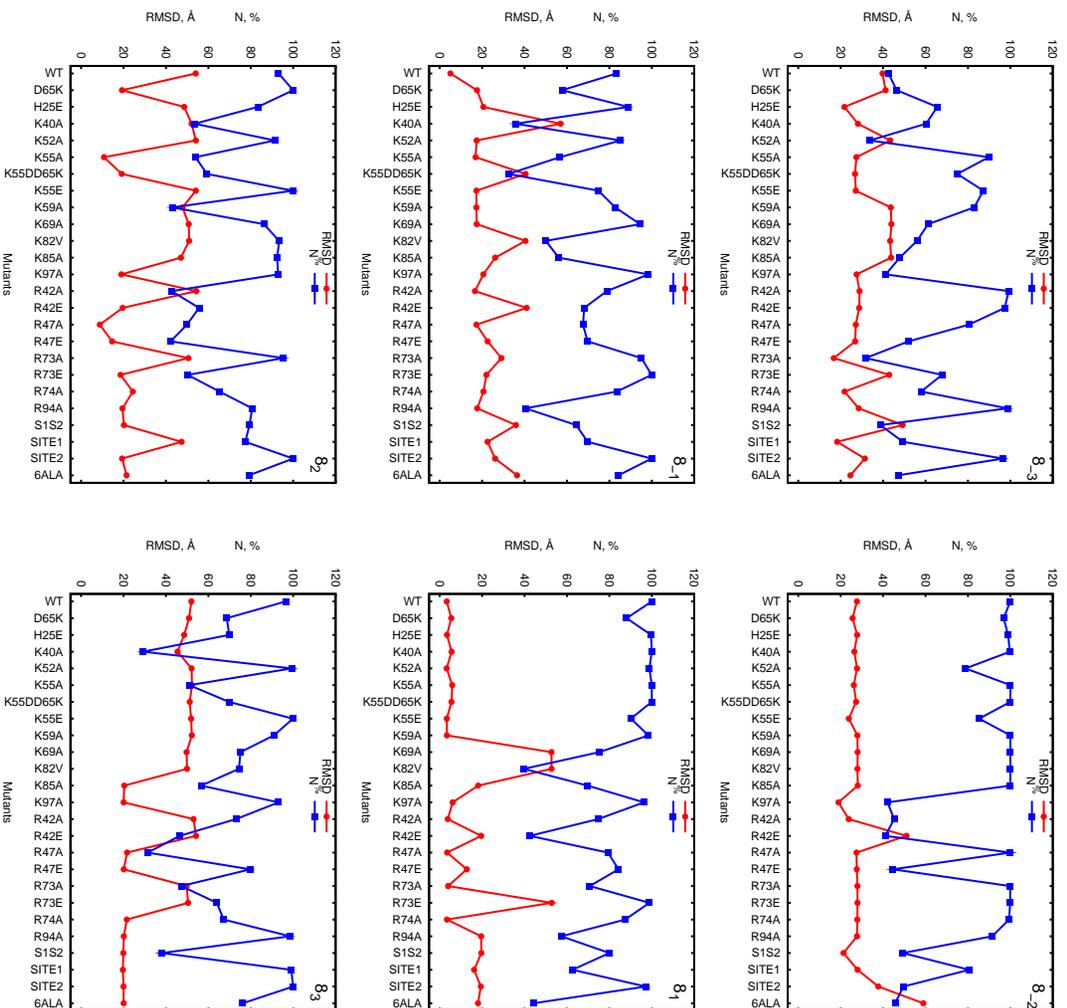


Figure 6.2: RMSD of the docked solution of each mutant to nucleosome conformations (8_{-3} , 8_{-2} , 8_{-1} , 8_1 , 8_2 , 8_3) to the docked WT GH5 in conformation $8_0 = 7_0$ (red). The blue line shows the percentage $N\%$ of the solutions which are represented by this highest docked solution.

The pattern of the mutant's positions for mode 8 (Fig. 6.2) indicates more irregular curves. Even the conformations 8_{-1} and 8_1 , which have a dominant binding mode of WT, show considerable deviations for some of the mutants. The mutants belonging to the *distinct* group are K82V, R42E, R73A, S1S2 and 6ALA in 8_{-1} , K69A and R73E in 8_1 . *Intermediate* binding mode have K85A, R94A in 8_{-1} , R42E, R94A and 6ALA in 8_1 . Both the *distinct* and *intermediate* positions of the mutants suggest that in mode 8 single mutations of the residues constituting the binding sites K69site and R42site also contribute to the observed deviations. We will not discuss in detail the deviations for the other conformations in mode 8, because the WT study showed that they have different implications on the linker histone binding. The interesting point here is that only conformation 8_1 depicts a pattern similar to the conformations 7_1 , 7_2 , 7_3 , 7_4 and 7_5 with exception of K82V and R42E. This means that only these conformations might lead to a stable nucleosome-linker histone complex, while the others could lead to a metastable complex depending on the level of IDNA opening.

6.2.2 Dynamics

The positional distribution of the mutant complexes with respect to the WT dominant binding mode examined in the previous Section 6.2.1 does not give any information on the dynamics of the encounter complex formation. Here, we will partially reveal some aspects of it through the residence time per trajectory t^{res} . This is the simulation time spent of a particle within distance windows of 1 Å up to 150 Å. Although, it was proposed that the nucleosome conformations 7_2 , 7_3 and 8_1 are perhaps the most robust for linker histone binding, we will still keep to the equilibrium conformation 7_0 as a reference structure. The average (over 5 random generators) residence time t^{res} per trajectory in 1 Å slab for all mutants and conformations is given in Fig. 6.3. It is clearly seen that the linker histone spends most of its diffusional time in a trajectory at a center-to-center distance $d \sim 61\text{-}63$ Å. The WT linker histone binds to the nucleosome (Sec. 5.6) at a separation close to this value for all nucleosome conformations. In Fig. 6.3 we see similar behaviour for the mutants as well, suggesting that all they form a diffusional encounter complex at the nucleosomal DNA (nDNA) in the plane of the nucleosome (xy in Fig. 5.6). However, the value of t^{res} at ~ 60 Å spans approximately two orders of magnitude for all mutants. This would imply that the interactions close to the nucleosome are not so favourable with respect to the dynamics for some of the mutants. In addition, the plot depicts a pronounced wing on the left peak side proposing that some mutants spend time at distances in the range 30–55 Å. Such positions of the linker histone mutants are definitely not between

CHAPTER 6. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE MUTANTS TO THE NUCLEOSOME

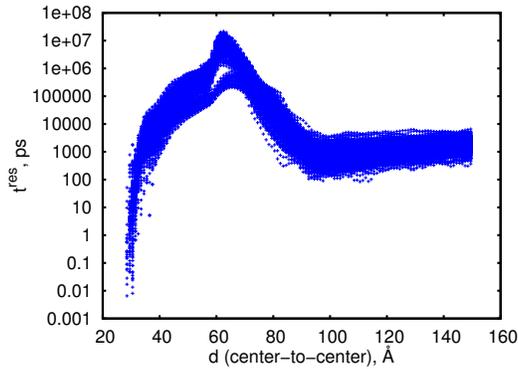


Figure 6.3: Residence time t^{res} per trajectory (ordinate) versus center-to-center distance d (abscissa) for all mutants and conformations.

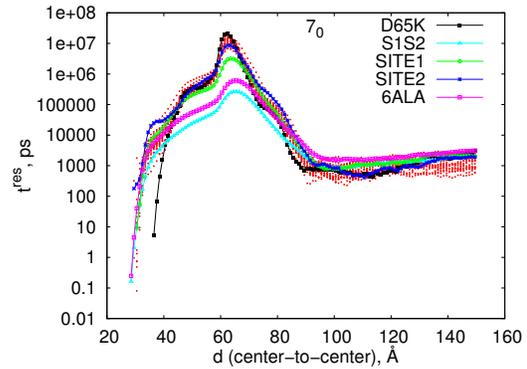


Figure 6.4: Residence time t^{res} per trajectory vs d for nucleosome conformation 7_0 . Highlighted are the curves for D65K (black), S1S2 (cyan), 6ALA (magenta), SITE1 (green) and SITE2 (blue).

the linker DNAs where a 'stem' structure forms [71]. Apparently, these mutants appear to bind either below or above the core body of the nucleosome (z axis in Fig. 5.6), which is more positively charged due to the histone octamer [137]. Such binding of the WT linker histone, however, does not fall into the 'stem' picture [32], although a model based on a photoactivable cross-linking data proposed an asymmetric position of the GH5 inside the gyres of the nDNA and lDNA [138]. Therefore, the time spent in this area can be attributed either to a formation of a unfavourable encounter complex of mutants having reduced positive charge or to a metastable local minimum along the diffusional path of the particle to its encounter complex positions identified in Section 6.2.1.

Figure 6.4 shows the same plot, but only for 7_0 nucleosome conformation. The curve shape for the other conformations does not show significant difference. On the graph highlighted are the time distributions for the mutants D65K, S1S2, SITE1, SITE2 and 6ALA. Seven simultaneous single mutations of S1S2 (Table 6.1) reduce dramatically its residence time in the range 40–80 Å. Similar behaviour is observed for 6ALA and SITE1, whereas SITE2 follows the WT distribution. The main difference between 6ALA and S1S2 (apart from R74A and R73A mutations, respectively) is the mutation of R94A in S1S2, which decreases the residence time of the latter. In black D65K shows a deviation from the curves in the range 35–40 Å, but have the highest peak among all the mutants. Since residue Asp65 is not directly exposed to the surface of the GH5, but it is directed towards Lys55 [17], replacement of

6.2. BD DOCKING AND ANALYSIS OF THE RESULTS

its negative charge to positive would result in a electrostatic repulsive interaction between K55 and K65. To avoid this we modeled K65 pointing out from the GH5 surface. In such a way D65K can actively participate in the tNCP-GH5 interactions around 60 Å rather than between 30 – 40 Å.

In order to compare qualitatively theory with experiment, the residence times obtained by the simulations (Fig. 6.5) and by the FRAP data [17] (Fig. 6.6) were plotted. The first picture gives the total average residence

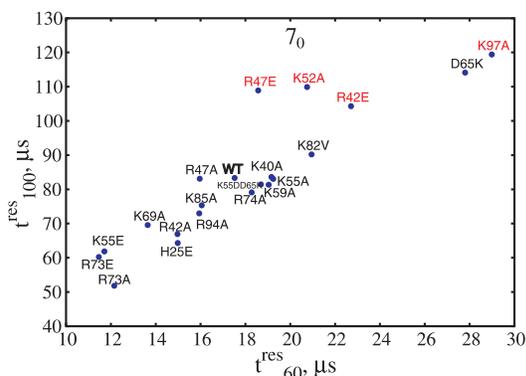


Figure 6.5: Total residence time per trajectory up to 100 Å (ordinate) and 60 Å (abscissa) interparticle distance.

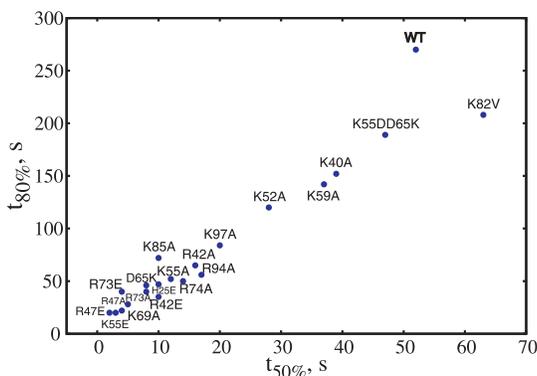


Figure 6.6: FRAP time at 80% recovery (ordinate) versus 50% recovery (abscissa). Data from Brown *et al.* [94]

times per trajectory up to distances 100 Å (y axis) and 60 Å (x axis), while the second picture depicts the experimental FRAP recovery time for 80% (y axis) and 50% (x axis) [94]. The times for S1S2, SITE1, SITE2 and 6ALA are very short and qualitatively agree with experiment and are not shown. The experimental time is related to the distance, which a mutant with attached green fluorescent protein (GFP) diffuses through, in order to cover the photobleached region. Slow fluorescence recovery would mean short passage distance and long time, i.e. the mutant stays closely bound to the nucleosome. In a similar way, the simulation residence time gives information on the strength of the attractive interactions within a spherical surface around the center of the nucleosome. However, important to note is that the time and length scales of the experiment and the simulation differ considerably. Therefore, both plots can be compared only qualitatively. Diagonal behaviour of the residence times is observed on both plots. The experimental WT time is longer than the time for the relevant mutants plotted, which suggests that all these residue contribute to binding to a different extent (Fig. 6.6). Only K82V deviates from this behaviour, which is seen on both figures. Interestingly, the simulation data are spread on both direction around WT in contrast to the experiment. Completely different behaviour

CHAPTER 6. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE MUTANTS TO THE NUCLEOSOME

is observed for R42E, R47E, K52A, D65K and K97A and moderately different for K40A, R47A, K55A, K59A, R74A and K55DD65K. The others are in qualitative agreement with the experiment. The intuitive longer time for D65K not observed in the experiment has been attributed to the structural position of K65 [17]. The mutant K97A spends more time than the WT up to 60 Å as well as up to 100 Å. Residue K97 links the GH5 with the basic C-terminal domain of linker histone H5 and, thus, it should not be in direct contact with the nucleosome. Replacement to a neutral residue normally should lead to a lower affinity to binding, which is not observed for K97A and K52A. This can suggest that these residues do not contribute directly to binding. It is, however, more surprising the residence time of mutants R47E and R42E. Such dramatic change in the residence time indicates that both residues R47 and R42 favour significantly the binding process, although the mutants R42A and R47A spend less residence time up to 60 Å (Fig. 6.5). One would expect that addition of negative Glu should lead to repulsion and not attraction between the tNCP and the GH5. The positively charged sur-

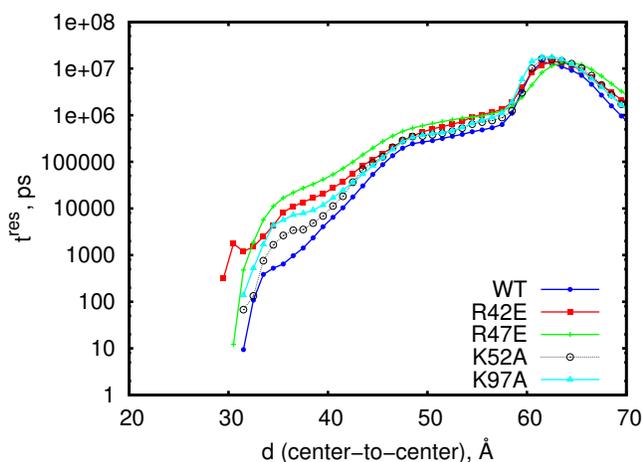


Figure 6.7: Residence time t^{res} per trajectory (ordinate) versus center-to-center distance d (abscissa) for the WT and mutants R42E, R47E, K52A, K97A. These mutants show deviation from the experimental data (Fig. 6.5)

face of the nucleosome histone cores might attract both mutants and this can increase their residence times close to the tNCP (see Fig. 6.7). It is well seen that R47E and R42E spend lots of time between 30 – 45 Å with R42E having a slight peak around 30 Å. This is a clear indication that these mutants diffuse close to the positive histone core. On the other hand, slight bumps in this area are also seen for K52A and K97A, which could be attributed to the same feature. In experimental conditions *in vivo* the histone core surface may be screened by adjacent nucleosomes without having direct contact with the aqueous environment.

6.2.3 Energetics

Another determinant, which can give insights into the strength of the diffusional encounter complex is the interaction energy between the molecules. The average electrostatic interaction energy of each mutant for conformation 7_0 is given in Fig. 6.8. We see that the most unfavourable encounter complex is formed by the multiple mutants S1S2, 6ALA, SITE1 and SITE2. On the other hand, the most favourable complex is formed by the mutant D65K. Surprisingly, a mutation from positive to neutral residue (K59A, K52A and K55A) leads to a slightly lower interaction energy than for the WT. This means that the residues K52, K55 and K59 do not contribute to binding in agreement with the FRAP experimental study of Brown *et al.* [17]. The order of the single residues mutated to Ala and displaying contribution to binding is K69, R94, R42, R47, K85, R73, K97, R74 and K40. The last residue K40 has been also assigned to be a nonbinding residue [17]. The graph (Fig. 6.8) clearly demonstrates that the multiple mutants of seven (S1S2) and six (6ALA) residues decrease substantially the interaction energy in contrast to the moderate decrease by three (SITE1) and four (SITE2) simultaneous residue mutations. The higher contribution to binding of only three residues in SITE1 (K69, R73, K85) than the four in SITE2 (K40, R42, K52, R94) can be explained by the more important single residues in SITE1, i.e. the binding site K69site.

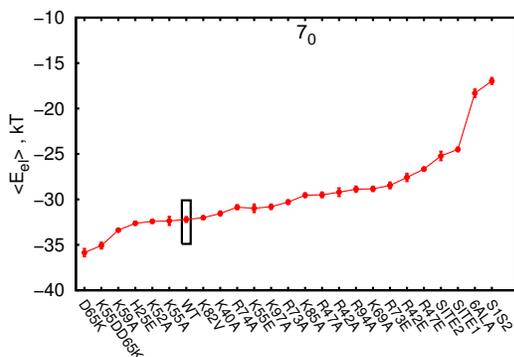


Figure 6.8: Average electrostatic energy of the diffusional encounter complex for each mutant modelled. Highlighted is the energy of the WT, the bars give the standard deviation.

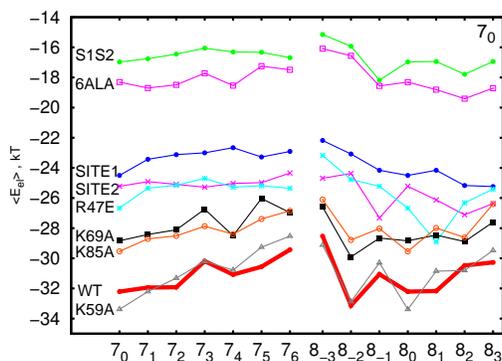


Figure 6.9: Average electrostatic interaction energy versus the nucleosome conformations. Displayed are the energy profiles for the WT and some of the mutants.

The effect of the different nucleosome conformations on the electrostatic interaction energy is laid out in Fig. 6.9. In general, $\langle E^{el} \rangle$ shows a deviation

of maximum $5 k_B T$ for all conformations. The trend observed in Fig. 6.8 for τ_0 is also conserved for the other nucleosome conformations indicating the highest interaction energies for S1S2 and 6ALA. This implies that the energetics of the diffusional encounter complexes is not dramatically influenced by the different conformational changes of the linker DNAs. In first sight this seems confusing, but as it was shown in Sections 6.2.1 and 6.2.2 all mutants form diffusional encounter complexes between the linker DNAs for all nucleosome conformations and this complex is either symmetric or asymmetric with respect to the dyad axis. The interaction energy computed by the effective charges of the mutant moving on the grid of the nucleosome should not deviate much if the mutants are placed close to the nDNA and lDNA at the same time.

6.2.4 Statistics

The binding strength of each residue either to nDNA or lDNA can be determined by the weight factors (Eq. 5.3). The influence of the mutants on the residue weight factors for τ_0 is plotted in Fig. 6.10 and Fig. 6.11. Residues 69 and 97 are belonging to the binding sites K69site and R42site, respectively, whereas residue 59 belongs to the nonbinding site on the loop between helix 2 and helix 3 (Fig. 3.3). Residue 85 is on the loop between the beta sheets close to the hydrophobic Val87 and, thus, these 4 residues represent the surface of the GH5 from different perspectives, which are sufficient to describe the mutational effects on the GH5 orientation with respect to the nucleosome. The weight factors ω_n and ω_l were scaled to the maximum possible weight value ω_{\max} , i.e. on the plots 1 and 0 mean high and low binding strengths, respectively, to either nDNA or lDNA. The points observed on the plots give the binding strength of a certain residue (69, 97, 59 and 85 in Fig. 6.10 and Fig. 6.11) belonging to the labelled mutants either to nDNA (ordinate) or lDNA (abscissa). Only the points deviating from the behaviour observed for the WT are labelled. For example, residue 69 (Lys in WT) binds preferably to the nDNA with high strength in most of the mutant BD simulations except for the multiple S1S2, SITE1 and 6ALA and some single mutations like R42E, R42A, R47E, R94A, K69A. The mutant K69A reduces twice the binding strength to the nDNA, but increases it slightly to the linker DNA. On the other hand, the mutations S1S2, R73E, 6ALA and R42E shift the K69site towards lDNA, R42site together with residue 85 to the solvent, while K59site contacts the nucleosomal DNA for S1S2 and R42E (see Fig. 6.10 and Fig. 6.11). However, mutation of R42 to Ala does not display such a distinguishable effect on the GH5 orientation. Similar behaviour is observed also for the R47E, which rotates K69site to the solvent with K59site and

6.2. BD DOCKING AND ANALYSIS OF THE RESULTS

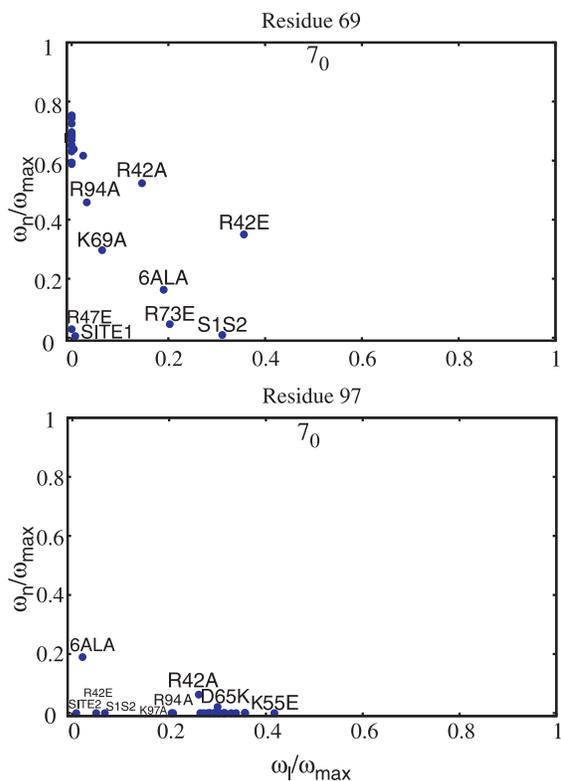


Figure 6.10: Scaled weight factors for residues 69 (upper) and 97 (down) for conformation 7_0 calculated for different docked mutants.

R42site binding simultaneously to lDNA. This suggests that single mutation to glutamic acid (E) on important residues for binding (see Fig. 5.11) can lead to a dramatic orientational change of GH5 with respect to the nucleosome, whereas the replacement of positive to neutral alanine is not sufficient to overcome the positive potential field of the surrounding residues in order to induce a large displacement from the WT position. Single mutants to Ala showing slight influence on the GH5 sites are R42, R94, K69, R42, K97, R47, K85, K40 and K55. The latter three are seen in the plot of residue 85, which spreads mainly on the lDNA site. All these residues have been considered to contribute significantly to binding in the experimental and our WT BD simulation studies [17] (Fig. 5.11) with exception of K40 and K55. The mutants K40A and K55A only reorient the K85 loop between lDNA and nDNA, but do not change the main binding sites.

So far we have discussed the effect of the mutants on the encounter complex formed at nucleosome conformation 7_0 . In order to quantify the binding strength at different nucleosome conformations, we plotted the scaled weight factors for the mutants K69A, K97A, K59A, and S1S2 (Fig. ??, Fig. 6.13, Fig. 6.14). Since the weight factor describes binding strength of a certain residue, we used the residues 69, 97 and 59 as a representatives for each site

CHAPTER 6. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE MUTANTS TO THE NUCLEOSOME

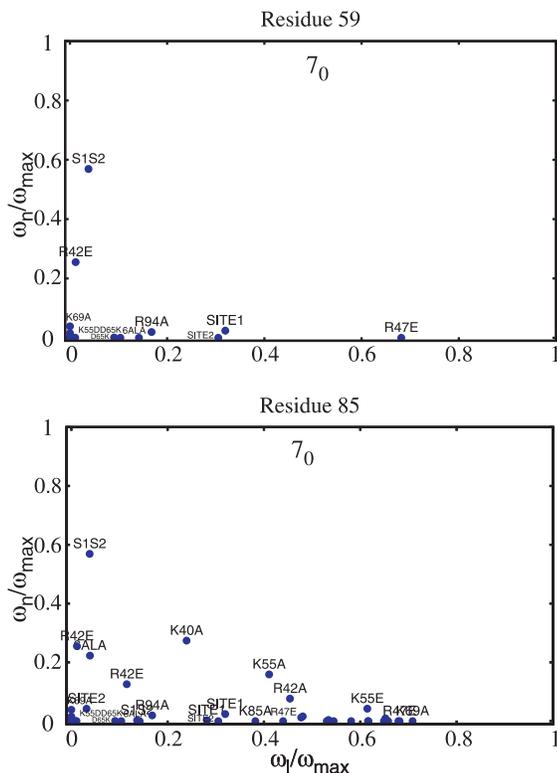


Figure 6.11: Scaled weight factors for residues 59 (upper) and 85 (down) for conformation 7_0 calculated for different docked mutants.

on the surface of GH5. Thus, each plot gives the binding strength of the residue to all of the nucleosome conformations for a given mutant. At the places where point clusters are formed, the labelling is not important because the result does not deviate from the equilibrium conformation 7_0 . For K69A residue Ala69 prefers to bind to nDNA for most of the conformations, but for the open conformations 7_6 , 8_1 , 8_2 , 8_3 K69site is directed to the lDNA (Fig. ??). On the other hand, Lys97 stays close to the lDNA as proposed by the WT docking and only for 7_6 and 8_2 turns to the nDNA. Residue K59 remains exposed to the solvent for all conformations. This means that K69A affects mainly K69site position only for the more open conformations, i.e. this mutation acts locally, and brings a global change for 7_6 . Mutant K97A shows deviation from the dominant WT binding mode for the closed conformations 8_{-2} and 8_{-3} , which shift the R42site to the solvent and K69site to the lDNA (Fig. 6.13). Interestingly, for the most open conformation of mode 8, 8_2 and 8_3 , R42site and K59site exchange their role for binding to lDNA. Mutant K59A also brings an orientational exchange, but between K69site and R42site contacting lDNA and nDNA, respectively, for the extreme open conformations 7_6 , 8_2 and 8_3 and the closed for mode 8 (Fig. 6.13). The data for the single mutations suggest that the mutants slightly reduce the binding

6.2. BD DOCKING AND ANALYSIS OF THE RESULTS

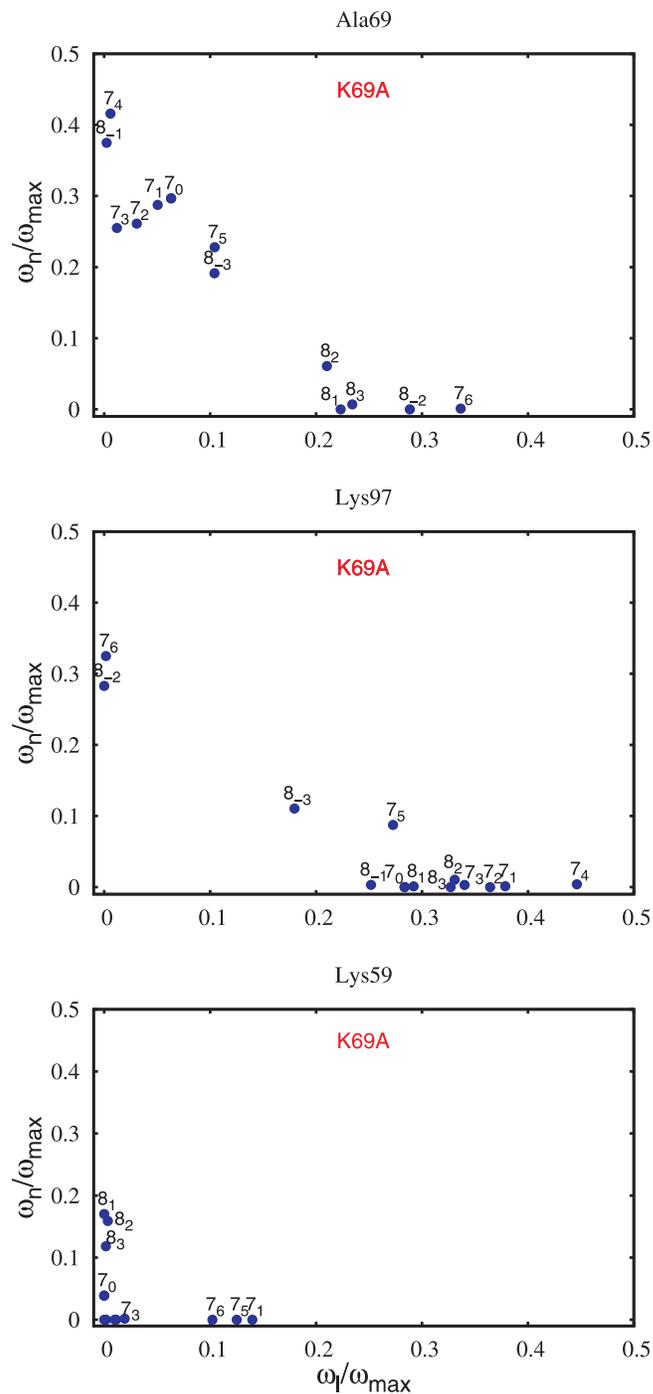


Figure 6.12: Scaled weight factors for the mutant K69A expressed by one residue from each site on GH5 (K69site, R42site, K59site). Each point represents different nucleosome conformation.

CHAPTER 6. BROWNIAN DYNAMICS DOCKING OF THE LINKER HISTONE MUTANTS TO THE NUCLEOSOME

strength of GH5, but does not change its orientation for the most of the nucleosome conformations. Orientational change upon mutation is observed mainly for the extreme conformations of both modes. Mutation of seven

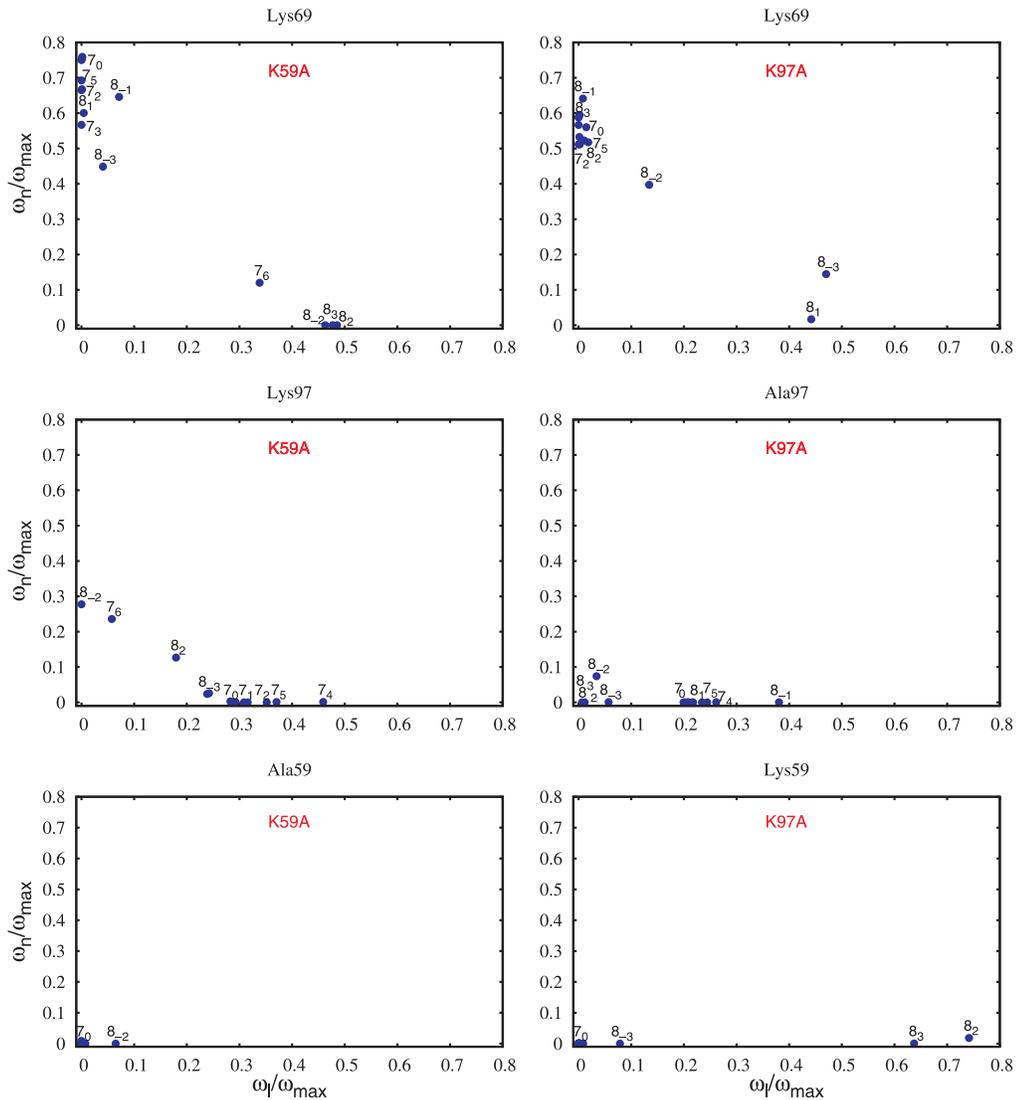


Figure 6.13: Scaled weight factors for the mutants K59A (left) and K97A (right) given by one residue from each site on GH5 (K69site, R42site, K59site). Each point represents different nucleosome conformation.

residues in S1S2 leads to a decrease in the binding strength for all conformations (Fig. 6.14). In S1S2 two residues in the K69site (K69, R73) and two in the R42site (R42, R94) are replaced by Ala. However, the K69site still

detailed classification scheme for the highest docked representative for each mutant giving a general picture for the diffusional encounter complexes. It was shown that the multiple mutations contribute significantly to the different complexes formed, while the single mutants displayed characteristics similar to the dominant binding mode for conformations $7_2 \rightarrow 7_5$ and 8_1 . If one assumed that the single mutants cannot change significantly the binding mode then this observation would suggest that the conformations 7_2 , 7_3 and 8_1 are the most favourable for the linker histone binding. This is also in agreement with the literature, where the crystal structure of the equilibrium conformation 7_0 has been resolved without the presence of the linker histone [4].

Second, the dynamics of the encounter complex formation was represented by the residence time per trajectory within a certain interparticle distance. The simulation data were compared qualitatively with the FRAP experiment [17] and agreed for most of the mutants. However, the mutants R47E and R42E showed longer residence time than the WT, which was attributed to binding close to the positively charged histone core. In experimental conditions the histone core surface might be screened due to the presence of adjacent nucleosomes and, thus, preventing a close contact between R47E, R42E and the nucleosome.

Third, the electrostatic interaction energy revealed the order of the residues stabilizing the encounter complex and pointed out the small energetic deviations for all nucleosome conformations. The findings are in agreement with the FRAP experimental data of Brown *et.al.* [17].

Fourth, zooming into a finer scale of the diffusional encounter complexes the contribution of each positive residue on the surface of GH5 to binding was quantified by weight factors, which are indicators for binding strength. The data showed that the single mutants of positive residues reduce slightly the binding strength to the nucleosome and induce an orientational change for the extreme conformations. The multiple mutants influence significantly the binding mode and a logical trend between all nucleosome conformations cannot be deduced.

In summary, the effect of mutations on the binding mode of GH5 is not additive, i.e. the contribution of multiple mutations to binding is not equal to the sum of the contributions of each single mutant. Single mutants cannot change dramatically the binding position of GH5, but they can induce an orientational change in the diffusional encounter. The dominant binding position is conserved for most of the nucleosome conformations in mode 7, while in mode 8 only 8_1 shows similar characteristics to the results in mode 7. The residues belonging to K69site showed a slightly higher binding strength than the residues constituting the R42site.

7

Association rates of GH5 to the nucleosome by Brownian Dynamics

The formation of biological complexes between proteins, proteins and small molecules, and proteins and nucleic acids is critical to many biological processes, including cell signaling, gene transcription, enzyme catalysis, and the immune response. Molecular association is governed by both the kinetic and the thermodynamic properties of the molecules and of the medium in which they are immersed. Inside a cell, the medium is packed with a wide variety of different molecules and is considered to be crowded. Biomacromolecular complexes vary widely in their affinities and lifetimes, ranging from obligate and permanent to transient and short-lived complexes. In this chapter, only bimolecular association to form a transient complex will be considered. Complexation is usually characterized in terms of affinity, as weak (and loose) or strong (and tight). The variation in affinity is often largely determined by the variation in dissociation rate. Association rates can, however, also vary over many orders of magnitude between complexes and can be critical in the biological context. For example, the snake toxin fasciculin must not only strongly inhibit acetylcholinesterase (an enzyme that is critical to neural transmission) but also reach its target quickly [139]. Similarly, the intracellular inhibitor barstar protects the bacterium *Bacillus amyloquefaciens* from the enzyme barnase, which it excretes to act as an extracellular ribonuclease [140]. Furthermore, the speed at which the lac repressor binds to its chromosomal lac operator regulates gene expression in the cell [141].

The backbone of nucleic acids contains negatively charged phosphate groups. This negative electrostatic potential (Fig. 5.7) leads to attraction of nucleic acids to proteins with positive binding sites. Therefore, the formation of a nucleic acid-protein complex is strongly governed by electrostatic interactions, which enhance the association rate. Such rate enhancement was predicted by applying a method called PARE to an atomistic model of protein-RNA (U1A-U1SLII) interactions [142]. The results based on chang-

ing the ionic strength and making mutations have been shown to be consistent with experiment [142].

In this chapter, the nucleosome-linker histone H5 complex formation is investigated from a kinetic point of view by Brownian Dynamics simulation. Association rates to the dominant diffusional encounter complex are computed for 11 different structures of GH5 with the program SDA5 (Sec. 4.2.2).

7.1 Brownian Dynamics set up

Association rate calculations were carried out for 11 linker histone structures diffusing towards 3 nucleosome conformations (7_0 , 7_6 and 8_3). Conformations 7_6 and 8_3 showed a distinct effect on the diffusional encounter complex and therefore it is reasonable to quantify the effect on association rates on them as well. The linker histone mutants K40A, K55A, K69A, R42A, R47A, R73A, R74A, K85A, R94A, K97A and the WT were used in the simulations. These single mutated residues are part of the positively charged linker histone surface. In addition, the basal association rates¹ of the WT to 7_0 , 7_6 and 8_3 were computed. We will use the notation of 'association rates' for rates computed with forces present and 'basal association rates' for rates computed without forces.

Electrostatic potential calculations

Electrostatic calculations for all structures were performed with a newer version of APBS (APBS 1.1.0) [101]. The input parameters for solving the Poisson-Boltzmann equation on a cubic grid were slightly modified (in comparison with the listed parameters in Chapter 5). The boundary condition was changed from single Debye-Hückel (sdh) to multiple Debye-Hueckel (mdh) sphere and the method by which the point charges are mapped onto the grid was changed from trilinear interpolation using the nearest neighbour grid points (spl0) to cubic B-spline discretization using also the next-next nearest neighbour (spl4). This was aimed at giving better accuracy, although significant differences on the results between both electrostatic calculations should not be observed. All the other parameters remained the same. The electrostatic potentials of the mutants were computed with APBS on smaller grids, 97^3 points (1 Å spacing), in comparison to the much larger grids used (200^3 , 1 Å spacing) in the BD docking computed by UHBD. In such a way the BD computational time can be reduced. The other parameters for the

¹Basal association rate is calculated without forces between the molecules

linker histones were not altered. The parameters for computing the effective charges remained the same.

Electrostatic desolvation calculations

For the association rate computation, the short-range interactions cannot be neglected, because the rates are computed with respect to the fully bound complex. Therefore, the electrostatic desolvation potential accounting for the unfavourable displacement of water upon binding should be included. It was computed on a cubic grid and used in Eq. 4.12 with 220^3 and 60^3 points for the nucleosome and linker histone, respectively. The grid spacing was set to 1 Å and the ionic strength to 100 mM.

Bound complex definition

The diffusional encounter complex, an intermediate state marking the endpoint of diffusion of two biomolecules toward each other, plays an important role in determining the association rates. However, its structure cannot be directly determined experimentally [143]. Usually, the rates are computed for the diffusional encounter complex (assuming the process is diffusion-controlled) using loose native contacts criteria based on an experimentally bound complex (Sec. 4.2.2). In this study we used the structure of the dominant encounter complex determined by the BD (Chap. 5) as a basis for the reaction criteria definition. Such an approximation can be considered to be accurate by means of the highly attractive electrostatic potential between the nucleosome and the linker histone as well as by the ignored repulsive short-range desolvation penalty in the BD docking (Chap. 5).

The donor-acceptor contacts were computed within a distance $d^{\text{pairs}} = 6$ Å. In the simulations only the independent contacts defined by the distance $d^{\text{ind}} = 6$ Å were counted (Sec. 4.2.2). The tNCP-GH5 complex obtained in the conformation 7_0 for the WT was used for the contact definition for all simulations. It had 8 independent out of 1027 dependent pairs.

In the simulations, the short-range repulsive interactions were considered by the desolvation penalty grid, while the hydrophobic desolvation interactions were neglected. The time step was changed from 0.25 ps to 1 ps in comparison with the BD docking runs. The monitoring of the reaction criteria was realized in a window distance of 20 Å ranging from 3 to 23 Å with a step of 0.5 Å. For each system (nucleosome-mutant) the number of runs was set to 2500 and 4 simulations with different random generators were carried out, i.e. in total 10000 trajectories were carried out. For the calculation of the basal rates a single run with 10000 trajectories was performed.

The computational time varied from 113 to 512 CPU hours per simulation. Truncation time t^{hit} was not used in these simulations.

7.2 Association rates results

First, association rate results for the reference nucleosome conformation 7_0 will be shown and after that these results will be compared with the extreme conformations 7_6 and 8_3 . The rate dependence of the ionic strength was not evaluated in the current study and a constant ionic strength of 100 mM was used for all calculations.

7.2.1 Rates to the reference structure 7_0

Figure 7.1 shows the association rates for the WT-nucleosome (7_0) complexation. The computed rates are presented with respect to the distance for a formation of 1, 2, 3 and 4 contacts. The absolute values of the rates are in the range of Smoluchowski rate ($10^9 \text{ M}^{-1}\text{s}^{-1}$) (Sec 2.5.1), which is the upper limit for diffusional association between two uniform spheres without forces between them. Several aspects in regard to Smoluchowski rate should be considered. First, the geometrical shape of the nucleosome is not a sphere and the linker DNAs narrow the binding path of the linker histone.

Although the GH5 resembles a sphere, the geometrical consideration of

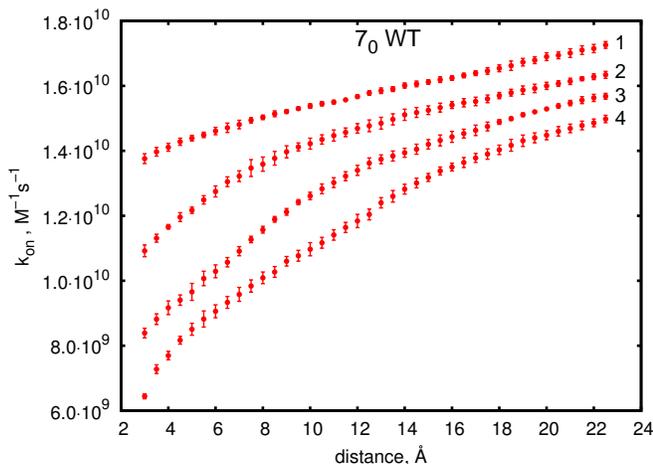


Figure 7.1: Association rates computed for the WT to 7_0 nucleosome conformation with electrostatic forces present. The rates are displayed for a formation of 1, 2, 3 and 4 contacts with respect to the contacts distance criteria. Standard deviations are given in bars.

the lDNAs could give rise to steric clashes resulting in a smaller rate than the Smoluchowski's one. Second, short-range electrostatic desolvation hinders the formation of a fully bound complex, which can reduce the absolute association rates towards the experimental rates as it has been shown for

7.2. ASSOCIATION RATES RESULTS

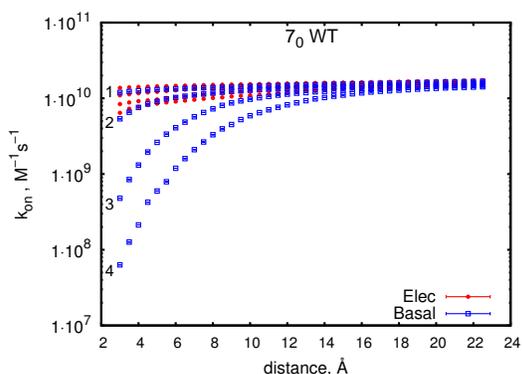


Figure 7.2: Association rates (red) and basal association rates (blue) computed for the WT to 7_0 nucleosome conformation (log scale). The basal rates are displayed for a formation of 1, 2, 3 and 4 contacts with respect to the contacts distance criteria.

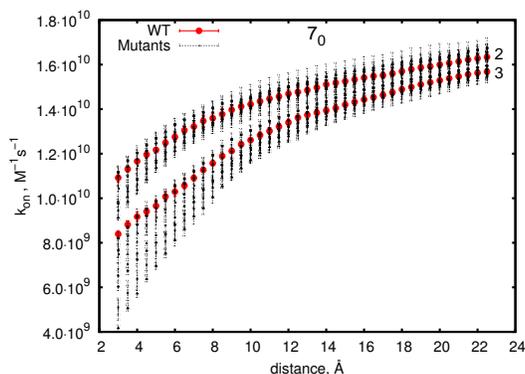


Figure 7.3: Association rates of the WT (red) and 10 mutants (black) for 2 and 3 contacts. Standard deviations are given in bars.

acetylcholinesterase-fasciculin kinetics [104]. On the other hand, since the molecules are oppositely charged and the diffusive process up to the encounter complex is governed by long-range electrostatic interaction, this can lead to high association rates. The basal association rates can be used as a reference point for quantifying the electrostatic contribution to the binding kinetics. Figure 7.2 compares the basal association rates with the rates plotted in Fig. 7.1. Two features are well seen, first, the association rates are higher than the basal association rates and, second, the basal association rates are also in the range of Smoluchowski rate. While the first feature is not surprising and clearly indicates rate enhancement due to the attractive electrostatic interactions, the second observation suggests that even only in the presence of exclusion grids the second protein still binds fast to the nucleosome. In order to quantify this effect, a rough estimate of the basal association rate k_{on} given by Eq. 2.32 was done. In the BD simulations the relative translational diffusion constant was set to $0.0185 \text{ \AA}^2/\text{ps}$ and the basal rate at a center-to-center distance of 62 \AA , which is the position of the diffusional encounter complex, was calculated. The result yielded $k_{\text{on}} = 2.4 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ and this value is comparable with the basal association rate for a formation of 4 contacts at a distance $\sim 4 \text{ \AA}$ (Fig. 7.2). Important to note is that the above calculation is only approximate for the case of two uniformly interacting spheres. Although this is not the case in the simulations, one would expect the shape of the nucleosome to lower the basal rate. However,

CHAPTER 7. ASSOCIATION RATES OF GH5 TO THE NUCLEOSOME BY BROWNIAN DYNAMICS

the basal association rate is higher than the analytical solution. This implies that geometrical considerations should be taken into account in order to understand this observation. In addition, the diffusive entrapment effect [45] can also contribute to an increase of the rates.

Rate enhancement due to electrostatic interactions may result in an increase of 100-fold in k_{on} as it was found for about 25 protein-protein complexes [144]. The rigid body dynamics does not include any conformational rearrangements and motions of the molecules during the diffusional process. In general, there is more than one intermediate state in the association process, because protein-protein binding consists of multiple steps: diffusion, conformer selection and refolding or induced fit [145]. A significant induced fit was found in the case of fasciculin 2 (Fas2) binding to acetylcholinesterase (AChE), two proteins that bind with a very high association rate constant. It was found that the conformation of Fas2 able to bind AChE is not stable in the unbound form of Fas2 and that the association process should follow a conformational change of a stable form of Fas2 that is not complementary to AChE [146, 147]. Hence, the flexibility can also affect the binding kinetics to either reducing or increasing the association rates. Hydrodynamic interactions can have similar effect, but they will not be discussed.

To our knowledge experimental data on the association rates for linker histone binding to the nucleosome are lacking and, hence, we will emphasize on the relative contribution of each mutant to the binding kinetics. Since the diffusional encounter is best described by at least two contacts [143, 148] only the k_{on} for two and three contacts of the WT and mutant proteins are depicted in Fig. 7.3. The WT data are shown in red, whereas the mutants are represented by black. We see that at short distances most of the mutants bind slower than the WT. Increasing the number of contacts leads to divergence of the mutant rates with respect to the WT rates at short distances. The reason is the additional polar (hydrogen) contact which has to be formed and in the absence of one charged residue on the surface of GH5, the probability of such a formation decreases. It is an effect of both, charge and side-chain geometry, which affect the contact formation in the diffusional encounter complex.

The next plot (Fig. 7.4) compares the association rates at 5.5 Å distance for 2 and 3 contacts. At this distance a mutational association rate study of barnase and barstar [149] has shown good agreement between experiment and simulation for 2 contacts and at a distance of 7 Å for 3 contacts. However, in the study [149] electrostatic desolvation was not used. When desolvation penalty to the diffusional process is applied then a distance of 7.5 Å has given the best results for the rates [104]. Here, we are mainly interested in the relative contribution of the rates and therefore the choice of the distance is not of major importance. Surprisingly, in both curves K97A shows faster

7.2. ASSOCIATION RATES RESULTS

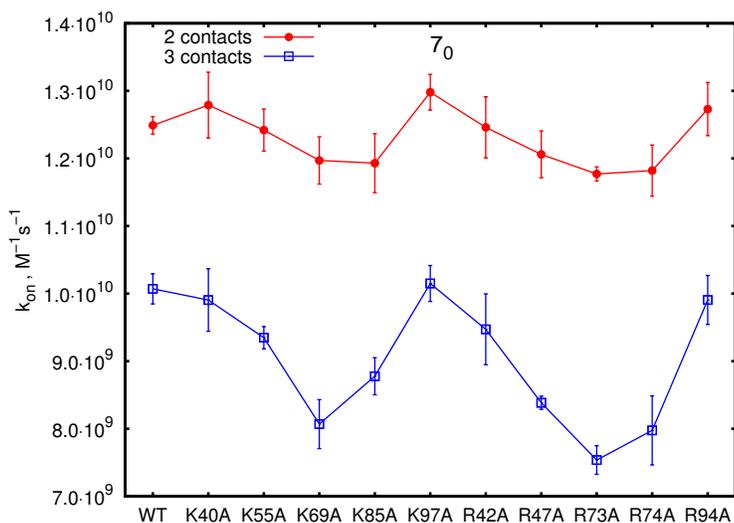


Figure 7.4: Association rates at distance 5.5 \AA for 2 (red) and 3 (blue) contacts are plotted versus the different protein structures. Standard deviations are given in bars.

binding than WT. As it was discussed in Section 6.2.2 residue K97 links the GH5 with its C-terminal domain. Therefore, it might be that K97 does not participate directly to binding, but has only an overall effect to the R42site. This can explain its association rate close to the WT (within the error). Unchanged rates display also K40A and R94A. While the former was considered as nonbinding in the BD docking, the latter is a part of the binding R42site. R94A was also identified as having high binding strength to lDNA (Fig. 5.11) and its low contribution to the binding kinetics is surprising. Also R42A from R42site shows a small relative decrease in the association rate in contrast to K85A, which has the smallest rate for two contacts (Fig. 7.4). On the other hand, the mutants constituting the K69site (K69A, R73A, R74A and R47A) show a distinctively unfavourable behaviour in terms of kinetics with respect to the WT. This means that positive mutations on helix 3 disrupt significantly the kinetics of the linker histone-nucleosome system, although the binding docking mode is slightly affected (Chap. 6). We can speculate that the hardly noticeable relative effect on the association rates for the mutants on the R42site is due to conformational changes on R42site upon binding, which are not taken into account in our model. Site R42site is located on loops, which are more flexible than the alpha helices, where K69site is situated (helix3). Another explanation can be related to the way in which the proteins bind to the DNA (specifically or nonspecifically), which can explain

the observation of association kinetic rates higher than the Smoluchowski rate [150] (Fig. 7.3). These studies suggest three-dimensional (3D) diffusion of the protein to the DNA followed by one-dimensional (1D) diffusion of the protein along the DNA to form a bound complex. This type of diffusion is referred to in the literature as *facilitated diffusion*. Slutsky and Mirny [151] proposed that for an optimal search for the target DNA, a protein should spend half of its time in 3D diffusion and the other half in 1D diffusion, sliding along the DNA. Their study aimed at quantitatively investigating the specific and nonspecific binding of proteins to DNA. However, a theoretical lattice Monte Carlo study [152] of transcription factors (TFs) binding to DNA molecules showed that even if only 15% of the diffusional search time is spent freely in solution, the timescale of target location is consistent with experimental measurements. In this diffusional search, the TFs might exhibit conformational changes, which could affect the association rate constant. Such conformational changes during the searching and sliding mechanism have been investigated to detect the shortest binding time to the DNA consistent with thermodynamics [153]. Also, a protein can jump from one DNA segment to another without dissociating, a process called *intersegment transfer*. In this way, the protein can bind specifically to the target site and a rate enhancement can be observed [154]. Lattice simulations [155] showed that increasing the nucleic acid chain density increases the protein diffusion in the case of intersegment transfer. Moreover, the diffusion coefficient appeared to be reciprocal to the chain density in 1D sliding on the DNA. Consequently, if the GH5 first slid along 1DNA1 with K69site and then binds specifically to the dominant binding position performing a rotation of K69site and R42site towards nDNA and 1DNA1, respectively, then the R42site would come into contact to 1DNA1 only at the end of the 1D diffusion. In this way, the association rates for the mutants on R42site would not deviate much from the WT, something what we observe in Fig. 7.4. Unfortunately, full trajectories were not recorded during the BD runs and we cannot confirm or reject this hypothesis.

7.2.2 Rates to the extreme conformations 7_6 and 8_3

Association rates of the WT binding to three nucleosome conformations are plotted in Fig. 7.5. The extreme conformation 7_6 shows the highest rate, while conformation 8_3 exhibits slower binding behaviour. It should be pointed out that the highest docked position of WT found on conformations 7_6 and 8_3 is located on the 1DNA2 and not in the 1DNA1 (see Chapter 5). The higher rate for 7_6 is reasonable, because the more open conformation would not restrict sterically the approach of the linker histone. Ramzi and Zhou [156]

7.2. ASSOCIATION RATES RESULTS

showed analytically that the association rate for a protein binding nonspecifically by facilitated diffusion to a short linear DNA is higher than to a circular DNA. The linker DNA1 in conformation 7_6 is more straight with respect to IDNA1 in 7_0 (Fig. 5.5). On the other hand, IDNA1 in conformation 8_3 is more bent with respect to IDNA1 in 7_0 , although the dyad is easier accessible. This can explain the differences in the association rates. However, the absolute magnitude of the rates at distance 6 Å varies from 1.17×10^{10} to $1.35 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ for the conformations, which is even smaller including the errors. Generally, the nucleosome conformation does not influence significantly the binding kinetics of the WT, which can be also seen in Fig. 7.6 for the basal association rates. The plot illustrates convergence of the basal association

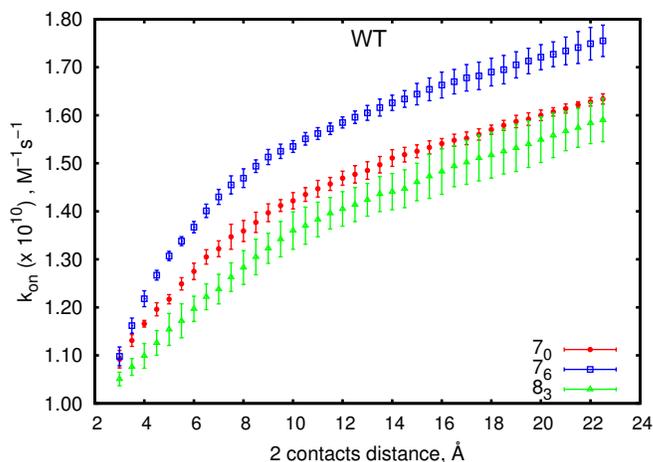


Figure 7.5: Association rates of the WT for 2 contacts for nucleosome conformations 7_0 (red), 7_6 (blue) and 8_3 (green). Standard deviations are given in bars.

rates at small distances for all three nucleosome conformations. At a distance of around 6 Å the basal association rate of the WT is independent on the nucleosome conformations. Comparing the rates with and without electrostatic forces, one can see that the relative association rates between 7_0 , 7_6 and 8_3 are larger than the relative basal association rates. Nevertheless, the magnitude of such a deviation is small and obvious conclusions about the effect of the nucleosome conformations on the WT association rates cannot be drawn.

Figure 7.7 indicates the association rate dependence of linker histone mutants as well as of the nucleosome conformations for 2 contacts formation at 5.5 Å. The curves follow the shape of the association rates of WT in 7_0 for mutants K69A, K85A, K97A, R47A and R74A. The association rates of the other mutants display some deviation from the WT pattern, which does not seem to affect significantly the absolute value of the association rates.

CHAPTER 7. ASSOCIATION RATES OF GH5 TO THE NUCLEOSOME BY BROWNIAN DYNAMICS

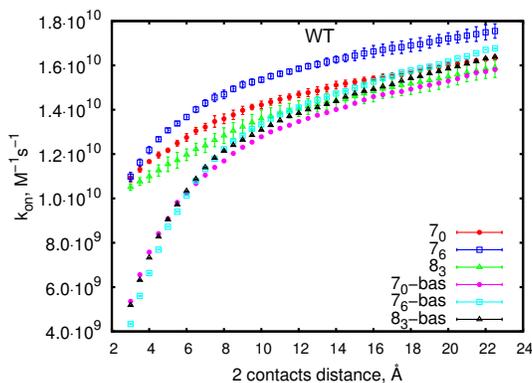


Figure 7.6: Association rates and basal association rates of the WT for 2 contacts for nucleosome conformations 7_0 , 7_6 and 8_3 . Standard deviations are in bars.

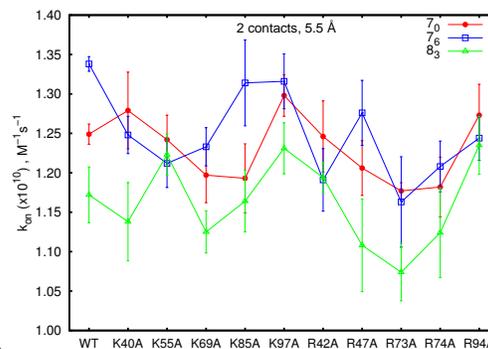


Figure 7.7: Association rates at 2 contacts distance for 11 linker histone mutants and 3 nucleosome conformations at a distance of 5.5 Å. Standard deviations are given in bars.

7.2.3 Summary

The association rates computed in this chapter indicated very fast binding mode of the GH5 to the nucleosome close to the upper analytical limit derived by Smoluchowski (Sec. 2.3). Comparing the association rates with the basal association rates (computed without forces between the molecules) electrostatic rate enhancement due to the highly charged molecules was observed. In addition, the high association rates could be explained by the searching mechanism of DNA binding protein, which has been shown to enhance the rates [150]. The residues participating in the K69site (found to bind to nDNA) showed significant reduction in their association rates upon mutation. On the other hand, mutations of residues belonging to R42site (contacting lDNA) displayed rates very close to the WT rate indicating modest contribution to the binding kinetics. We proposed a hypothesis by which K69site could, firstly, bind to the linker DNA and perform a 1D diffusion along it until the binding site is reached. Then the linker histone could rotate and, thus, K69site and R42site would adjust to nDNA and lDNA, respectively. This mechanism might be also responsible for the rate enhancement and could explain the small contribution of the R42site to it. The computed association rates did not show significant dependence on the nucleosome conformations.

The results obtained revealed only partially which factors influence the binding kinetics of the linker histone to the nucleosome. The complexity of the interactions within chromatin fiber requires experimental data on the linker histone association rates, which the computational methods can be

7.2. ASSOCIATION RATES RESULTS

compared with. At the same time, more investigations are needed in determining the contributions of the C-terminal domain of H5 and the ionic strength dependence of the association rate as well as on the sliding mechanism of the GH5 along DNA.

8

Flexibility in Brownian Dynamics simulation

One of the drawbacks of the Brownian Dynamics method used in this study (Sec. 4.2) is the restriction of the particle motion to only six degrees of freedom - translation and rotation. Large conformational changes during the diffusional process can, however, influence the bimolecular dynamics and interactions as it was shown in the previous chapters. Without violating the approximations taken in the BD algorithm, we introduce a method accounting for large conformational changes on the solute particle in the BD simulation. Firstly, we give a short overview of sampling techniques, then the methodology and implementation in the Brownian Dynamics code is discussed and in the end we show preliminary results on the application to the nucleosome-linker histone complex formation.

8.1 Metropolis Monte Carlo algorithm

Monte Carlo (MC) methods are stochastic computational methods applied to big systems having a large number of coupled degrees of freedom. In general, the MC algorithm relies on a random number picked at each simulation step and according to it moving the system to a new state in the phase space. This method is important in statistical physics, for example, for computing thermal averages. However, in order to be accurate, this sampling should comprise a large region in the phase space, i.e. a large number of states should be generated randomly. When the generated states are not completely random, but follow a probability function, which assists the system towards its thermal equilibrium, then only a representative set of states can be sufficient to describe the system properties well. This sampling is called *importance sampling*.

If a configuration in the phase space is denoted by X then the probability

of this configuration can be expressed by a Boltzmann factor

$$p(X) \sim \frac{1}{Z} \exp\left(-\frac{E(X)}{k_B T}\right), \quad (8.1)$$

where $E(X)$ is the energy of the system in this particular configuration X and Z is the partition function (Eq. 2.4). Configurations with probabilities given by Eq. 8.1 can be generated either completely randomly or in a successive way $X_n, X_{n+1}, X_{n+2}, \dots$. The latter approach is called *Markov chain* and each configuration depends only on the previous one. The initial configuration X_0 is randomly chosen and the next one is constructed through a transition probability $T(X_{n-1} \rightarrow X_n)$, which gives the probability to move from conformation X_{n-1} to X_n . For this chain of configurations with equilibrium probability p^{equil} , the *detailed balance* condition should hold

$$p^{\text{equil}}(X_n)T(X_n \rightarrow X_m) = p^{\text{equil}}(X_m)T(X_m \rightarrow X_n). \quad (8.2)$$

The above equation implies that the sampling path is reversible, which means that the probability of any configuration remains the same during the simulation. If $T(X_n \rightarrow X_m) \neq 0$ and $T(X_m \rightarrow X_n) \neq 0$ then Eq. 8.2 can be rewritten in the form

$$\frac{T(X_n \rightarrow X_m)}{T(X_m \rightarrow X_n)} = \exp\left(-\frac{E(X_m) - E(X_n)}{k_B T}\right). \quad (8.3)$$

When a step from configuration X_n to X_m is not possible, the transition probability reads $T(X_n \rightarrow X_m) = 0$. The reverse relation $T(X_m \rightarrow X_n) = 0$ should be fulfilled as well, which means that any two arbitrary configurations of the system must be adjacent in order to have a successful step. Thus, the probability ratio between both configurations (states) depends on the energy difference between them.

An algorithm based on the Markov chain process, called *Metropolis method* [157], takes advantage of the probability energy dependence of each state and evaluates whether a new step X_m can be accepted or rejected (X_n) as follows:

- if $E(X_m) < E(X_n)$ then the state with energy $E(X_m)$ is accepted
- if $E(X_m) > E(X_n)$ then the probability ratio is evaluated

$$P_{nm} = \frac{p(X_m)}{p(X_n)} = \exp\left(-\frac{E(X_m) - E(X_n)}{k_B T}\right), \quad (8.4)$$

and a random uniform number $b \in [0, 1]$ is drawn out and if $P_{nm} > b$ then the state with energy $E(X_m)$ is accepted, otherwise it is rejected.

The way arbitrary adjacent states are chosen does not affect the final equilibrium distribution as long as the detailed balance holds. However, it will have an effect on how fast this equilibrium is reached and the more states are generated, the longer the time will be.

8.2 Methodology

The flexibility of biomolecules is difficult to model in atomic detail when biological processes at time scales larger than hundreds of nanoseconds are of interest. In Brownian Dynamics, the second protein performs Brownian steps according to the Ermak-McCammon algorithm (Eq. 4.4) (see Fig. 4.1), where it is assumed that the time step Δt is much larger than $\tau_v = mD/k_B T$, where τ_v is the momentum relaxation time. The parameter m is the particle mass and D is the diffusion constant. Actually, only the time for analysis of BD has to satisfy this requirement. This requirement says that the positional relaxation time is much longer than the velocity relaxation time ensuring a diffusional process takes place. On the other hand, the time step Δt must not exceed the time at which the forces and torques cannot be considered constant. Hence, the BD dynamics procedure is highly dependent on the choice of the time step, which is crucial for performing physically correct sampling of the configurational space. During the diffusional process in the BD, the time step Δt is defined by Δt_1 and Δt_2 , where $\Delta t_2 > \Delta t_1$. The time step Δt_2 aims at reducing the computational time at sampling in regions which are far away from the target location, i.e. it has a value several times larger than Δt_1 at distance x_2 . The actual value of Δt_2 varies linearly with the distance, approaching Δt_1 at distance x_1 ($x_1 < x_2$). Therefore, not only the time steps Δt_1 and Δt_2 , but also the distances x_1 and x_2 are important parameters influencing the computational time as well as the sampling accuracy.

A method by which the flexibility of the target molecule can be taken into account without affecting the separation of slow and fast processes in the BD is developed. The general idea is as follows: the energy landscape of a biomolecule has many local and global minima, which stand for different conformational states. These conformational states can be adopted by the biomolecule at any time with different probabilities. If we know that two conformational states are separated by a barrier the crossing of which requires long time sampling, e.g. ns or μ s time scale, then we can neglect the transition path and look only at the two conformational states. If next, the biomolecule again moves to another long living state, which depends only on its previous state then this motion can be described as a Markov chain process. However, it might appear that the molecule is in a long living state

between two other such states. Consequently, if we consider a three points chain and the biomolecule is positioned on the middle point (state) then it can move either right, left or stay in the same state if the barriers to cross are too high. In this way the conformational states can be represented as discrete states in the configurational space, where the biomolecule jumps from state to state with the same rate. Thus, if these states are known in advance from computational techniques for modelling flexibility like MD, NMA, etc., they can be incorporated in the BD procedure as discrete objects which exchange with each other with a certain rate. Details about the implementation are given below.

8.2.1 Implementation

The new method was implemented in the SDA5C, a version of SDA, a Brownian Dynamics docking program, recording complexes satisfying predefined constraints (Sec. 4.2.2). The programming code of SDA is *Fortran*. The target molecule, molecule 1, can be represented by different conformations obtained by other methods. Then, all these conformations can be treated as different molecules and grids and charges can be assigned to each of them. The idea is that the second molecule, molecule 2, will diffuse towards molecule 1 and, after a certain time t^{conf} spent within a center-to-center distance less than x_1 , molecule 1 will change its conformation if certain criteria are satisfied (discussed below), i.e. different exclusion and force grids will be assigned to it. This process will be repeated until the simulation ends. The way it is done is as follows: initially, the number of the conformations ($ngrid^1$), i.e. the grids, and the initial conformation ($nstart$) of molecule 1 are assigned. In addition, two other parameters are given at input; the time t^{conf} at which a new conformation for the region of the time step Δt_1 is chosen² and the method ($gmethod$) by which a new grid (conformation) is selected. The numbers $nstart$ and $ngrid$ are important, because the first gives the starting molecule 1's conformation for each trajectory, while the second yields the order by which the conformation's selection algorithm is applied. At each t^{conf} , the adjacent conformations (exclusion grids) of molecule 1 are checked for overlaps with the current state $\Omega(\mathbf{r}, \boldsymbol{\theta})$ of molecule 2, and those with overlaps are rejected from the selection algorithm. Let the current conformation of molecule 1 at time t^{conf} has an interaction energy $E^i(\mathbf{r}_i)$ with the molecule 2. Having the molecule 2 fixed, its interaction energies with the adjacent conformations of molecule 1, if available, are $E^{i-1}(\mathbf{r}_{i-1})$ and $E^{i+1}(\mathbf{r}_{i+1})$. The

¹ $ngrid \geq 1$

²Only if selection criteria are satisfied, see below

selection algorithm is applied only if the number of new possible conformations at time t^{conf} is greater than zero. Four different methods for selecting a new conformation were implemented:

Energy algorithm ($gmethod = 1$)

When an *energy algorithm* is chosen ($gmethod = 1$) the conformation with the lowest interaction energy $E^{\text{min}}(\mathbf{r})$ is selected,

$$E^{\text{min}}(\mathbf{r}) = \min [E^{i-1}(\mathbf{r}_{i-1}), E^i(\mathbf{r}_i), E^{i+1}(\mathbf{r}_{i+1})] \quad (8.5)$$

and the BD step is performed. This algorithm samples the configuration space towards an energy minimum. However, it has the disadvantage of biasing the system downhill to a local minimum from which it cannot escape.

Modified Metropolis algorithm ($gmethod = 2$)

The *modified Metropolis algorithm* ($gmethod = 2$) uses the standard Metropolis method (Sec. 8.1) when only two conformations $E^i(\mathbf{r}_i)$ and $E^{i-1}(\mathbf{r}_{i-1})$ (or $E^i(\mathbf{r}_i)$ and $E^{i+1}(\mathbf{r}_{i+1})$) are available. When three conformations, $E^{i-1}(\mathbf{r}_{i-1})$, $E^i(\mathbf{r}_i)$ and $E^{i+1}(\mathbf{r}_{i+1})$ exist then the modified Metropolis method is performed

- if $E_{\text{adjacent}}^{\text{min}}(\mathbf{r}) = \min [E^{i-1}(\mathbf{r}_{i-1}), E^{i+1}(\mathbf{r}_{i+1})] \equiv E^{\text{min}}(\mathbf{r})$ then $E^{\text{min}}(\mathbf{r})$ is accepted
- if $E^i(\mathbf{r}_i) \equiv E^{\text{min}}(\mathbf{r}) = \min [E^{i-1}(\mathbf{r}_{i-1}), E^i(\mathbf{r}_i), E^{i+1}(\mathbf{r}_{i+1})]$ then Metropolis algorithm is applied for conformations with energies $E^i(\mathbf{r}_i)$ and $E_{\text{adjacent}}^{\text{min}}(\mathbf{r})$.

This method is more efficient than the *energy algorithm* when the configuration space is large.

Random energy algorithm ($gmethod = 3$)

The *random energy algorithm* ($gmethod = 3$) prohibits the current conformation $E^i(\mathbf{r}_i)$ and the system always moves to a new state. If there are two new conformations $E^{i-1}(\mathbf{r}_{i-1})$ and $E^{i+1}(\mathbf{r}_{i+1})$ then the state with energy $E_{\text{adjacent}}^{\text{min}}(\mathbf{r}) = \min [E^{i-1}(\mathbf{r}_{i-1}), E^{i+1}(\mathbf{r}_{i+1})]$ is selected. This algorithm has the disadvantage that it can be also trapped in a local minimum and fluctuate between two states without convergence.

Random algorithm ($gmethod = 4$)

The *random algorithm* ($gmethod = 4$) relies on a uniform random number $b \in [0, 1]$ generated at time t^{conf} . A new conformation is chosen with probability of $p_{\text{three}} = 1/3$ and $p_{\text{two}} = 1/2$ among three and two conformations, respectively. The *random algorithm* does not include computation of interaction energies at each t^{conf} and it makes it faster than the previous three selection procedures. However, it is not based on rigorous physics.

8.2.2 Discussion and limitations

The method described can be applied to a large range of biomolecules, which undergo conformational changes upon interaction with other biomolecules. There is no restriction on the number of conformations ($ngrid$), although very large biomolecules may cause computational memory problems. Equally time distributed conformations are reasonable for strongly interacting molecules, while for weak molecular interactions this is not necessary. In principle, due to the discretization of the conformational states the ideal physical situation for more accurate BD run would be when the transition path between the states is much shorter in time than the time spent within a certain conformation. This implies that the transition rate is very high. Then one would not worry about the influence of the intermediates, not sampled, states on the dynamics during BD simulation. Such states can be sampled if the resolution of the conformational path is increased, but, first, it would not be computationally feasible and, second, the increased resolution have a limit: the time $t^{\text{conf}} \geq \Delta t_1$. This limit cannot be overcome without violating the approximations taken in the BD algorithm. It is worth mentioning that for a sufficient long simulation and for a large number of $ngrids$ convergence to the equilibrium encounter complex should be achieved. In addition to the interaction energies and complexes recorded, the new version of SDA, records the number of successive trials³, the trajectory number, the interaction energies (random numbers for $gmethod = 4$ instead), the simulation time and the conformation at each t^{conf} .

³'Successive' trial means when a selection algorithm during the run is applied regardless of the outcome.

8.3 Application to the nucleosome and the linker histone

The method described was applied to tNCP-GH5 BD docking using SDA5C program. Nucleosome conformations were chosen from the NMA analysis (Sec. 5.2). Five conformations were chosen from mode 7 ($7_0, 7_1, 7_2, 7_3, 7_4$) and mode 8 ($8_{-2}, 8_{-1}, 8_0, 8_1, 8_2$). If we assume that the absolute values of the eigenfrequencies are not arbitrary, but have physical meaning, then the period \mathcal{T} for each mode can be calculated and the time distribution between the conformations evaluated. The eigenfrequencies of mode 7 and 8 are $\omega^7 = 1.69 \text{ cm}^{-1}$ and $\omega^8 = 1.46 \text{ cm}^{-1}$, respectively. Then, the periods read

$$\mathcal{T}^7 = \frac{1}{c\omega^7} = 19,7 \text{ ps} \quad \mathcal{T}^8 = \frac{1}{c\omega^8} = 14.9 \text{ ps}, \quad (8.6)$$

where c is the speed of light. On each harmonic approximation 15 conformations were derived from the NMA (Fig. 5.3) having 14 time intervals between them. Then each time interval is given by

$$\Delta t_{\text{NMA}}^7 = 1.41 \text{ ps} \quad \Delta t_{\text{NMA}}^8 = 1.06 \text{ ps}. \quad (8.7)$$

These time steps are very close to the time step $\Delta t_1 = 1 \text{ ps}$, which is usually used in the BD simulation. This calculation estimates that the transition from one conformation to another is approximately 25% slower in mode 7 than in mode 8. It also suggests that two BD runs for mode 7 and 8 should be run with different t^{conf} in order to have accurate relative dynamics between the modes. However, as it was already discussed solvent damping and anharmonic effects might contribute significantly to NMA dynamics. Intuitively, this contribution would result in longer time intervals between the different conformations. Since we observe large conformational changes of the linker DNAs we can assume that these conformations happen in the nanosecond time scale. In order to test the new method accounting for flexibility during a BD run, three different times $t^{\text{conf}} = 1, 10, 100 \text{ ns}$ were assigned and for each of them three selection algorithms $g\text{methods} = 1, 2, 4$ were applied. In total 18 BD runs were carried out with the same input parameters as in SDA4C. The relevant parameters were $x_1 = 130 \text{ \AA}$, $x_2 = 180 \text{ \AA}$, $\Delta t_1 = 0.25 \text{ ps}$ and $\Delta t_2 = 20 \text{ ps}$. In the simulations conformation 7_0 was chosen as a start conformation ($n\text{grid} = 1$) and 7_4 was the last ($n\text{grid} = 5$) in mode 7. In mode 8 conformation 8_0 was selected for beginning of each run ($n\text{grid} = 3$), 8_{-2} ($n\text{grid} = 1$) and 8_2 ($n\text{grid} = 5$) were the extreme conformations.

First, results using $g\text{method} = 1$ will be shown. Each point on the plots indicates the conformation selected after applying a selection algorithm. Figure 8.1 and Fig. 8.2 give the interaction energy (electrostatic) for mode 7

between the GH5 and a nucleosome conformation (selected) at a center-to-center distance d and at the simulation time, respectively. Each plot illustrates three BD runs ($t^{\text{conf}} = 1, 10, 100\text{ns}$). According to the *energy algorithm* the energy E^{el} is the lowest energy conformation (labelled on the plot) out of the conformations participating in the selection algorithm. It is well seen that in the simulation with the shortest transition time (1 ns), the largest number of trials is performed. For the three simulations the energy tends to decrease with the distance and at a certain distance no more selection trials are performed. Actually this is the point where either the simulation is truncated due to reached time of $t^{\text{hit}} = 0.2\text{ ms}$ for all sampled trajectories or due to overlaps no more trials are possible. Here the plots show the data of the last trajectory (3) in the simulation, which is the only one with selection trials performed. This plot (Fig. 8.2) shows only the energy distribution for all trials, but does not show the order of the selected conformations. The trial energy dependence of the simulation time is given in Fig. 8.2. The GH5 approaches the nucleosome searching for an energetically favourable nucleosome conformation, which for 1 ns , 10 ns , 100 ns are conformations 7_0 ($ngrid = 1$), 7_2 ($ngrid = 3$) and 7_4 ($ngrid = 5$), respectively. It means that the choice of the transition time is important for the final result, although for very long simulations with a large number of trajectories the data should converge to a single conformation. Figure 8.2 is only indicative and does not aim at showing the chosen conformation (labelled points) at each trial. In Fig. 8.9 this will be discussed and shown at higher resolution for $t^{\text{conf}} = 1\text{ ns}$. Around a microsecond simulation time for 1 ns transition time, the GH5 was within the distance x_1 and many trials were done. Generally, both plots indicate that the selection algorithm assists the GH5 to find the most favourable path to the encounter complex.

Figure 8.3 and Fig. 8.4 display similar behaviour for mode 8 with less selection trials performed for each run⁴. The reason might be more overlaps due to the more closed conformations 8_{-2} and 8_{-1} . In mode 8, the linker histone finishes its diffusional motion on conformations 8_2 , 8_2 and 8_1 for the runs with $t^{\text{conf}} = 1\text{ ns}$, $t^{\text{conf}} = 10\text{ ns}$ and $t^{\text{conf}} = 100\text{ ns}$, respectively. In this mode, GH5 prefers the more open conformations than the equilibrium one 8_0 .

The *modified Metropolis algorithm* ($gmethod = 2$) gives slightly better convergence towards the minimum state (Fig. 8.5, Fig. 8.6, Fig. 8.7, Fig. 8.8) for both modes, where again the trials of the last trajectory (3) are plotted. The last accepted conformations for modes 7 and 8 are 7_4 , 7_2 , 7_4 and 8_1 , 8_2 and 8_1 for 1 ns , 10 ns , 100 ns , respectively. These results again indicate that

⁴Here also the last trajectory (3) is plotted.

8.3. APPLICATION TO THE NUCLEOSOME AND THE LINKER HISTONE

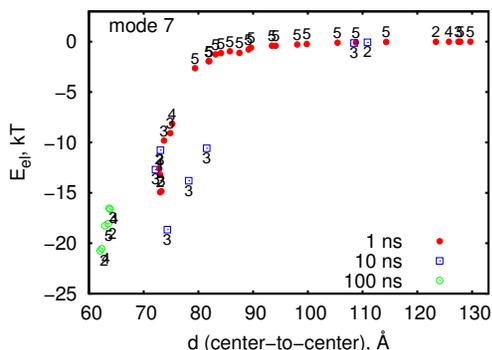


Figure 8.1: Interaction electrostatic energy dependence versus the distance d for $gmethod = 1$ and three different times t^{conf} . Each point gives a performed trial for a certain conformation (labelled) of mode 7.

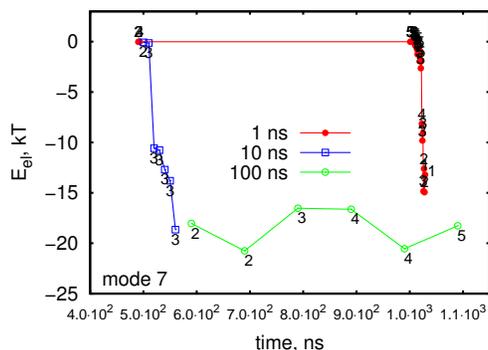


Figure 8.2: Interaction energy versus the simulation time for $gmethod = 1$ and three different times t^{conf} . Each point is a trial with certain outcome (labelled) of mode 7.

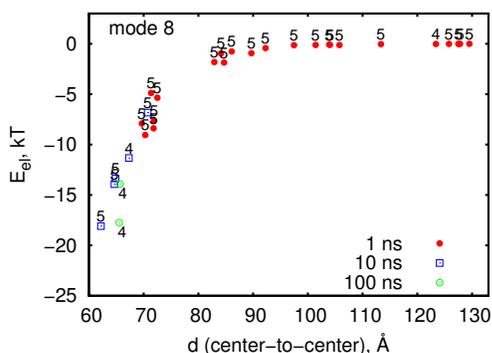


Figure 8.3: Interaction electrostatic energy dependence versus the distance d for $gmethod = 1$ and three different times t^{conf} . Each point gives a performed trial for a certain conformation (labelled) of mode 8.

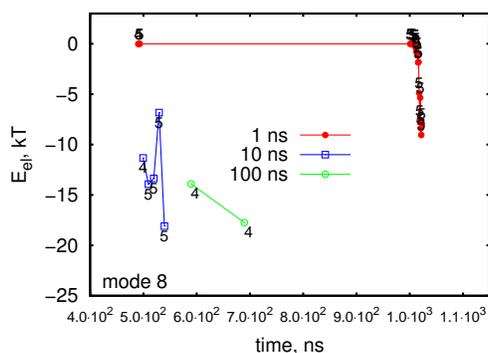


Figure 8.4: Interaction energy versus the simulation time for $gmethod = 1$ and three different times t^{conf} . Each point is a trial with certain outcome (labelled) of mode 8.

CHAPTER 8. FLEXIBILITY IN BROWNIAN DYNAMICS SIMULATION

the GH5 prefers to bind to more open conformation than the equilibrium one. It shows that for mode 7 both algorithms lead to the same nucleosome conformation for transition times 1 ns and 10 ns. The same is observed for mode 8 as well: 8₂ and 8₁ for 1 ns and 10 ns, respectively. This implies that for longer t^{conf} the *energy* and the *modified Metropolis algorithms* behave similarly.

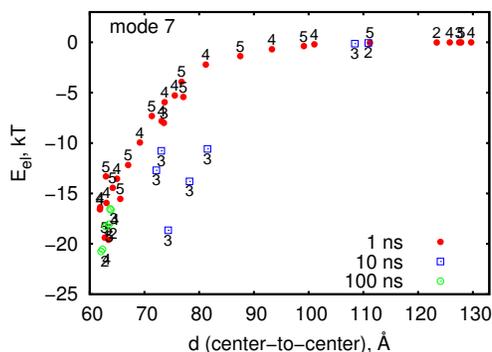


Figure 8.5: Interaction electrostatic energy dependence versus the distance d for $gmethod = 2$ and three different times t^{conf} . Each point gives a performed trial for a certain conformation (labelled) of mode 7.

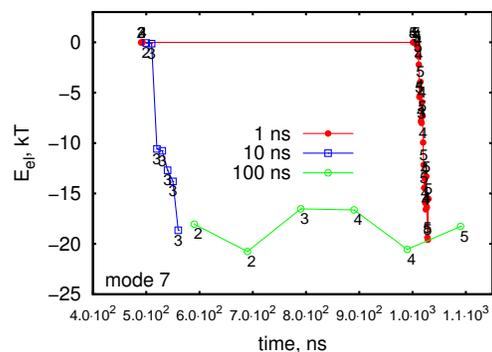


Figure 8.6: Interaction energy versus the simulation time for $gmethod = 2$ and three different times t^{conf} . Each point is a trial with certain outcome (labelled) of mode 7.

A comparison between both methods for $t^{\text{conf}} = 1 \text{ ns}$ at simulation times close to the end of the simulation is shown on Fig. 8.9 and Fig. 8.10 for modes 7 and 8, respectively. At first sight one can see the energy difference of the chosen conformations between mode 7 and 8. The latter displays higher interaction energies than the former, which can be attributed to the binding path differences between both modes. In the simulation time range depicted, the linker histone quantifies its dominant binding mode deduced in Chapter 5 ($\sim 60 \text{ \AA}$) only for mode 7 using $gmethod = 2$ (see Fig. 8.5). This can explain the lower energy of $gmethod = 2$ in the sampling path of mode 7 in comparison with $gmethod = 1$ (Fig. 8.9). Actually, for longer simulation time both curves should converge. The same should be valid for mode 8 as well, although $gmethod = 2$ displays longer sampling than $gmethod = 1$ (Fig. 8.10). However, the selection trials are ended at closer distance to the nucleosome for $gmethod = 1$ than $gmethod = 2$ (Fig. 8.3 and Fig. 8.7). It is difficult to deduce which method is better in this case, but both sample at the end only conformations 8₁ and 8₂. Conformation 8₂ was found to have a different binding mode from the dominant one, which might be the reason

8.3. APPLICATION TO THE NUCLEOSOME AND THE LINKER HISTONE

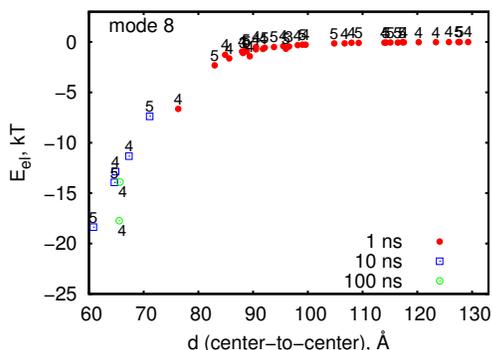


Figure 8.7: Interaction electrostatic energy dependence versus the distance d for $gmethod = 2$ and three different times t^{conf} . Each point gives a performed trial for a certain conformation (labelled) of mode 8.

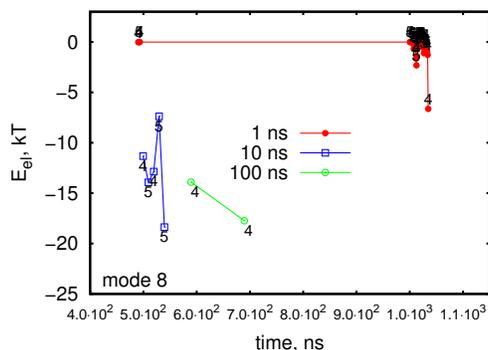


Figure 8.8: Interaction energy versus the simulation time for $gmethod = 2$ and three different times t^{conf} . Each point is a trial with certain outcome (labelled) of mode 8.

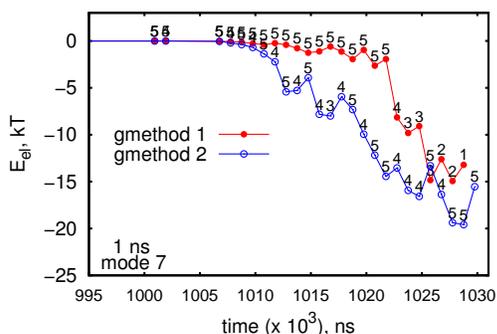


Figure 8.9: Comparison between $gmethod = 1$ (red) and $gmethod = 2$ (blue) for mode 7 with $t^{conf} = 1 ns$. The interaction energy dependence with respect to the simulation time is depicted.

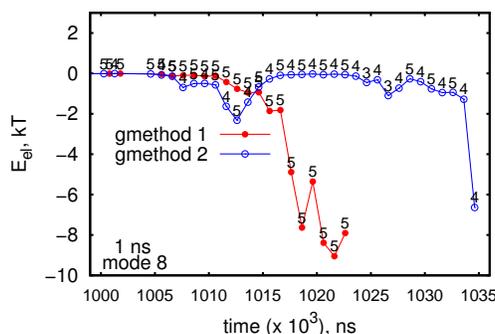


Figure 8.10: Comparison between $gmethod = 1$ (red) and $gmethod = 2$ (blue) for mode 8 with $t^{conf} = 1 ns$. The interaction energy dependence with respect to the simulation time is depicted.

CHAPTER 8. FLEXIBILITY IN BROWNIAN DYNAMICS SIMULATION

for the behaviour observed in Fig. 8.10.

The *random algorithm* ($gmethod = 4$) used also indicated attraction of the GH5 towards the nucleosome. However, in contrast to the previous two applied, it does not follow the interaction energy landscape of the biomolecules. Only results for transition time of 1 ns close to the last selected conformations for mode 7 and 8 are shown in Fig. 8.11 and Fig. 8.12, respectively. The graphs display the dependence of the center-to-center distance d on the simulation time. Clearly seen is that the GH5 moves closer to the nucleosome and reaches final conformations of 7_0 and 8_1 for modes 7 and 8, respectively. This is in agreement with the result of the $gmethod = 1$ for mode 7 and $gmethod = 2$ for mode 8. Interestingly, the plot of mode

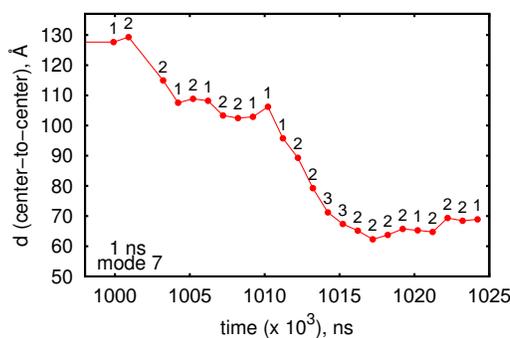


Figure 8.11: Distance-time dependence of the selection trials for $gmethod = 4$ with 1 ns transition time for mode 7.

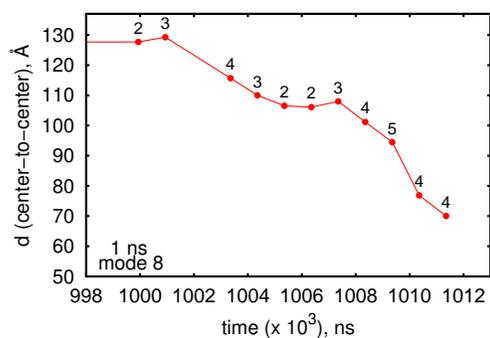


Figure 8.12: Distance-time dependence of the selection trials for $gmethod = 4$ with 1 ns transition time for mode 8.

7 (Fig. 8.11) illustrates that conformations 7_3 and 7_4 are not invoked at all, although conformation 7_2 was chosen two times. The reason, apart from the randomness, might be that conformations 7_3 and 7_4 overlap with the linker histone. Here again the linker histone approaches the nucleosome at a smaller selection distance for mode 7 than mode 8. It suggests that mode 8 is more restricted in terms of configurations than mode 7.

At the end, we are interested in the binding mode of the linker histone and how it is influenced by the different algorithms applied. Due to the more selection trials for $t^{\text{conf}} = 1$ ns we will compare the binding modes only for these BD runs. The recorded complexes were 500 and they were clustered according to the procedure described in Chapter 5. The highest docked representatives were compared to the dominant binding mode. Remarkably well were reproduced the binding position and orientation of the GH5 with respect to the obtained nucleosome conformation. Only for mode 8 and

$gmethod = 1$, the obtained position of the GH5 on conformation 8_2 was on linker DNA2, which is actually what was obtained by the 'standard' BD docking. The average interaction energies at the docking position for all algorithms used do not deviate much from the dominant binding mode and they are around $30 k_B T$. This indicates, as seen on the plots, that after the last selection trial the GH5 continues searching for a more favourable position on the nucleosome. This might imply that at a certain position, the linker histone does not allow much conformational freedom of the nucleosome in agreement with the results on the NMA (Sec. 5.6), where GH5 suppresses the linker DNA1.

8.4 Summary

In this chapter, a new method for treating flexibility in all-atom Brownian Dynamics was developed. One of the interacting molecules is represented by a set of conformations, which can exchange between each other with a constant rate during BD simulation. Four algorithms for selecting a conformation were implemented: the *energy algorithm* selects the conformation with the lowest interaction energy between the interacting molecules, the *modified Metropolis algorithm* uses the standard Metropolis method with a slight modification for a choice between three conformations, the *random energy algorithm* selects an adjacent conformation according to its interaction energy and the *random algorithm* uses a random uniform number to select a conformation. The conformation selection is invoked within a certain distance after a specified transition time. There are no restrictions on the number of conformations as well as on the transition time as long as it is computationally feasible and the approximations of the Brownian Dynamics algorithm hold, respectively.

The method was applied to the tNCP-GH5 system and the nucleosome was represented as a set of conformations obtained by the NMA. The method applied reproduced remarkably well the docked results obtained in Chapter 5. This is an indication that the new implementation, although, with less statistical confidence can lead to the same docking position in reasonable computational time. However, only in two out of 18 BD runs the final nucleosome conformation was the reference 7_0 . The linker histone was docked 5 times on conformation 8_1 , 4 times on conformations 7_4 and 8_2 and 3 times on conformation 7_2 . Although with short sampling this outcome supports the hypothesis that the GH5 prefers more open conformations than the equilibrium 7_0 , which has been resolved without the linker histone [4]. For longer sampling, it is clear that the new methodology should give better results. It cannot be deduced which selection method is better, but definitely over a

CHAPTER 8. FLEXIBILITY IN BROWNIAN DYNAMICS SIMULATION

long simulation Metropolis method should lead to global convergence.

9

Molecular Dynamics of the chromatosome particle

The diffusional encounter complex of the nucleosome-linker histone H5 complex obtained by the Brownian Dynamics was further refined to higher resolution using Molecular Dynamics. In this chapter we show preliminary data on the dynamics of the nucleosome-linker histone complex on nanosecond time scale.

9.1 MD setup

The MD simulations were carried out with the NAMD software package [107] using Amber99SB force field [106]. Details on the numerical procedure are given in Section 4.3. The structure of the dominant binding mode of GH5 together with the nucleosome (Chap. 5), called chromatosome particle, was used as a starting conformation of the MD setup. It has 25027 atoms including hydrogens. The structure was neutralized by adding 226 Na⁺ ions and was placed in a truncated octahedron box with explicit waters, 50 Na⁺ ions and 50 Cl⁻ ions. The ions represent the system at physiological conditions. The solvation was done with the Leap program from the Amber software package [158] using TIP3PBOX [159] with 4 Å solvation layer. In total the system contained 198 303 atoms. Periodic boundary conditions (PBC) were applied during the simulation. The particle mesh Ewald (PME) algorithm for treatment the full electrostatic interactions with 1 Å grid spacing was used.

Minimization

The system was minimized in 11 subsequent runs with different force constants (500, 100, 50, 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0 kcal/molÅ²) for the constrained heavy atoms. Each minimization run was carried out for 1000 cycles using

CHAPTER 9. MOLECULAR DYNAMICS OF THE CHROMATOSOME PARTICLE

the steepest descent method to relieve bad steric interactions (100 cycles) followed by the conjugate gradient (900 cycles) to let the system go downhill to the energy minimum.

Equilibration

Three subsequent equilibration runs for 25, 100 and 250 ps, respectively, were performed on the system. The time step was set to 1 fs. In equilibration 1 and 2 all non-hydrogen atoms and not water were restrained and *NVT* and *NPT* (Langevin Piston) ensembles were constructed, respectively. Equilibration 3 was performed also in *NPT* ensemble, but without constraints of the heavy atoms. In the runs the Langevin temperature was set to 300 K and the Langevin damping coefficient to 2 ps^{-1} . SHAKE algorithm [111] was applied to all bonds to hydrogen atoms.

MD runs

The MD simulation was carried out using Berendsen *NPT* bath coupling [114] with constant temperature, $T = 300 \text{ K}$, constant pressure, $P = 1 \text{ atm}$ and pressure compressibility, $\beta = 4.75 \times 10^{-5} \text{ bar}^{-1}$. The usage of the SHAKE algorithm allows to have a time step δt of 2 fs. The total simulation time of the system was $t^{\text{sim}} = 56.35 \text{ ns}$. During the first 10 ns, snapshots were recorded every 1000 steps, subsequently they were recorded every 2000 timesteps.

9.2 Results

Since we are interested in the protein-nucleosome interactions rather than the water behaviour, all results described concern only the protein and DNA and water effects will not be discussed. The RMSDs of all C_α and P atoms superimposed for the whole system as well as some parts of it are plotted in Fig. 9.1, Fig. 9.2 and Fig. 9.3. All the plots display the RMSD from the diffusional encounter complex and the simulation time for minimization and equilibration is not included. The first Figure 9.2 shows that the core histones stay stably bound to the DNA, while the latter displays much larger fluctuations in the range of 5-8 Å. At around 15 ns there is an abrupt jump of the DNA backbone (Fig. 9.1). In order to find out whether it is related to linker DNA fluctuations, we plotted the RMSD of IDNA1 and IDNA2 in Fig. 9.2. It is seen that both DNAs fluctuate a lot in the range of 1-6 Å, but between 10 ns and 15 ns the RMSD of IDNA1 decreases in contrast to IDNA2. Inspection of the trajectory revealed that not only the linker DNAs, which are 10 bp long, but also part of the DNA wrapped around the histone

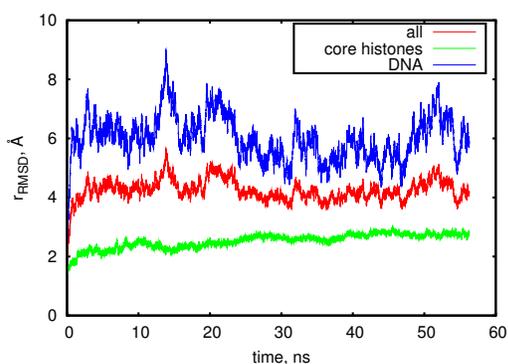


Figure 9.1: RMSDs (backbone) versus simulation time for the whole system (red), the core histones (green) and the whole DNA (blue).

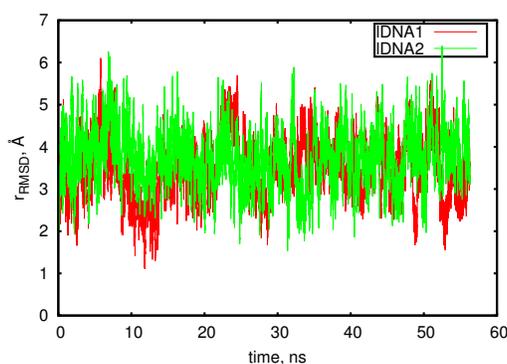


Figure 9.2: RMSDs of IDNA1 (red) and IDNA2 (green).

cores shows pronounced movement apart from the core. Such flexibility of DNA might be important for the initial stage of DNA unwrapping from the nucleosome core [19, 20, 160]. The linker histone RMSD is depicted

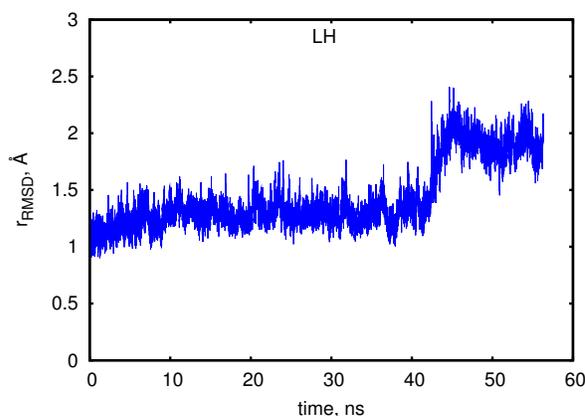


Figure 9.3: RMSD of the linker histone (LH) H5 versus the simulation time. A pronounced jump is observed at around 40 ns.

in Fig. 9.3. During the first ~ 40 ns the linker histone is displaced only slightly from the position at the diffusional encounter. This implies that this position found by the BD docking is a stable minimum and the GH5 does not tend to move towards the dyad or dissociate. Interestingly, at around 42 – 45 ns the GH5 suddenly experienced a positional and conformational change leading to a RMSD increase of approximately 1 Å. Such behaviour

can be attributed to an induced fit, where upon binding, one of the binding partners passes to another conformation [146, 147]. Such a transition could be related to the function of the linker histone to protect DNA from nuclease digestion [32]. However, whether this observation is really an induced fit in the sense of characterising the binding process itself is unclear. It might be that this conformation of the linker histone could also occur in equilibrium, but because of its low population, the probability of occurrence is negligible. Here, we should note that to our knowledge there is no experimental evidence of such a hypothesis or of an induced fit of the GH5 on the nucleosome.

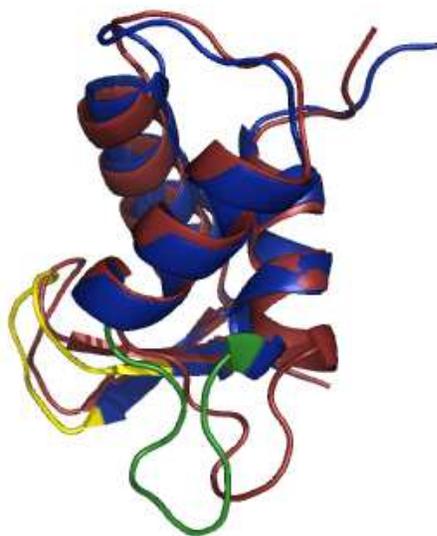


Figure 9.4: Linker histone conformation at the diffusional encounter (ruby) superimposed to its structure at the end of the MD run (blue). The loops, which underwent conformational change are highlighted in green (between helices 1 and 2) and in yellow (beta sheet loop).

To shed more light on the conformational change of the GH5 at the end of the MD simulation with respect to its structure at the beginning, both structures were superimposed as shown in Fig. 9.4. The calculated backbone RMSD (C_α, N, C, O) was 2.1 Å (all atom RMSD=3.8 Å). Well seen are the beta sheet and helix 1 and 2 loops, which underwent an observable conformational change. More pronounced is the structural modification of the loop between helix 1 and 2 and this also partially involves α helix 1. In Fig. 9.5 the backbone RMSD of the residues constituting GH5 is shown. The largest RMSD is exhibited by S24, which is located at the N-terminal of GH5. Also, the last residue K97 on the C-terminal tail displayed a RMSD of about 5 Å. The residues situated on the helix 1 and 2 loop (green) show a considerable RMSD, especially G43, R42, K40 and A38. In addition, the residue K37 located at the end of α helix 1 experienced a large conformational change with a RMSD of 5.6 Å (see Fig. 9.4). On the other hand, the amino acids laying

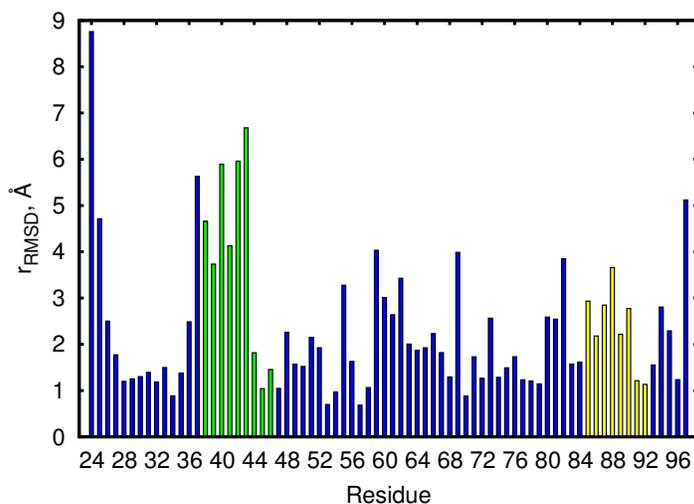


Figure 9.5: Backbone residue RMSD of the GH5 at the last MD snapshot with respect to the GH5 at the end of the BD run. Highlighted are the loops between helix 1 and helix 2 (green) and between the beta strands (yellow).

on the beta sheet loop (yellow in Fig. 9.5) displayed a moderate RMSD in the range of 1-3.8 Å. The hydrophobic residue V87 is displaced by 2.8 Å in direction opposite to the helix 3. The other observed RMSDs greater than 3 Å are due to long side chain residues as Lys and Arg. The plot clearly indicates a collective conformational change of the loop connecting helix 1 and helix 2 with a significant contribution of the last positive residue (K37) on α helix 1.

The position of the whole system at the end of the BD and MD simulations is displayed in Fig. 9.6. Both configurations were superimposed on the core histones with a backbone RMSD of 2.6 Å. The linker histone readjusted and moved closer to the IDNA1 in comparison to the reference structure. Moreover, the flexible loop (helix 1 and 2) penetrated into the major groove of the IDNA1 in such a way that the plane of the loop together with the hydrophobic loop (Val87, Ala89) aligned parallel to the phosphate backbone (see Fig. 9.7). For comparison the aligned diffusional encounter complex is depicted in the same perspective in Fig. 9.8. In addition, the interaction between the GH5 and IDNA1 led to a distinct twisting of the linker DNA1. Looking at the linker DNAs (Fig. 9.6) one can see that they are more open after the MD simulation than in the diffusional encounter conformation. A plot showing how the distance between them changes during the simulation time is depicted in Fig. 9.9. Well seen is that the linker DNAs are very dynamic over large spatial scales. Sharp decrease of the RMSD is observed at times of 43 – 46 ns and 55 – 56 ns indicating for motions bringing near the IDNAs. It can be seen that mainly IDNA2 contributes to this behaviour (see Fig. 9.10) moving towards the dyad. Linker DNA1, which is bound to the linker histone did not exhibit so dramatic change with respect to the

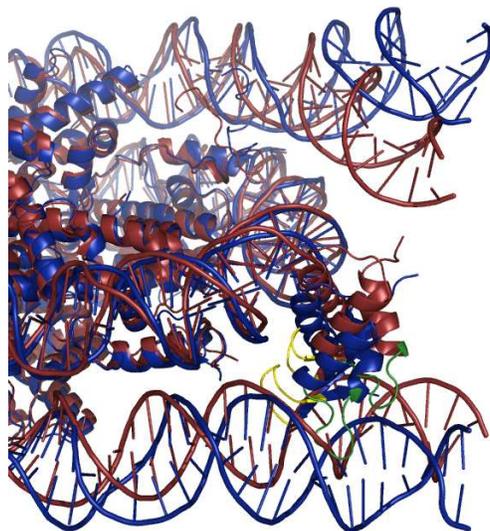


Figure 9.6: Nucleosome-linker histone complex before (ruby) and after (blue) the MD simulation. The structures were aligned on the core histones and the backbone RMSD was 2.6 Å.

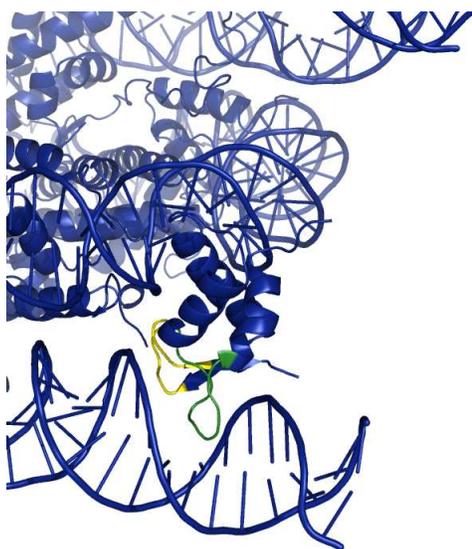


Figure 9.7: Nucleosome-linker histone complex at the last MD snapshot. The flexible loops (helix 1 and 2, beta sheet) penetrate in the major groove of DNA1 (colored green and yellow, respectively).

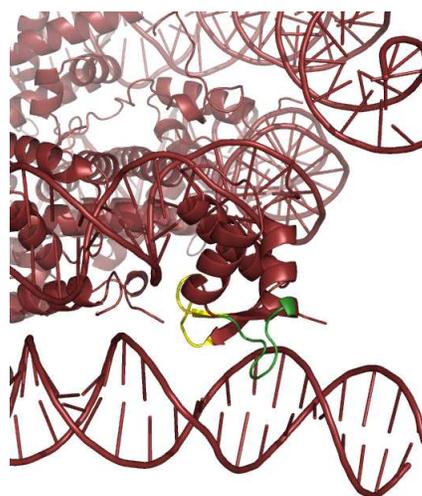


Figure 9.8: Nucleosome-linker histone complex at the end of the BD run. The perspective of the view is the same as in Fig. 9.7.

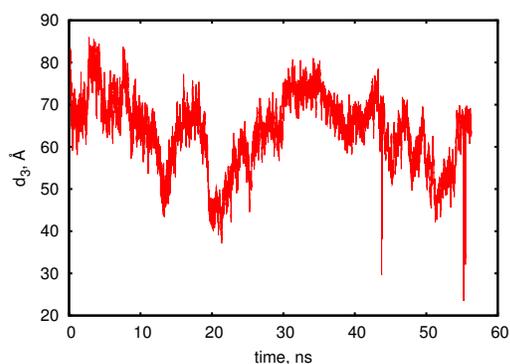


Figure 9.9: Distance d_3 between the end of IDNA1 and the end of IDNA2 during the simulation time.

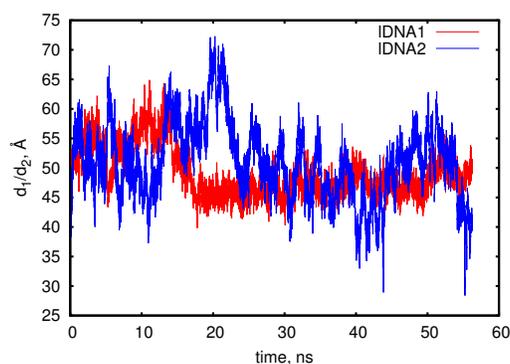


Figure 9.10: Distance d_1/d_2 between the end of IDNA1 (blue) and IDNA2 (red) and the dyad during the MD run, respectively.

dyad. Moreover, after 20 *ns* it 'embraced' the linker histone and went closer to the nucleosome dyad. This observation is in agreement with the NMA data (Sec. 5.6) proposing suppression of the IDNA1 motion due to presence of the linker histone.

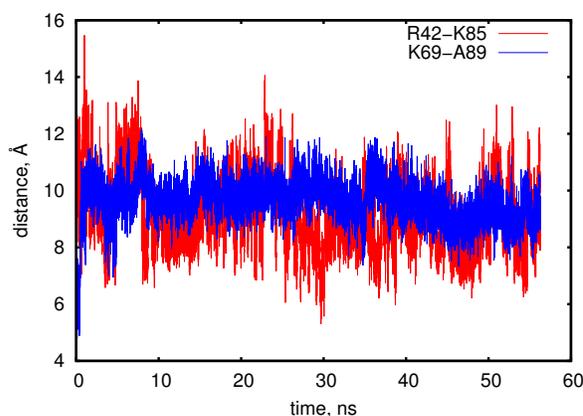


Figure 9.11: Distances between residues within the GH5 during the time course. Residue R42 (helix 1 and 2 loop) and residue K85 (beta sheet loop) are shown in blue, while residues K69 (helix 3) and A89 (beta sheet loop) are in red.

In order to quantify the conformational changes within GH5, the distances between the helix 1-2 loop (R42) and beta sheet loop (K85) as well as between helix 3 (K69) and the hydrophobic loop (A89) were plotted (Fig 9.11). The atoms used for the distance calculation were N_ζ , C_ζ and C_β for Lys, Arg

and Ala, respectively. The atoms located on the positive side chains are in direct contact to the solvent, while the C_β of Ala89 is orientated towards helix 3 in the crystal structure of H5. The long side chains of R42 and K85 fluctuate a lot displaying larger deviations from the distance (8.4 Å) in the crystal structure (Fig 9.11). As we have already seen the structural changes of the loops containing R42 and K85 contribute to these fluctuations. The hydrophobic residue Ala89 moved away from helix 3 and the distance between A89 and K69 reached a value of ~ 10 Å. It is almost double in comparison to the distance of 5.2 Å in the crystal structure. Such opening could be attributed to a mechanism by which the hydrophobic loop tries to contact IDNA1 while K69site (helix 3) still keeps close to the nucleosomal DNA. In this vicinity the hydrophobic loop might not feel the unfavourable water effect, which could contribute to the A89-K69 separation.

9.3 Summary

Preliminary results on the nucleosome-linker histone interactions were shown in this chapter. After finding its binding path to the nucleosome via diffusion, the GH5 remained in contact with the linker DNA1 during 56 ns of MD simulation. This suggests that the linker histone is located in a stable potential well on the energy landscape of the system. However, internally the GH5 experienced large conformational change of the loop linking α helices 1 and 2 and a moderate displacement of the beta sheet loop apart from helix 3. Whether this conformation forms due to an induced fit effect [147] or a native state of the linker histone with low population remains unclear. Moreover, both loops penetrated into the major groove of IDNA1 aligning parallel to the backbone-nucleotides plane. In addition, this interaction caused twisting of IDNA1. Such mechanical deformation on DNA upon binding of GH5 can be attributed to the mechanical forces acting on IDNA1 due to the presence of the linker histone [60, 161]. This interaction might be crucial for the formation of a long living chromosome particle [70, 94]. As proposed by Misteli *et al.* [94] the residence time is several minutes, which cannot be easily accessed even with an atomic detail Brownian Dynamics.

The binding site K69site on the nucleosomal DNA did not show any significant motions and remained in close contact with the nDNA. The other linker DNA2 sampled large configurational space experiencing fluctuations in a broad spatial range up to 73 Å from its conformation in the diffusional encounter. This is a clear indication of the highly dynamic nature of DNA in chromatin [122]. The behaviour of the linker DNAs confirmed our observations from the NMA of highly flexible chains when they are free and of

suppression of motion in the presence of linker histone. These preliminary data give some insight into the dynamic nature of DNA at the nucleosomal level in the presence of GH5 and on the structural changes accompanying the binding dynamics.

10

Conclusion and Outlook

The present thesis addresses questions concerning the dynamic nature of complex biomolecular systems and, particularly, the driving forces and interactions involved in bimolecular complex formation.

In the cell nucleus, the nature of DNA packaging (chromatin) is governed by a tremendous number of biomolecular processes comprising length and time scales from 10^{-10} m to 10^{-5} m and 10^{-15} s to 10^{-1} s, respectively. To unravel and understand these processes in terms of biological function, physical principles must be applied. However, the available experimental techniques and methods capture only a small part of the complex biomolecular network and experiments, describing completely processes occurring in various spatial and temporal scales, are not trivial to perform. With the advent of highly powerful computers, a rapid increase of computational methods and tools for modelling such processes with improved accuracy has been observed. With the combination of various physically based computational methods, one can model and predict, with atomic detail, biomolecular interactions spanning over more than several hundreds of nanosecond.

In this thesis, the biomolecular complex formation between the nucleosome (a protein-DNA particle of component of constituting the chromatin) and the linker histone (protein contributing to chromatin compaction) has been investigated thoroughly using the well-established computational physics based methods: Normal Mode Analysis (NMA), Brownian Dynamics (BD) and Molecular Dynamics (MD) Simulation. Thus, dynamic behaviour on time scales from picoseconds to microseconds has been assessed. In addition, a new method for incorporating collective molecular degrees of freedom in the rigid-body Brownian Dynamics algorithm has been successfully developed. A physically precise picture of the structure and dynamics of the nucleosome-linker histone (chromatosome) complex has been proposed.

The determination of the nucleosome structure to a resolution of 1.9 Å [5] (147 bp DNA) has greatly contributed to our understanding of possible mechanisms of DNA-histone binding, nucleosome repositioning and forma-

tion of higher-order chromatin structure [3]. With the crystal structure of the tetranucleosome [4], a new step forward has been made provoking scientists for explanation of the diverse experimental and theoretical data on chromatin. An obvious drawback of the tetranucleosome structure is the lack of a linker histone, which has been experimentally shown to exist within chromatin [17, 18, 31, 58, 69, 94]. The available experimental structure of the globular domain of linker histone H5 (GH5) [72] made it possible to model nucleosome-linker histone interactions in atomic detail.

We identified the chromatosome structure by extensive Brownian Dynamics (BD) simulations mimicking the physical process of diffusion combined with Normal Mode Analysis (NMA). The linker DNAs (lDNAs) of the nucleosome, 10 bp each, exhibited the most pronounced conformational changes obtained by the NMA and based on these conformations the GH5 was docked to the nucleosome to find out the diffusional encounter complex, a complex formed before the last stage of fully bound molecules. A dominant binding mode of GH5 with respect to the nucleosome, placed asymmetrically one helical turn away from the dyad point close to lDNA1, was found for 8 nucleosome conformations (out of 13). Two binding sites of GH5, the K69site (R47, K69, R73, R74) and R42site (R42, K97, R94), binding to the nucleosomal DNA (nDNA) and lDNA1, respectively, were revealed. Residue K69 was found to contribute most to binding, K59 least. All these findings are in agreement with an experimental FRAP study [17]. The other 5 extreme conformations of the nucleosome showed different GH5 binding mode having, on the most open structures (7_6 , 8_2 and 8_3), GH5 bound to the lDNA2. The dependence of the linker histone H5 binding mode on two groups of distinct nucleosome conformations could be related to the existence of a heteromorphic chromatin structure [58], which can simultaneously adapt solenoid and zig-zag conformations [58]. In this way, the GH5 could be present on the dominant binding mode in the tighter zig-zag chromatin, whereas the position on lDNA2 could be in the solenoid chromatin structure. A subsequent NMA on the diffusional encounter complex pointed out a restricted sampling of lDNA1 in the configurational space in comparison to its motions on a free nucleosome. Clearly, the presence of GH5 indicated a suppression of the lDNA1 fluctuations.

Using the same algorithm, the effect of mutations, i.e. a replacement of one amino acid residue by another, on the binding mode of GH5 was investigated. Experimental FRAP study [17] has identified the amino acids (residues) on a linker histone H1⁰ (97% similar to H5) which contribute most and least to binding to the nucleosome. The experimental mutational data have been mapped onto the crystal structure of GH5 and Brown *et.al* [17] have proposed a model for the binding sites and position of GH5 on the nu-

cleosome. This model was computationally tested by Brownian Dynamics and very good agreement with the experimental data was observed. First, the effect of mutations on the binding mode of GH5 was found to be nonadditive, i.e. the contribution of multiple mutations to binding is not equal to the sum of the contributions of each single mutant. Second, single mutants could not change dramatically the binding position of GH5, but they could induce an orientational change in the diffusional encounter of GH5. On the other hand, simultaneous mutations of positive residues on the surface of GH5 affected significantly the binding mode of GH5, decreasing its strength. In addition, the dominant binding position was conserved for most of the nucleosome conformations in mode 7 (obtained by the NMA), while for mode 8 only conformation 8_1 showed similar behaviour. The residues belonging to K69site showed a slightly higher binding strength than the residues constituting the R42site. This could be attributed to a two-step binding process, where firstly K69site binds to the nDNA and, second, R42site readjusts to form contact with the linker DNA1. All these findings are complementary to other methods trying to reveal the dynamics and interaction in the linker protein network within chromatin [70].

The structure of the diffusional encounter complex is trapped in a local energy minimum along the diffusional path of the GH5, but information on how fast this minimum has been reached is missing. Due to the highly dynamic nature of DNA in chromatin, the speed of reaction could be very important for the formation of a stable nucleosome-linker histone complex. To address this issue, we investigated the binding kinetics of GH5 and 11 of its mutants to three different nucleosome conformations (7_0 , 7_6 and 8_3). Our simulations yielded results very close to the Smoluchowski rate ($k_{\text{on}} \sim 10^9 \text{M}^{-1}\text{s}^{-1}$), which, after comparing with the basal association rates, were attributed to electrostatic enhancement of the association rates. In addition, the high association rates can be explained by the searching mechanism of DNA binding protein, which has been shown to enhance the association rates [150]. The residues participating in the K69site, which was found to bind to nDNA, showed significant reduction in their association rates upon mutation. On the other hand, mutations of residues belonging to R42site, which contacts lDNA, display rates very close to the wild type (WT) rate indicating modest contribution to the binding kinetics. We proposed a hypothesis by which K69site binds firstly to the linker DNA and performs a 1D diffusion along it until it reaches the binding site, where GH5 rotates and K69site and R42site contact nDNA and lDNA, respectively. This can enhance the association rate and can explain the small contribution of the R42site to it. Another reason could be the neglected flexibility in the BD simulations, which might be very important for the residues on R42site located on loops. Further investigation

for incorporating flexibility as well as for modelling conformational changes is needed.

A new method for treating explicitly flexibility in all-atom Brownian Dynamics was developed. One of the interacting molecules was represented by a discrete set of conformations with a transition rate described by a Markovian process. Four algorithms for selecting a new conformation were implemented. The first algorithm called *energy algorithm* selects the conformation with the lowest interaction energy between the current positions of the interacting molecules. The second algorithm called *modified Metropolis algorithm* uses the standard Metropolis method with a slight modification for a choice when three, instead of two, conformations are available. The *random energy algorithm* is the third one, which selects only adjacent conformations according to their interaction energy with the other molecule. And the last *random algorithm* uses a random uniform number to select the new conformation. The conformation selection is invoked within a certain distance between both molecules after an initially specified transition time. There are no restrictions on the number of conformations or on the transition time as long as it is computationally feasible and the approximations of the Brownian Dynamics algorithm hold, respectively. The method was applied to the tNCP-GH5 system and the nucleosome was represented by a set of conformations obtained by the NMA. The method reproduced remarkably well the results obtained in Chapter 5 in terms of binding position of GH5 on the nucleosome. This is an indication that the new implementation can, although with less statistical confidence, lead to the same docking position in reasonable computational time. However, in terms of nucleosome conformations, the GH5 bound most frequently at the end of the diffusional process to δ_1 , and, the second and third most frequent conformations were δ_2 and δ_4 . These results suggested that the GH5 prefers to bind to more open nucleosome conformations than the equilibrium δ_0 , which has been resolved without the presence of the linker histone [4]. In general, the new methodology should give better results for a longer sampling and, for such a case, the Metropolis method should lead to global convergence.

The obtained diffusional encounter complex is the last step preceding the formation of a fully bound complex. In order to reveal how this process occurs, short-range interactions including atomic flexibility should be taken into account. We carried out Molecular Dynamics (MD) simulation on the chromosome structure. Preliminary results on the nucleosome-linker histone interactions were shown. After finding its binding path to the nucleosome via diffusion, the GH5 remained in contact with the linker DNA1 during 56 ns of MD study. This suggests that the linker histone is located in a stable potential well on the energy landscape of the system. However,

internally the GH5 experienced large conformational change of the loop linking α helices 1 and 2 and a moderate displacement of the beta sheet loop apart from helix 3. Whether this conformation forms due to an induced fit effect [147] or a native state of the linker histone with low population remains unclear [162]. Moreover, both loops penetrated into the major groove of IDNA1 aligning parallel to the backbone-nucleotides plane. In addition, this interaction caused twisting of IDNA1. Such mechanical deformation on DNA upon binding of GH5 can be attributed to the mechanical forces acting on IDNA1 due to the presence of the linker histone [60, 161]. This interaction might be crucial for the formation of a long living chromosome particle [70, 94]. As proposed by Misteli *et al.* [94] the residence time is several minutes, which cannot be easily accessed even with an atomic detail Brownian Dynamics.

The binding site K69site on the nucleosomal DNA did not show any significant motions and remained in close contact with the nDNA. The other linker DNA2 sampled large configurational space experiencing fluctuations in a broad spatial range up to 73 Å from its conformation in the diffusional encounter. This is a clear indication of the highly dynamic nature of DNA in chromatin [122]. The behaviour of the linker DNAs confirmed our observations from the NMA of highly flexible chains when they are free and of suppression of motion in the presence of linker histone. These preliminary data give some insight into the dynamic nature of DNA at the nucleosomal level in the presence of GH5 and on the structural changes accompanying the binding dynamics.

In summary, the atomic structure of the chromosome particle has been characterized by a combination of physically based computational approaches. In addition, the dynamics and interactions in the formation of the chromosome particle have been identified and the contribution of flexible linker DNAs to the complexation recognized. It was found that the linker histone binds asymmetrically on the nucleosome with respect to dyad axis, forming a stable structure, preferably on open nucleosome conformations (with respect to the angle between linker DNAs). In overall, the presented work is a valuable part of the process of understanding the structure and function of the higher-order chromatin fiber.

Outlook

Simulations of macromolecular complexation are computationally demanding and require the use of approximations such as the neglect of molecular flexibility. Furthermore, establishing the effects on macromolecular association of the heterogeneous and crowded cellular environment is a challenge for

both computational and experimental approaches. Surmounting these hurdles requires the development of multiple-scale and coarse grained models with more accurate molecular interaction force fields as well as the development of highly parallelized software and new computing hardware to permit detailed simulations over many orders of time and length scales. The current study provides an efficient description of the association process between the linker histone and the nucleosome in atomic detail incorporating flexibility. The results obtained can be used as an input in modelling chromatin on a coarse-grained level in which several atoms, residues or the whole nucleosome are represented as a single geometrical object, the interactions involved are included and the dynamics is simulated either by BD or Monte Carlo [22–24, 58, 76, 79]. Attempts to predict the conformation of chromatin fiber have been made at an atomistic level as well [77].

The process of protein-DNA recognition and, particularly within chromatin, is affected by different factors as the histone tails [24], the presence of mono- and multivalent ions, the effect of histone acetylation [18] and the nucleosome repeat length [85]. Although the contribution of each of them to the compaction of chromatin fiber is still under debate, a promising way to improve the model proposed is to investigate these effects on the binding interaction and kinetics of the chromatosome particle complexation. Going further, systems of two and three nucleosomes including the linker histone could be modelled in atomic detail with different degree of flexibility. The symmetry/asymmetry distribution of the linker histone on the nucleosome is also an open question and needs conclusive experimental evidences. The dissociation pathway of the linker histone from the nucleosome can give insights on the stability of the chromatosome particle as well as on the dissociation constant. However, due to the long time scale of the complex, atomic detail simulations can be computationally demanding. A way to overcome this could be the application of a biased force to the linker histone, which would enforce a dissociation process. Another partially unresolved issue is the effect of the C-terminal domain of the linker histone on the dynamics of the linker DNAs and whether it plays only a role in bridging together the linker DNAs on the nucleosome or it has other functions regarding the linker histone-nucleosome interactions. The sketched open questions need future investigations.

A

Computation of the Brownian Dynamics forces and torques

The electrostatic force computed at each BD step (see Sec. 4.2.2) is given by

$$\mathbf{F}^{\text{el}}(\mathbf{r}) = \mathbf{F}_2^{\text{el}}(\mathbf{r}) - \mathbf{F}_1^{\text{el}}(\mathbf{r}), \quad (\text{A.1})$$

where $\mathbf{F}_2^{\text{el}}(\mathbf{r})$ is the total electrostatic force acting on molecule 2 due to the presence of molecule 1, and $\mathbf{F}_1^{\text{el}}(\mathbf{r})$ is the force exerted on molecule 1 due to molecule 2. The force $\mathbf{F}_2^{\text{el}}(\mathbf{r})$ is given by

$$\mathbf{F}_2^{\text{el}}(\mathbf{r}) = \sum_l \left[\frac{1}{2} q_{l2}^{\text{eff}}(\mathbf{r}_l) \mathbf{f}_{l2}(\mathbf{r}_l) + [q_{l2}^{\text{eff}}(\mathbf{r}_l)]^2 \mathbf{f}_{l2}^{\text{ed}}(\mathbf{r}_l) \right], \quad (\text{A.2})$$

where $\mathbf{f}_{l2}(\mathbf{r}_l)$ is the electrostatic force¹ (in $k_B T / (e\text{\AA})$) acting on a charge $q_{l2}^{\text{eff}}(\mathbf{r}_l)$ due to the electrostatic potential of molecule 1 and $\mathbf{f}_{l2}^{\text{ed}}(\mathbf{r}_l)$ is the electrostatic desolvation force (in $k_B T / (e^2\text{\AA})$) experienced by the charge $q_{l2}^{\text{eff}}(\mathbf{r}_l)$ due to the desolvation (grid) penalty of molecule 1. In a similar way, the force $\mathbf{F}_1^{\text{el}}(\mathbf{r})$ yields

$$\mathbf{F}_1^{\text{el}}(\mathbf{r}) = \sum_m \left[\frac{1}{2} q_{m1}^{\text{eff}}(\mathbf{r}_m) \mathbf{f}_{m1}(\mathbf{r}_m) + [q_{m1}^{\text{eff}}(\mathbf{r}_m)]^2 \mathbf{f}_{m1}^{\text{ed}}(\mathbf{r}_m) \right]. \quad (\text{A.3})$$

The components of \mathbf{f} are computed as a finite-difference derivative of the electrostatic/(electrostatic desolvation) potential φ on a 3D grid (cubic in the thesis, see Fig. A.1)

$$f^i = -\frac{1}{h} \sum_{jk} \Delta\varphi^{ijk} \Delta s^{jk}, \quad (\text{A.4})$$

where h is the grid spacing, $\Delta\varphi^{ijk}$ is the electrostatic/(electrostatic desolvation) potential difference along i in a cube with a diagonal $[(i,j,k);(i +$

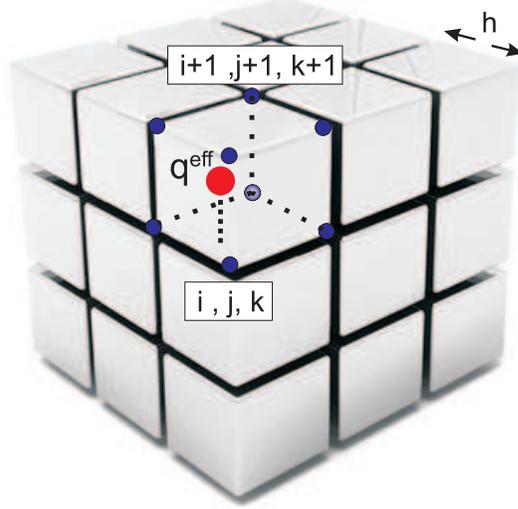


Figure A.1: Schematic view of a 3D (cubic) grid encompassing each molecule. At the cube corners (blue points) values of the relevant potential are assigned. The PB electrostatic potential is computed using partial atomic charges of the biomolecule. When a charge q^{eff} of molecule 1 (2) is within the grid of molecule 2 (1), the force components acting on it due to the grid 2 (1) are computed using Eq. A.4. The grid spacing is denoted by h .

$1, j+1, k+1)$] and Δs^{jk} is a dimensionless quantity accounting for the exact position of a charge q^{eff} within the cube (Fig. A.1).

The torque due to the electrostatic forces at each BD step is computed as

$$\begin{aligned} \mathbf{T}(\mathbf{r}) &= \sum_l \mathbf{r}_{2l} \times \left[\frac{1}{2} q_{l2}^{\text{eff}}(\mathbf{r}_l) \mathbf{f}_{l2}(\mathbf{r}_l) + [q_{l2}^{\text{eff}}(\mathbf{r}_l)]^2 \mathbf{f}_{l2}^{\text{ed}}(\mathbf{r}_l) \right] - \\ &- \sum_m \mathbf{r}_{1m} \times \left[\frac{1}{2} q_{m1}^{\text{eff}}(\mathbf{r}_m) \mathbf{f}_{m1}(\mathbf{r}_m) + [q_{m1}^{\text{eff}}(\mathbf{r}_m)]^2 \mathbf{f}_{m1}^{\text{ed}}(\mathbf{r}_m) \right] + \quad (\text{A.5}) \\ &+ \mathbf{r} \times \mathbf{F}_1^{\text{el}}(\mathbf{r}), \end{aligned}$$

where \mathbf{r}_{2l} and \mathbf{r}_{1m} are the distances to the effective charges from the center of molecule 2 and 1, respectively, and \mathbf{r} is the center-to-center distance between

¹Actually $\mathbf{f}_{l2}(\mathbf{r}_l)$ is the electric field at the position \mathbf{r}_l .

the molecules. In the simulations performed, the nucleosome (molecule 1) was kept fixed and only the rotations and translations of the linker histone (molecule 2) were accounted for at each BD step. In the case of rotating molecule 1, an additional torque term in Eq. A.5 should be considered. The computation of the forces due to the hydrophobic desolvation interaction follows a similar procedure using 3D hydrophobic interaction grids around both molecules and a buried solvent accessible surface area (SASA) [105]. Since such interactions were not included in the simulations performed, computational details are omitted.

Bibliography

- [1] Annalisa Izzo, Kinga Kamieniarz, and Robert Schneider. The histone h1 family: specific members, specific functions? *Biol Chem*, 389(4):333–343, Apr 2008. 1, 19
- [2] Nicole Happel and Detlef Doenecke. Histone h1 and its isoforms: contribution to chromatin structure and function. *Gene*, 431(1-2):1–12, Feb 2009. 1, 19
- [3] K. E. van Holde. *Chromatin*. Springer-Verlag, 1989. 1, 3, 19, 112
- [4] Thomas Schalch, Sylwia Duda, David F Sargent, and Timothy J Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141, Jul 2005. 1, 19, 21, 39, 40, 74, 99, 112, 114
- [5] Curt A Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol*, 319(5):1097–1113, Jun 2002. 1, 3, 21, 40, 111
- [6] Maarten Kruithof, Fan-Tso Chien, Andrew Routh, Colin Logie, Daniela Rhodes, and John van Noort. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat Struct Mol Biol*, 16(5):534–540, May 2009. 1, 22, 55
- [7] Brent D. Brower-Toland, Corey L. Smith, Richard C. Yeh, John T. Lis, Craig L. Peterson, and Michelle D. Wang. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of dna. *Proc Natl Acad Sci U S A*, 99(4):1960–1965, Feb 2002. 1
- [8] Kohji Hizume, Tonau Nakai, Sumiko Araki, Eloise Prieto, Kenichi Yoshikawa, and Kunio Takeyasu. Removal of histone tails from nucleosome dissects the physical mechanisms of salt-induced aggregation, linker histone h1-induced compaction, and 30-nm fiber formation of the nucleosome array. *Ultramicroscopy*, Mar 2009. 1, 23
- [9] D. V. Lebedev, M. V. Filatov, A. I. Kuklin, A. Kh. Islamov, E. Kentzinger, R. Pantina, B. P. Toperverg, and V. V. Isaev-Ivanov. Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: Dna structure exhibits biphasic fractal properties. *FEBS Lett*, 579(6):1465–1468, Feb 2005. 1
- [10] P. Baudy and S. Bram. Chromatin fiber dimensions and nucleosome orientation: a neutron scattering investigation. *Nucleic Acids Res*, 5(10):3697–3714, Oct 1978. 1
- [11] D. Axelrod, D.E. Koppel, J. Schlessinger, E. Elson, and W.W. Webb. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys J*, 16:1055–1069, 1976. 1
- [12] Jennifer Lippincott-Schwartz, Nihal Altan-Bonnet, and George H Patterson. Photobleaching and photoactivation: following protein dynamics in living cells. *Nat Cell Biol*, Suppl:S7–14, Sep 2003. 1
- [13] T. Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Ann. Physik*, 437:55–75, 1948. 1

BIBLIOGRAPHY

- [14] Jörg Langowski. Protein-protein interactions determined by fluorescence correlation spectroscopy. *Methods Cell Biol*, 85:471–484, 2008. 1
- [15] D. Magde, E. L. Elson, and W. W. Webb. Thermodynamic fluctuations in a reacting system: Measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.*, 29: 705–708, 1972. 1
- [16] E. L. Elson and D. Magde. Fluorescence correlation spectroscopy 1. conceptual basis and theory. *Biopolymers*, 13(1):1–27, Jan 1974. 1
- [17] David T Brown, Tina Izard, and Tom Misteli. Mapping the interaction surface of linker histone h1(0) with the nucleosome of native chromatin in vivo. *Nat Struct Mol Biol*, 13(3):250–255, Mar 2006. 2, 23, 39, 49, 50, 51, 57, 58, 64, 65, 66, 67, 69, 74, 112
- [18] Malte Bussiek, Katalin Tóth, Nathalie Schwarz, and Jörg Langowski. Trinucleosome compaction studied by fluorescence energy transfer and scanning force microscopy. *Biochemistry*, 45(36):10838–10846, Sep 2006. 2, 23, 45, 54, 112, 116
- [19] Alexander Gansen, Alessandro Valeri, Florian Hauger, Suren Felekyan, Stanislav Kalinin, Katalin Toth, Jörg Langowski, and Claus A M Seidel. Nucleosome disassembly intermediates characterized by single-molecule fret. *Proc Natl Acad Sci U S A*, Aug 2009. 2, 103
- [20] W. J A Koopmans, R. Buning, T. Schmidt, and J. van Noort. spfret using alternating excitation and fcs reveals progressive dna unwrapping in nucleosomes. *Biophys J*, 97(1):195–204, Jul 2009. 2, 103
- [21] Katalin Toth, Nathalie Brun, and Jörg Langowski. Chromatin compaction at the mononucleosome level. *Biochemistry*, 45(6):1591–1598, Feb 2006. 2, 23, 45
- [22] Jörg Langowski and Dieter W Heermann. Computational modeling of the chromatin fiber. *Semin Cell Dev Biol*, 18(5):659–667, Oct 2007. 2, 22, 53, 54, 116
- [23] Holger Merlitz, Konstantin V Klenin, Chen-Xu Wu, and Jörg Langowski. Facilitated diffusion of dna-binding proteins: Simulation of large systems. *J Chem Phys*, 125(1):014906, Jul 2006. 2, 116
- [24] Gaurav Arya and Tamar Schlick. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J Phys Chem A*, 113(16):4045–4059, Apr 2009. 2, 22, 116
- [25] Martin Depken and Helmut Schiessel. Nucleosome shape dictates chromatin fiber structure. *Biophys J*, 96(3):777–784, Feb 2009. 2
- [26] M. Ivanchenko, S. Zacharieva, J. Yaneva, and J. Zlatanova. Linker histone binding to superhelical dna: electrophoretic and filter binding studies. *Biochimie*, 80(12): 1031–1034, Dec 1998. 2
- [27] Y. B. Zhou, S. E. Gerchman, V. Ramakrishnan, A. Travers, and S. Muyldermans. Position and orientation of the globular domain of linker histone h5 on the nucleosome. *Nature*, 395(6700):402–405, Sep 1998. 2, 23, 39, 49, 51

-
- [28] Li Fan and Victoria A Roberts. Complex of linker histone h5 with the nucleosome and its implications for chromatin packing. *Proc Natl Acad Sci U S A*, 103(22): 8384–8389, May 2006. 2, 22, 23, 39, 51
- [29] M. M Srinivas Bharath, Nagasuma R Chandra, and M. R S Rao. Molecular modeling of the chromatosome particle. *Nucleic Acids Res*, 31(14):4264–4274, Jul 2003. 2, 22, 23, 39, 49, 51
- [30] Feng Cui and Victor B Zhurkin. Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to at-rich and methylated dna. *Nucleic Acids Res*, Mar 2009. 2, 22, 39, 50, 51
- [31] Philip J J Robinson, Woojin An, Andrew Routh, Fabrizio Martino, Lynda Chapman, Robert G Roeder, and Daniela Rhodes. 30 nm chromatin fibre decompaction requires both h4-k16 acetylation and linker histone eviction. *J Mol Biol*, 381(4):816–825, Sep 2008. 3, 23, 40, 112
- [32] M. M. Duggan and J. O. Thomas. Two dna-binding sites on the globular domain of histone h5 are required for binding to both bulk and 5 s reconstituted nucleosomes. *J Mol Biol*, 304(1):21–33, Nov 2000. 3, 22, 23, 46, 64, 104
- [33] Martin Kampmann. Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential. *Mol Microbiol*, 57(4):889–899, Aug 2005. 3
- [34] Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett*, 77(9):1905–1908, Aug 1996. 4, 26, 40, 42
- [35] E M Lifshitz L D Landau. *Statistical Physics*, volume 5. Butterworth-Heinemann, 3 edition, 1980. 5, 6
- [36] Kerson Huang. *Statistical Mechanics*. Wiley, 2 edition, 1987. 5
- [37] Robert Brown. A brief account of microscopical observations, made in the months of june, july, and august, 1827 on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Reprinted in Edin. New Phil. J.*, 5:358–371, 1928. 7
- [38] Robert M Mazo. *Brownian Motion: Fluctuations, Dynamics, and Applications*. Oxford University Press, USA, 2009. PaperbackS. 7, 8, 10
- [39] M. V. Smoluchowski. Sur le chemin moyenne parcouru par less molecule des gaz et sur son rapport la theorie de la diffusion. *Bull. Internat. Acad. Krakovie*, 202, 1906. 7
- [40] P Langevin. Sur la theorie de mouvement brownien. *C. R. Acad. Sci. Paris.*, 1908: 530, 146. 9
- [41] M. V. Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Z Phys Chem*, 92:129–168, 1917. 12, 14, 30
- [42] O. G. Berg and P. H. von Hippel. Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Biophys Chem*, 14:131–160, 1985. 12

BIBLIOGRAPHY

- [43] S. A. Allison, S. H. Northrup, and J. A. McCammon. Simulation of biomolecular diffusion and complex formation. *Biophys J*, 49(1):167–175, Jan 1986. 12
- [44] Cristina Albarran, Jose A Navarro, Fernando P Molina-Heredia, Piedad del S Murdoch, Miguel A De la Rosa, and Manuel Hervas. Laser flash-induced kinetic analysis of cytochrome f oxidation by wild-type and mutant plastocyanin from the cyanobacterium *nostoc* sp. pcc 7119. *Biochemistry*, 44(34):11601–11607, Aug 2005. 12
- [45] S. H. Northrup and H. P. Erickson. Kinetics of protein-protein association explained by brownian dynamics computer simulation. *Proc Natl Acad Sci U S A*, 89(8):3338–3342, Apr 1992. 14, 80
- [46] Domantas Motiejunas and Rebecca C. Wade. "Structural, Energetic, and Dynamic Aspects of Ligand Receptor Interactions" in *Comprehensive Medicinal Chemistry II*, volume 4. Elsevier: Oxford, 2007. 15
- [47] D.L. Ermak and J.A. McCammon. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69:1352–1360, 1978. 16, 27
- [48] Jian Sun and Harel Weinstein. Toward realistic modeling of dynamic processes in cell signaling: quantification of macromolecular crowding effects. *J Chem Phys*, 127(15):155105, Oct 2007. 16
- [49] Ken A Dill, Thomas M Truskett, Vojko Vlachy, and Barbara Hribar-Lee. Modeling water, the hydrophobic effect, and ion solvation. *Annu Rev Biophys Biomol Struct*, 34:173–199, 2005. 17
- [50] David Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, Sep 2005. 17
- [51] Brian N Dominy, Dieter Perl, Franz X Schmid, and Charles L Brooks. The effects of ionic strength on protein stability: the cold shock protein family. *J Mol Biol*, 319(2):541–554, May 2002. 17
- [52] Matthew S Lin, Nicolas Lux Fawzi, and Teresa Head-Gordon. Hydrophobic potential of mean force as a solvation function for protein structure prediction. *Structure*, 15(6):727–740, Jun 2007. 17
- [53] Jordanka Zlatanova. *Chromatin Structure and Dynamics: State-of-the-art*. Elsevier, 2004. 19
- [54] David J Tremethick. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*, 128(4):651–654, Feb 2007. 19
- [55] Peter H von Hippel. From "simple" dna-protein interactions to the macromolecular machines of gene expression. *Annu Rev Biophys Biomol Struct*, 36:79–105, 2007. 19, 54

-
- [56] Philip J J Robinson, Louise Fairall, Van A T Huynh, and Daniela Rhodes. Em measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A*, 103(17):6506–6511, Apr 2006. 19, 23, 39, 54
- [57] M. H. Parseghian and B. A. Hamkalo. A compendium of the histone h1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem Cell Biol*, 79(3):289–304, 2001. 19
- [58] Sergei A Grigoryev, Gaurav Arya, Sarah Correll, Christopher L Woodcock, and Tamar Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A*, Jul 2009. 20, 22, 23, 54, 112, 116
- [59] J. Widom. A relationship between the helical twist of dna and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc Natl Acad Sci U S A*, 89(3):1095–1099, Feb 1992. 21
- [60] Nils B Becker and Ralf Everaers. Dna nanomechanics in the nucleosome. *Structure*, 17(4):579–589, Apr 2009. 21, 108, 115
- [61] Thomas C Bishop. Geometry of the nucleosomal dna superhelix. *Biophys J*, 95(3):1007–1017, Aug 2008. 21
- [62] Jordanka Zlatanova, Thomas C Bishop, Jean-Marc Victor, Vaughn Jackson, and Ken van Holde. The nucleosome family: dynamic and growing. *Structure*, 17(2):160–171, Feb 2009. 21
- [63] T. Wocjan, K. Klenin, and J. Langowski. Brownian dynamics simulation of dna unrolling from the nucleosome (dagger). *J Phys Chem B*, Dec 2009. 21
- [64] Danilo Roccatano, Andre Barthel, and Martin Zacharias. Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers*, 85(5-6):407–421, 2007. 21, 41
- [65] H. Schiessel. The nucleosome: a transparent, slippery, sticky and yet stable dna-protein complex. *Eur Phys J E Soft Matter*, 19(3):251–262, Mar 2006. 21
- [66] Karine Voltz, Joanna Trylska, Valentina Tozzini, Vandana Kurkal-Siebert, Jörg Langowski, and Jeremy Smith. Coarse-grained force field for the nucleosome from self-consistent multiscaling. *J Comput Chem*, 29(9):1429–1439, Jul 2008. 21
- [67] David J Schwab, Robijn F Bruinsma, Joseph Rudnick, and Jonathan Widom. Nucleosome switches. *Phys Rev Lett*, 100(22):228105, Jun 2008. 21
- [68] I. M. Kulic and H. Schiessel. Dna spools under tension. *Phys Rev Lett*, 92(22):228101, Jun 2004. 21
- [69] Philip J J Robinson and Daniela Rhodes. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol*, 16(3):336–343, Jun 2006. 21, 39, 112

BIBLIOGRAPHY

- [70] Jordanka Zlatanova, Corrine Seebart, and Miroslav Tomschik. The linker-protein network: control of nucleosomal dna accessibility. *Trends Biochem Sci*, 33(6):247–253, Jun 2008. 22, 23, 41, 43, 52, 108, 113, 115
- [71] J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster, and C. L. Woodcock. Nucleosomes, linker dna, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*, 95(24):14173–14178, Nov 1998. 22, 23, 64
- [72] V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. Crystal structure of globular domain of histone h5 and its implications for nucleosome binding. *Nature*, 362(6417):219–223, Mar 1993. 22, 23, 40, 112
- [73] Piotr Widlak, Magdalena Kalinowska, Missag H Parseghian, Xu Lu, Jeffrey C Hansen, and William T Garrard. The histone h1 c-terminal domain binds to the apoptotic nuclease, dna fragmentation factor (dff40/cad) and stimulates dna cleavage. *Biochemistry*, 44(21):7871–7878, May 2005. 22
- [74] Xu Lu, Barbara Hamkalo, Missag H Parseghian, and Jeffrey C Hansen. Chromatin condensing functions of the linker histone c-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder. *Biochemistry*, 48(1):164–172, Jan 2009. 22
- [75] Michael J Hendzel, Melody A Lever, Ellen Crawford, and John P H Th'ng. The c-terminal domain is the primary determinant of histone h1 binding to chromatin in vivo. *J Biol Chem*, 279(19):20028–20034, May 2004. 22
- [76] Nick Kepper, Dietrich Foethke, Rene Stehr, Gero Wedemann, and Karsten Rippe. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophys J*, Jan 2008. 22, 53, 54, 116
- [77] Hua Wong, Jean-Marc Victor, and Julien Mozziconacci. An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. *PLoS ONE*, 2(9):e877, 2007. 22, 53, 54, 116
- [78] Gaurav Arya and Tamar Schlick. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci U S A*, 103(44):16236–16241, Oct 2006. 22
- [79] Frank Aumann, Filip Lankas, Maiwen Caudron, and Jörg Langowski. Monte carlo simulation of chromatin stretching. *Phys Rev E Stat Nonlin Soft Matter Phys*, 73(4 Pt 1):041927, Apr 2006. 22, 54, 116
- [80] Philipp M Diesinger and Dieter W Heermann. The influence of the cylindrical shape of the nucleosomes and h1 defects on properties of chromatin. *Biophys J*, 94(11):4165–4172, Jun 2008. 22, 53
- [81] Alex Gansen, Katalin Toth, Nathalie Schwarz, and Jörg Langowski. Structural variability of nucleosomes detected by single-pair forster resonance energy transfer: Histone acetylation, sequence variation, and salt effects (dagger). *J Phys Chem B*, 113(9):2604–2613, Mar 2009. 23

-
- [82] Xingwu Lu, Sandeep N Wontakal, Alexander V Emelyanov, Patrick Morcillo, Alexander Y Konev, Dmitry V Fyodorov, and Arthur I Skoultchi. Linker histone h1 is essential for drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes Dev*, Feb 2009. 23
- [83] Laura Cato, Katherine Stott, Matthew Watson, and Jean O Thomas. The interaction of hmgb1 and linker histones occurs through their acidic and basic tails. *J Mol Biol*, 384(5):1262–1272, Dec 2008. 23
- [84] Yasuhiro Hirano, Hirohide Takahashi, Masahiro Kumeta, Kohji Hizume, Yuya Hirai, Shotaro Otsuka, Shige H Yoshimura, and Kunio Takeyasu. Nuclear architecture and chromatin dynamics revealed by atomic force microscopy in combination with biochemistry and cell biology. *Phys Rev Lett*, 456(1):139–153, Apr 2008. 23
- [85] Andrew Routh, Sara Sandin, and Daniela Rhodes. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A*, 105(26):8872–8877, Jul 2008. 23, 54, 116
- [86] Zeina Al-Natour and Ahmed H Hassan. Effect of salt on the binding of the linker histone h1 to dna and nucleosomes. *DNA Cell Biol*, 26(6):445–452, Jun 2007. 23
- [87] Toshiro Kobori, Satoshi Iwamoto, Kunio Takeyasu, and Toshio Ohtani. Chromatin dynamics of unfolding and refolding controlled by the nucleosome repeat length and the linker and core histones. *Biopolymers*, 85(4):295–307, Mar 2007. 23
- [88] Julia N Yaneva, Elena G Paneva, Siyka I Zacharieva, and Jordanka Zlatanova. Histone h1 interacts preferentially with dna fragments containing a cisplatin-induced 1,2-intrastrand cross-link. *Z Naturforsch [C]*, 62(11-12):905–908, 2007. 23
- [89] Frederic Catez, Tetsuya Ueda, and Michael Bustin. Determinants of histone h1 mobility and chromatin binding in living cells. *Nat Struct Mol Biol*, 13(4):305–310, Apr 2006. 23, 43, 55
- [90] Christopher L Woodcock, Arthur I Skoultchi, and Yuhong Fan. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res*, 14(1):17–25, 2006. 23
- [91] Julia N Yaneva, Elena G Paneva, Siyka I Zacharieva, and Jordanka S Zlatanova. Linker histones do not interact with dna containing a single interstrand cross-link created by cisplatin. *Z Naturforsch [C]*, 61(11-12):879–883, 2006. 23
- [92] Kohji Hizume, Shige H Yoshimura, and Kunio Takeyasu. Linker histone h1 per se can induce three-dimensional folding of chromatin fiber. *Biochemistry*, 44(39):12978–12989, Oct 2005. 23
- [93] K. Tóth, N. Brun, and J. Langowski. Trajectory of nucleosomal linker dna studied by fluorescence resonance energy transfer. *Biochemistry*, 40(23):6921–6928, Jun 2001. 23

BIBLIOGRAPHY

- [94] T. Misteli, A. Gunjan, R. Hock, M. Bustin, and D. T. Brown. Dynamic binding of histone h1 to chromatin in living cells. *Nature*, 408(6814):877–881, Dec 2000. 23, 54, 65, 108, 112, 115
- [95] W. An, J. Zlatanova, S. H. Leuba, and K. van Holde. The site of binding of linker histone to the nucleosome does not depend upon the amino termini of core histones. *Biochimie*, 81(7):727–732, Jul 1999. 23, 40, 61
- [96] W. An, K. van Holde, and J. Zlatanova. Linker histone protection of chromatosomes reconstituted on 5s rDNA from *Xenopus borealis*: a reinvestigation. *Nucleic Acids Res*, 26(17):4042–4046, Sep 1998. 23
- [97] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins*, 33(3):417–429, Nov 1998. 26, 41
- [98] R. R. Gabdouliline and R. C. Wade. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods*, 14(3):329–341, Mar 1998. 28, 31, 46
- [99] Scott H. Northrup, Stuart A. Allison, and J. Andrew McCammon. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.*, 80:1517–1524, 1984. 30
- [100] Nathan A Baker. Poisson-boltzmann methods for biomolecular electrostatics. *Methods Enzymol*, 383:94–118, 2004. 31
- [101] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–10041, Aug 2001. 32, 45, 76
- [102] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon. Electrostatic and diffusion of molecules in solution: simulations with the university-of-houston-brownian dynamics program. *Comput. Phys. Commun.*, 62:187–197, 1991. 32, 45
- [103] R. R. Gabdouliline and R. C. Wade. Effective charges for macromolecules in solvent. *J. Phys. Chem.*, 100(6):3868–3878, Dec 1996. 32, 46
- [104] A. H. Elcock, R. R. Gabdouliline, R. C. Wade, and J. A. McCammon. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J Mol Biol*, 291(1):149–162, Aug 1999. 33, 79, 80
- [105] Razif R Gabdouliline and Rebecca C Wade. 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 events. *J Am Chem Soc*, 131(26):9230–9238, Jul 2009. 33, 119
- [106] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, Nov 2006. 35, 101

-
- [107] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16):1781–1802, Dec 2005. 35, 38, 101
- [108] P P Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys. (Leipzig)*, 64:253–287, 1921. 35
- [109] Lalith Perera Ulrich Essmann and Max L. Berkowitz. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995. 36
- [110] L Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:1–98, 1967. 36
- [111] Ciccotti G. Ryckaert, J.P. and H.J.C. Berendsen. Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of n-alkanes. *J. Comput. Phys*, 23:327–341, 1977. 37, 102
- [112] S. Miyamoto and P.A. Kollman. Settle - an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13:952–962., 1992. 37
- [113] Charles L. Brooks Axel Bruenger and Martin Karplus. Stochastic boundary conditions for molecular dynamics simulations of st2 water. *Chemical Physics Letters*, 105:495–500, 1984. 38
- [114] Berendsen H.J.C., Postma J.P.M., A. Vangunsteren W.F. and Dinola, and J.R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690., 1984. 38, 102
- [115] Yuhong Zhang Scott E. Feller and Richard W. Pastor. Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.*, 103:4613, 1995. 38
- [116] W. G. Hoover. *Computational Statistical Mechanics*. Elsevier: Amsterdam, 1991. 38
- [117] Frederic Catez, Huan Yang, Kevin J Tracey, Raymond Reeves, Tom Misteli, and Michael Bustin. Network of dynamic interactions between histone h1 and high-mobility-group proteins in chromatin. *Mol Cell Biol*, 24(10):4321–4328, May 2004. 39
- [118] S. Hayward, A. Kitao, and H. J. Berendsen. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, 27(3):425–437, Mar 1997. 41
- [119] M. Delarue and Y-H. Sanejouand. Simplified normal mode analysis of conformational transitions in dna-dependent polymerases: the elastic network model. *J Mol Biol*, 320(5):1011–1024, Jul 2002. 41
- [120] Ozge Kurkcuoglu, Zeynep Kurkcuoglu, Pemra Doruker, and Robert L Jernigan. Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. *Proteins*, 75(4):837–845, Jun 2009. 41

BIBLIOGRAPHY

- [121] Ken van Holde and Jordanka Zlatanova. Chromatin fiber structure: Where is the problem now? *Semin Cell Dev Biol*, 18(5):651–658, Oct 2007. 41
- [122] Michael G Poirier, Malte Bussiek, Jörg Langowski, and Jonathan Widom. Spontaneous access to dna target sites in folded chromatin fibers. *J Mol Biol*, 379(4):772–786, Jun 2008. 41, 55, 108, 115
- [123] Erik Lindahl, Cyril Azuara, Patrice Koehl, and Marc Delarue. Nomad-ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*, 34(Web Server issue):W52–W56, Jul 2006. 41
- [124] D. Daniel ben Avraham. Vibrational normal-mode spectrum of globular proteins. *Phys Rev B*, 47:14559–14560, 1993. 42
- [125] Todd J Dolinsky, Jens E Nielsen, J. Andrew McCammon, and Nathan A Baker. Pdb2pqr: an automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667, Jul 2004. 45
- [126] Sanbo Qin and Huan-Xiang Zhou. Do electrostatic interactions destabilize protein-nucleic acid binding? *Biopolymers*, 86(2):112–118, Jun 2007. 45
- [127] Dipanjan Bhattacharya, Aprotim Mazumder, S. Annie Miriam, and G. V. Shivashankar. Egfp-tagged core and linker histones diffuse via distinct mechanisms within living cells. *Biophys J*, 91(6):2326–2336, Sep 2006. 45
- [128] T. Roopa and G. V. Shivashankar. Direct measurement of local chromatin fluidity using optical trap modulation force spectroscopy. *Biophys J*, 91(12):4632–4637, Dec 2006. 45
- [129] Domantas Motiejunas, Razif Gabdoulline, Ting Wang, Anna Feldman-Salit, Tim Johann, Peter J Winn, and Rebecca C Wade. Protein-protein docking by simulating the process of association subject to biochemical constraints. *Proteins*, 71(4):1955–1969, Jun 2008. 47
- [130] P. J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*, 28(7):849–857, Jul 1985. 50, 51
- [131] R. C. Wade and P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem*, 36(1):148–156, Jan 1993. 50, 51
- [132] R. C. Wade, K. J. Clark, and P. J. Goodford. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem*, 36(1):140–147, Jan 1993. 50, 51
- [133] Gu Li, Marcia Levitus, Carlos Bustamante, and Jonathan Widom. Rapid spontaneous accessibility of nucleosomal dna. *Nat Struct Mol Biol*, 12(1):46–53, Jan 2005. 52

-
- [134] Maik Engholm, Martijn de Jager, Andrew Flaus, Ruth Brenk, John van Noort, and Tom Owen-Hughes. Nucleosomes can invade dna territories occupied by their neighbors. *Nat Struct Mol Biol*, 16(2):151–158, Feb 2009. 55
- [135] Schatz D.G. Odegard V.H. Targeting of somatic hypermutation. *Nat. Rev. Immunol.*, 6:573–583, 2006. 57
- [136] Palo Alto CA USA. <http://www.pymol.org> DeLano, W.L. The PyMOL Molecular Graphics System. (2008) DeLano Scientific LLC. 57
- [137] Jayanth V Chodaparambil, Andrew J Barbera, Xu Lu, Kenneth M Kaye, Jeffrey C Hansen, and Karolin Luger. A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nat Struct Mol Biol*, 14(11):1105–1107, Nov 2007. 64
- [138] D. Pruss, B. Bartholomew, J. Persinger, J. Hayes, G. Arents, E. N. Moudrianakis, and A. P. Wolffe. An asymmetric model for the nucleosome: a binding site for linker histones inside the dna gyres. *Science*, 274(5287):614–617, Oct 1996. 64
- [139] Daniel M. Quinn. Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. *Chem Rev*, 87:955–979, 1987. 75
- [140] M. Jucovic and R. W. Hartley. Protein-protein interaction: a genetic selection for compensating mutations at the barnase-barstar interface. *Proc Natl Acad Sci U S A*, 93(6):2343–2347, Mar 1996. 75
- [141] Johan Elf, Gene-Wei Li, and X. Sunney Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, May 2007. 75
- [142] Sanbo Qin and Huan-Xiang Zhou. Prediction of salt and mutational effects on the association rate of ula protein and ul small nuclear rna stem/loop ii. *J Phys Chem B*, 112(19):5955–5960, May 2008. 75, 76
- [143] R. R. Gabdoulline and R. C. Wade. On the protein-protein diffusional encounter complex. *J Mol Recognit*, 12(4):226–234, 1999. 77, 80
- [144] Yossi Shaul and Gideon Schreiber. Exploring the charge space of protein-protein association: a proteomic study. *Proteins*, 60(3):341–352, Aug 2005. 80
- [145] Raik Grünberg, Johan Leckner, and Michael Nilges. Complementarity of structure ensembles in protein-protein binding. *Structure*, 12(12):2125–2136, Dec 2004. 80
- [146] Jennifer M Bui, Zoran Radic, Palmer Taylor, and J. Andrew McCammon. Conformational transitions in protein-protein association: binding of fasciculin-2 to acetylcholinesterase. *Biophys J*, 90(9):3280–3287, May 2006. 80, 104
- [147] Jennifer M Bui and J. Andrew McCammon. Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci U S A*, 103(42):15451–15456, Oct 2006. 80, 104, 108, 115

BIBLIOGRAPHY

- [148] Razif R Gabdouliline and Rebecca C Wade. Biomolecular diffusional association. *Curr Opin Struct Biol*, 12(2):204–213, Apr 2002. 80
- [149] R. R. Gabdouliline and R. C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophys J*, 72(5):1917–1929, May 1997. 80
- [150] Stephen E Halford and John F Marko. How do site-specific dna-binding proteins find their targets? *Nucleic Acids Res*, 32(10):3040–3052, 2004. 82, 84, 113
- [151] Michael Slutsky and Leonid A Mirny. Kinetics of protein-dna interaction: facilitated target location in sequence-dependent potential. *Biophys J*, 87(6):4021–4035, Dec 2004. 82
- [152] Vahid Rezania, Jack Tuszynski, and Michael Hendzel. Modeling transcription factor binding events to dna using a random walker/jumper representation on a 1d/2d lattice with different affinity sites. *Phys Biol*, 4(4):256–267, Dec 2007. 82
- [153] Longhua Hu, Alexander Y Grosberg, and Robijn Bruinsma. Are dna transcription factor proteins maxwellian demons? *Biophys J*, 95(3):1151–1156, Aug 2008. 82
- [154] Tao Hu and B. I. Shklovskii. How a protein searches for its specific site on dna: the role of intersegment transfer. *Phys Rev E Stat Nonlin Soft Matter Phys*, 76(5 Pt 1):051909, Nov 2007. 82
- [155] Annika Wedemeier, Ting Zhang, Holger Merlitz, Chen-Xu Wu, and Jörg Langowski. The role of chromatin conformations in diffusional transport of chromatin-binding proteins: Cartesian lattice simulations. *J Chem Phys*, 128(15):155101, Apr 2008. 82
- [156] Ramzi Alsallaq and Huan-Xiang Zhou. Protein association with circular dna: rate enhancement by nonspecific binding. *J Chem Phys*, 128(11):115108, Mar 2008. 82
- [157] N. Metropolis, A.W. Rosenbluth, and M.N Rosenbluth. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953. 88
- [158] T.E. Cheatham III C.L. Simmerling J. Wang R.E. Duke R. Luo M. Crowley Ross C. Walker W. Zhang K.M. Merz B. Wang S. Hayik A. Roitberg G. Seabra I. KolossvÁary K.F. Wong F. Paesani J. Vanicek X. Wu S.R. Brozell T. Steinbrecher H. Gohlke L. Yang C. Tan J. Mongan V. Hornak G. Cui D.H. Mathews M.G. Seetin C. Sagui V. Babin D.A. Case, T.A. Darden and P.A. Kollman. Amber 10. *University of California, San Francisco*, 2008. 101
- [159] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*, 79:926–935, 1983. 101
- [160] T. Wocjan, K. Klenin, and J. Langowski. Brownian dynamics simulation of dna unrolling from the nucleosome. *J Phys Chem B*, 113(9):2639–2646, Mar 2009. 103
- [161] Nils B Becker and Ralf Everaers. Dna nanomechanics: how proteins deform the double helix. *J Chem Phys*, 130(13):135102, Apr 2009. 108, 115

- [162] David D Boehr, Ruth Nussinov, and Peter E Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, 5(11):789–796, Nov 2009. 115

Erratum

Page 27: The sentence “In case of biomolecules, this reduction of degrees of freedom ‘freezes’ global structural changes, as side chains movements, accompanying the diffusional motion.” should be removed.

Page 31: The sentence “All the forces are *modelled* on a 3-dimensional grid, *which encompasses both* biomolecules, and at each Brownian step the force between ...” should be read “All the forces are *computed* on a 3-dimensional grid *encompassing this biomolecule*. At each Brownian dynamics step, the force between ...”

Page 33: The sentence in the text below Eq. 4.13 “The Debye-Hückel parameter is κ , α is a scaling factor and the second term in Eq. 4.12 is computed *as* ...” should be read “The Debye-Hückel parameter is κ , α is a scaling factor and the second term in Eq. 4.12 is computed *analogously to* ...”

Page 65: A sentence should be added before the sentence “The times for S1S2, SITE1, SITE2 and...”, which reads “The ratio of the WT experimental FRAP time at 80 % to 50 % recovery is approximately 5 Å and since the same ratio is obtained for the simulation residence times at 100 Å to 60 Å, these distances were chosen for comparison and analysis.”

Page 69: Fig. ?? should be read Fig. 6.12.

Page 93, line 8: The frequency of mode 8 should be read “2.23 1/cm” instead of “1.46 1/cm”.

Page 93: The times in Eq. 8.7 should be multiplied by 0.5.

Page 94, line 13: The sentence “This plot (Fig. 8.2) ...” should be read “This plot (Fig. 8.1) ...”.