Università degli Studi di Genova
Istituto Italiano di Tecnologia

---

Dipartimento di Fisica

# Modelling Electrostatic Interactions and Solvation in Chromatin:

## from the single nucleosome towards the chromatin fibre

Artemi Bendandi

Dottorato in Fisica

XXXIII Ciclo

*Thesis Advisors:*
Prof. Alberto Diaspro
Dr. Walter Rocchia
Dr. Silvia Dante

Για τη Βουρβουρού,
για τους νεκρούς που μου πήρε ο Άδης,
και για τη θεά Άρτεμη.

*All my life, I could feel an insistent gnawing in the back of my mind.*
*It was a yearning, a thirst for discovery that could be neither numbed, nor sated.*

# Abstract

Chromatin is a complex of proteins and DNA found in the nuclei of eukaryotic cells. It reinforces the DNA and its topology tunes DNA transcription and gene expression. It is formed by nucleosomes, structures composed of an octameric protein core and approximately 147 base pairs of DNA. Chromatin is an extremely complex system, the behaviour of which is ruled by both mechanical and electrostatic factors that are depend on its structure, and biomolecular interactions occurring in the cell nucleus. In this thesis, I analyse chromatin compaction from an electrostatic perspective and focus on the role of electrostatics and solvation as determinants of the topology of chromatin. I examine the effect of the histone tails and propose a methodology to connect electrostatic calculations to the structural and functional features of protein-DNA systems. This methodology can also be combined with coarse-grained representations. I study the electrostatic forces acting on the phosphate atoms of the DNA backbone. I investigate the electrostatic origins of effects such as different stages in DNA unwrapping, nucleosome destabilisation upon histone tail truncation, and the role of specific arginines and lysines undergoing Post-Translational Modifications. I find that the positioning of the histone tails can oppose the attractive pull of the histone core, locally deform the DNA, and tune DNA unwrapping. I conduct an analysis of the porosity of nucleosomes and related to the importance of solvation phenomena. I complement and support my computational findings on nucleosome electrostatic interactions experimental Zeta Potential and Dynamic Light Scattering measurements on single nucleosomes under varying ionic concentrations, providing information on the surface charge and the size of nucleosomes. I present a comprehensive study of the electrostatic interactions between nucleosome pairs sampling different translations and rotations. My analysis aims to provide a cohesive description of nucleosome electrostatic interactions in the chromatin fibre, combining information on the energetics of different relative positions of nucleosomes, especially in very tight packing situations. In addition to numerical estimates of electrostatic interaction energy of nucleosomes at different relative distances and orientations, obtained within the Poisson-Boltzmann framework, I present their approximation by analytical asymptotic expressions, where nucleosomes are approximated as monopoles and dipoles centred in dielectric spheres immersed in an electrolytic solution. I am able to identify a non-linearity region around the nucleosomes, and to exploit the fact that that in points outside that region the electrostatic potential can be described by the linearised Poisson-Boltzmann Equation.

# Acknowledgements

Deep breath and on with everyone's favourite part of a thesis!

A kind of conclusion has been reached. It's not the conclusion the characters necessarily had in mind. It's not one that ties up every plot thread. No one in our story has enough power to make that happen. But the conclusion we've carved out reaches some kind of resolution and leaves us a bit better off than when we started. It is time for me to give credit where credit is due.

To Alberto Diaspro, for giving me opportunities that changed my entire life.
To Walter Rocchia, for everything.
To Silvia Dante, my scientific mom, for being immensely supportive and incredibly good to me.
To Aymeric Le Gratiet, for believing in me, for never forgetting where he comes from, and for our work together.
To Colin Sheppard, for being an inspiration, as a scientist and as a person.
To Alessandro Patelli, for being a comrade in science, and for beers at Sat.
To Andrea Spitaleri, for teaching me how to use GROMACS without prejudice, with unwavering patience, and with a smile.
To Sergii Siryk, for the things I could never have done myself.
To Miguel Soler and Nicola Scafuri, for the talks and the dinners.
To Roberto Mulargia, for helping me get through living in Genoa.
To Eleonora Uriati, for reminding me that I can be a force for good in this world (and for being our favourite pet sitter).
To Riccardo Marongiu, for being my "compagno di scrivania" and a nakama these years.
To Chiara Povero, for encouraging me with her words and inspiring me with her actions.
To my partner and my best friends, for knowing what I am and loving me anyway.
To Cake, my familiar, who can't read this because he's a cat, for being the only constant in my ever-changing world.
Alla mia famiglia, per le cose che hanno fatto bene, e per le cose che hanno sbagliato.

*In the beginning, there was darkness. And in addition to the darkness there was also me. And I moved upon the face of the darkness. And I saw that I was alone. Let there be Light.*

# Research Statement

The research I conducted during the three years of my PhD in Physics in the University of Genoa is categorised into two projects, carried out under the guidance of Prof. Alberto Diaspro in the Nanoscopy Lab & Nikon Imaging Centre, in collaboration with Dr. Walter Rocchia and the Concept Lab, both in the Italian Institute of Technology (IIT), in Genoa, Italy.

My primary research, culminating in this document, concerns the role and importance of electrostatic interactions in chromatin compaction, as part of a wider project investigating the factors that tune the process of folding in chromatin. This work was directly supervised by Dr. Walter Rocchia (computational and analytical parts), and by Dr. Silvia Dante (experimental part). The research conducted for this project was motivated by the observation that chromatin, a protein-DNA complex, is a highly charged system in which elevated electric charges come into very close proximity. Electrostatic interactions between histones and DNA and the contribution of solvation, are therefore fundamental determinants of the topology of chromatin compaction.

Our analysis of nucleosome electrostatics gave rise to a detailed analysis of the electrostatic interactions between the histone tails, the disordered terminal domains of the histone proteins, and the nucleosomal DNA. I developed a methodology for the study of protein-DNA electrostatic interactions, and applied it to the interactions between the histone tails and nucleosomal DNA. My method can be used to correlate electrostatic interactions to structural and functional features of protein-DNA systems, and can be combined with coarse-grained representations. In particular, I focused on the electrostatic field and resulting forces acting on nucleosomal DNA as a result of a spatial rearrangement of the histone tails. I investigated the electrostatic effects on different stages of DNA unwrapping, nucleosome destabilisation upon histone tail truncation, and the role of specific arginines and lysines undergoing Post-Translational Modifications. I validated my results with comparisons to previous experimental and computational observations.

We aim to provide a conclusive description of electrostatic interactions between nucleosomes, as a basis for the development of a mesoscale coarse-grained force field for chromatin. To this end, I measured the energy changes involved in electrostatic interactions between nucleosomes at different relative distances and rotations. This approach also gives us information on the most favourable nucleosome positions in close distances. I developed an analytical expression for the electrostatic interaction energy between two charged dielectric spheres immersed in an electrolyte, based on the Poisson Boltzmann Equation (PBE). I represented nucleosomes as spheres with a monopole and a dipole of appropriate intensity placed in their centre. I parametrised my model with numerical data on the electrostatic potential, calculated with the DelPhi PBE solver. I found that the particles are surrounded by a "non-linearity layer", where the electrostatic potential is high enough for the full non-linear PBE treatment to be required. Beyond this potential surface, I can describe electrostatic interactions between the particles using the linearised PBE, which is far less demanding in terms of computational resources.

We performed accompanying and complementary experiments, using the Zeta Potential method, with which I explored the dependence of the electrostatic potential on the shearing plane of the nucleosomes at varying ionic conditions. I correlated these measurements with Dynamic Light Scattering measurements, a method that gives us access to measurements of the mean radius of the nucleosomes to investigate the ionic concentration in which nucleosomes start to unravel and agglomerates start forming. I compared these experimental results with computational calculations, using the PBE, showing once again that the use of the full non-linear PBE is necessary when studying electrostatic properties of nucleosome interactions in their immediate vicinity.

During my three years as a PhD candidate in IIT I also took part in a research project concerning the development of a mathematical formalism based on the manipulation of the Mueller matrix with the so called coherency matrix and its applications in label-free microscopy. The determination of the coherency matrix elements from the experimental Mueller matrix is of interest due to the fact that this method provides the capability of dealing with scattering and noisy measurements for extracting crucial localised information on the structural organisation of complex random media at the sub-micrometer scale. This work is carried out by Prof. Colin Sheppard of the University of Wollongong, NSW, Australia, and Nanoscopy Lab & Nikon Imaging and Dr. Aymeric Le Gratiet of the Nanoscopy Lab & Nikon Imaging Centre. We investigated the properties of the coherency matrix, and interpreted the physical information on light that can be extracted from this analysis, studying different microscopy architectures in transmission, backscattering, and reflectance, with a particular emphasis on highly scattering media. We successfully applied the coherency matrix analysis on zebrafish samples at different stages of embryonic development, achieving improved contrast with respect to the straightforward Mueller matrix images, and obtaining optical segmentation of the samples. We compared the coherency matrix method with the gold standard Lu-Chipman decomposition of the Mueller matrix, both in a theoretical and an experimental framework. We showed that the coherency matrix method is more general, providing a sensitive discrimination of the experimental errors mixed with random complex media fingerprints. In addition, we demonstrated that our innovative method provides local tracking of polarimetric changes in a complex system, applying a straightforward mathematical treatment which does not pose a necessity for a priori information on the sample. The findings of this project are not reported in this document.

# Contents

# List of Figures

# List of Tables

## CHAPTER 1

# Chromatin Multiscale Modelling

Chromatin is a complex of proteins and DNA found in the nuclei of eukaryotic cells. It is formed by nucleosomes, structures in which approximately 147 base pairs of DNA wrap around histone proteins like thread around a spool. It is an extremely complex system, the behaviour of which is tuned by both mechanical and electrostatic factors that are related to its structure, and biomolecular interactions in the cell nucleus.

The topology of chromatin compaction tunes DNA transcription and misfolding has been proven to be a crucial factor in cell malfunction and several diseases. Simulations provide extremely useful insights on the different mechanisms and factors influencing compaction, using different levels of approximation to represent the system. In this chapter we mention several computational works that used as inputs parameter sets acquired through experiments or evaluated their results by comparing them with preexisting experimental data. We provide a succinct synopsis of some among the existing modelling approaches for chromatin, focusing on the physics-based ones, and on those that allow integration with experimental biophysical and/or biological knowledge.

We start from a description of chromatin, and an analysis of the structure of nucleosomes, chromatin's building blocks. We then move on to discuss the need for multiscale modelling solutions for the issue of chromatin compaction, and we categorise them by order of magnitude, from single nucleosomes to entire genomes. We discuss full-atom simulations of nucleosomes and nucleosome assembly and disassembly, and the role of linker DNA and histone H1. We mention the interesting structural repercussions of DNA sequence specificity. Combining simulations results with various experimental techniques can help shed light on the main determinants of chromatin compaction.

Theoretical descriptions are, therefore, paralleled with experimental techniques providing instrumental information for the validation and improvement of these models, paying particular attention to methods that only minimally perturb the observed system. We conclude with a discussion on experimental techniques that have been used in chromatin studies.

The literature review presented in this Chapter is part of [1].

## 1.1 Chromatin and Nucleosomes

If one were to stretch the DNA found inside a cell nucleus, they would end up with an approximately 2-metre long fibre. In order to fit inside the cellular nucleus, which measures approximately 6 μm in diameter, DNA needs to compact itself in a manner that permits efficient accessibility to DNA-binding proteins, while at the same time reinforcing and compacting the fibre. Compaction is achieved through the wrapping of DNA around certain proteins, the histones, forming the building blocks of the chromatin fibre, the nucleosomes. Chromatin is a molecule that demands multiscale analysis since changes as small as the absence of one DNA base pairs (bp) in nucleosomal or linker DNA can cause non-local changes in the topology of the fibre. Given the fundamental importance of chromatin organisation regarding gene expression, the question of discovering the manner in which the genome folds and compacts itself is one of the most fundamental in Biology.

Chromatin can be categorised into two broad sub-types: euchromatin and heterochromatin. Euchromatin is less densely packed, contains more active genes, meaning that it is more often transcribed, and it is usually found in the interior of the nucleoplasm. Heterochromatin is more densely packed, contains mostly inactive genes, and is usually found in the nuclear periphery. However, that does not mean that heterochromatin does not influence cellular processes. In fact, the interactions of heterochromatin with the internal lamina of the nuclear envelope are particularly interesting as they have been seen experimentally to provide structural support to the nucleus, and can cause several pathologies when disrupted [3–6]. Naturally, nuclei are extremely crowded environments, and simulations have suggested that temporal changes in the mean density of the nuclear environment affect chromatin organisation, and therefore transcription [7], and highlight the key role of entropy as a driver of the self-organisation of genomes in loops [8]. The role of entropic and kinetic effects in genome organisation through the mechanism of bridging-induced attraction mediated by DNA binding proteins as observed in Brownian dynamics and Molecular Dynamics (MD) simulations is discussed in [9] and [10].

The chromatin fibre has been widely proposed to assume two main forms: the 11-nm and the 30-nm fibre. Chromatin was believed to assume the latter configuration, but in recent years it has been suggested by experimental observations that chromatin structure is more dynamic and that various forms are found in the nucleus [11, 12]. These observations indicate that 30-nm fibre requires several factors to be present in vivo, such as a relatively low ionic concentration, a regular linker DNA length distribution, a lack of significant histone tail modifications such as acetylation, and no notable histone depletion in nucleosomes. These are naturally numerous requirements for a live cell to satisfy at the same time. New experimental methods attempt at shedding light on this discrepancy, such as ChromEMT, Chromatin Electron Microscopy Tomography [13]. With this method, a disordered 5nm to 24nm fibre architecture was observed. Studies have also attempted to use notions of fluid phase-separation to describe chromatin organisation at the level of the nucleus [14]. Overall, the study of chromatin is an intrinsically multiscale endeavour, since the effects of interactions spanning from atomic to chromosome-level influence

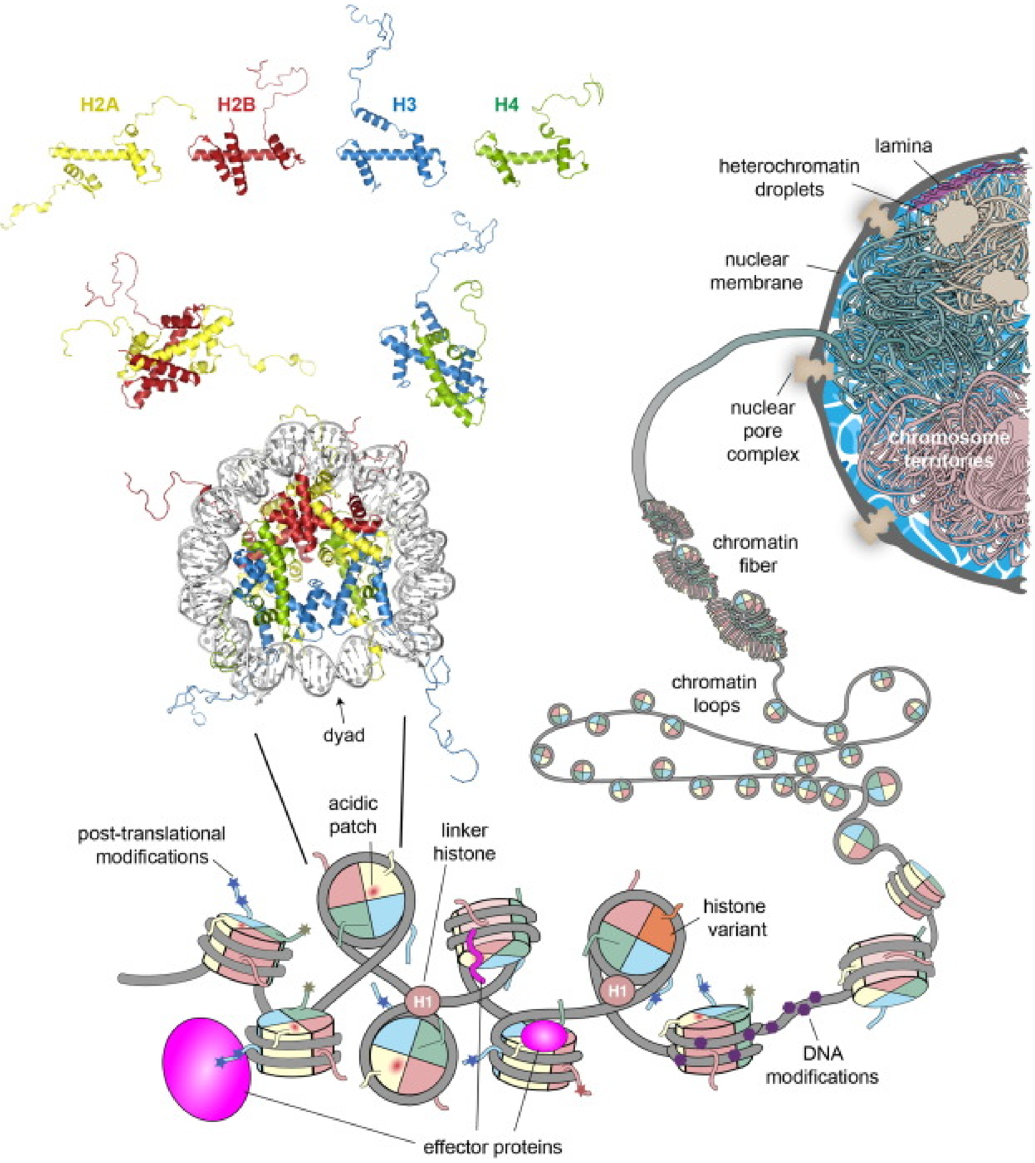


Figure 1.1: Schematic representation of chromatin structure. At the first level of DNA packing, we have nucleosomes, spools of protein and DNA. At the second level of packing, nucleosomes adopt a "beads on a string" configuration, forming the chromatin fibre. The tertiary level of packing, the way in which the chromatin fibre folds upon itself, remains one of the most important open questions in Biology. Illustration from [2]

its topology. Chromatin polymorphism is mostly driven by the delicate equilibrium of electrostatic interactions, solvation effects and mechanical constraints, such as steric exclusion and linker DNA length.

Nucleosomes are composed of a protein core, the histone octamer (consisting of H2A, H2B, H3, and H4 histone dimers), and 147 bp of DNA wrapped around the core in 1.64 turns. Each histone of the octameric core has a highly disordered N-terminal portion, the histone tail. The core domains of the histones form alpha helices [15] connected by short

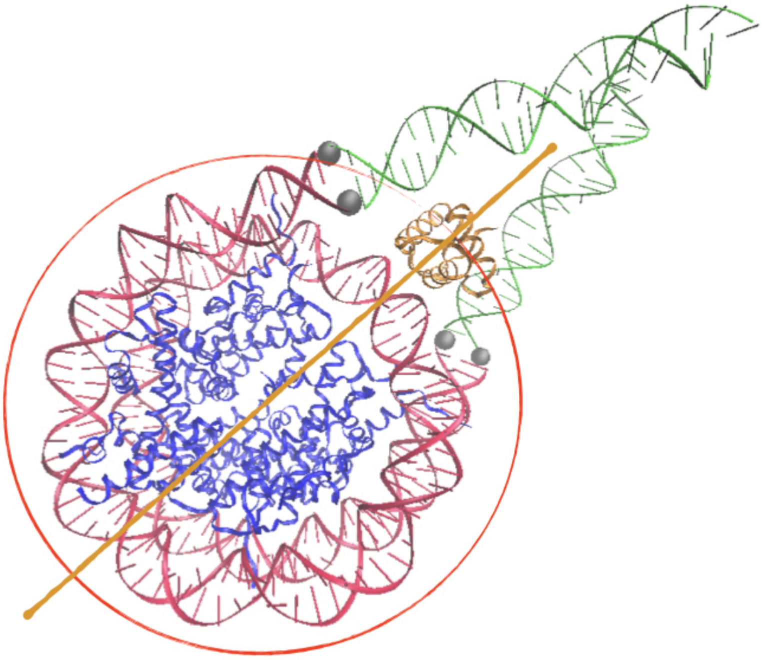


Figure 1.2: Nucleosome crystal structure with two 25bp-long linker DNA segments in complex with linker histone H1 (PDB code 5NL0). Nucleosomal DNA (147 bp) is coloured in cherry red, while linker DNA is coloured in green. Linker DNA connects each nucleosome to its neighbouring nucleosomes, and its length can vary across different organisms, or even cell types in the same organism and genomic locations in the same nucleus. Linker DNA length is one of the parameters that tune chromatin topology. DNA entry and exit sites, the points in which linker DNA is attached to the nucleosome, are denoted by silver spheres. The histones belonging to the core octamer are coloured in blue, while the linker histone is coloured in orange. The dyad axis, the pseudosymmetry axis dividing the plane of the nucleosome is drawn as an orange line. Figure constructed with VMD.

loops 1.2, mainly composed of positively charged residues [16]. Two more tails extend from the C-terminals of H2A histones, amounting to a total of ten unstructured dynamic domains [17], which contain approximately 30% of total histone mass. The histone tails bear a high concentration of positive charge and, along with counterions, help neutralize the high negative charge of the DNA backbone, generally promoting nucleosome stability [18]. Nucleosomes are connected to each other by varying lengths of linker DNA strands 1.2, but it has been calculated that the spooling of DNA around nucleosomes alone makes DNA shorter by 7 times [19]. The mean diameter of a nucleosome is 11nm, from which the 11-nm chromatin fibre term originates.

All of the N-terminal histone tails are rich in glycine residues, the backbone carbonyls of which can form specific contacts with DNA phosphates. The flexibility of glycines facilitates changes in the local curvature of nucleosomal DNA [20–22]. Arginine residues are also of particular interest, as they are positively charged and have been known to stabilise the DNA around the nucleosome at 14 contact points [23]. Nucleosomes are by no means static entities - on the contrary, they participate in phenomena such as nucleosome sliding, a process in which DNA gradually repositions itself around histones, while maintaining contact with the histone core. [24, 25] They also undergo partial or total disassembly [26, 27], in order to tune transcription, DNA damage response, and gene expression [28–30].

While Nucleosome Core Particles (NCPs) have been observed to be rather stable in dynamics, the histone tails present high variability in their conformations, as expected of intrinsically disordered domains. Shaytan et al.[20] performed a 1 µs-long full atom MD simulation, studying the dynamics of NCPs at varying ionic concentrations, observing, among else, that the majority of the contacts between the histones and the nucleosomal DNA are due to the histone tails, which rapidly adsorbed on the DNA surface during the simulation. The interactions and the number of contacts between histone tails and DNA strongly depend on their position with respect to the nucleosome dyad axis, and on the number of arginine, glycine and lysine residues. For example, the H3 N-terminal and H2A C-terminal tails were observed to form numerous contacts with the linker DNA and stabilise the DNA entry and exit regions of the NCP. Moreover, H3 tails have been observed in experiments [31] to form stable folded structures, possibly due to their high glycine residue content [32]. Histone tail interactions with the nucleosomal DNA can also locally alter the latter's geometry; for example, in a 100ns-long full-atom MD simulation of NCPs, it was seen that the largest fluctuation of the DNA groove width was caused by an arginine residue of the H2A N-terminal tail probing the DNA backbone [16].

## 1.2 Multiscale Modelling of Chromatin

The simultaneous advances in computational and experimental resources not only led to significant milestones, but have also opened new possibilities in chromatin studies. Because of the intrinsic multiscale nature of chromatin, there is a plethora of computational and experimental approaches, which focus on structures as small as the single nucleosome and its dynamics, up to the entire genome of an organism. These models try to describe and predict experimental observables such as different fibre-start patterns, as well as the effect of different linker DNA lengths on fibre topology. For chromatin modelling, especially at small and intermediate scales, approaches that rely on basic physical interactions for the description of electrostatics and solvation are of uttermost importance. The other indisputably essential ingredient is the mechanical connection; for example, the presence of high-curvature AT-rich segments (A-tracts) in linker DNA is known to influence nucleosome interaction and alter chromatin folding [33].

Chromatin models can be divided into two general categories, depending on the underlying initial assumptions and on the chosen building blocks: bottom-up and top-down models [34]. The preferred approach depends on the level of detail of interest, the level of theory that one wants to adopt for the model and, inescapably, the computational capabilities at hand. Bottom-up models take the nucleosome and linker DNA crystal structures as a starting point (Fig.1.1.A). The electrostatics and dynamics of these structures may be studied at the full atom level, and the derived results can be used to feed a coarse-grained model, which allows to draw conclusions for larger systems, such as oligonucleosomes or, sometimes, even larger structures (Fig.1.1.B) [35–39]. The parameters used in these coarse-grained models depend on the properties of interest and on those observed by the accompanying experiments. In order to parametrise these types of models, data is often used from all-atom structures and simulations, making their results dependent on the resolution of the structures and the performance of the force fields used.

In top-down models, the behaviour of the fibre is deduced from experimental observations and sequencing of large regions of chromatin, or even of the entire genome, from which a scheme of interactions is derived. Given the limitations in resolution and accuracy of experiments, top-down models cannot possess the same level of detail as bottom-up models. However, they provide a way to study global chromatin properties. These models may incorporate a multitude of, often ad-hoc, coarse grained descriptions to look into very specific chromatin features related to smaller scale structures, such as the kbp scale. Finally, in this category of models the use of notions from polymer physics is very common, representing chromatin as a polymer chain and its stages of compaction as phase transitions, imposing constraints in the forms of potentials. [40, 41]

Alternatively [42], chromatin models have been divided in categories based on whether they are built to match pre-existing data or emerge as representations of physical properties: data-driven models and ab initio models. Regarding data-driven models, some examples are given by approaches that try to generate chromosome structures based on Hi-C maps [43], translating contact probability to distance. In these cases, however, one needs to bear in mind that Hi-C maps, and sequencing techniques in general, often give an average picture of the genome. Ab initio models, on the other hand, take properties that have been observed or even hypothesised about chromatin as a starting point, and aim to reproduce them through the application of constraints and potentials [44, 45]. The mathematical nature of these models can sometimes lead to a simplification of biological factors at play.

Here, bearing in mind these general classifications, which are consistent with model classifications in many fields, we propose an exploration of various models based on the final order of magnitude that they are able to study, ranging from mononucleosome studies up to works examining the entire genome. Examining different orders of magnitude of chromatin, we present approaches that make use of different assumptions and are based on different types of data, illustrating the multifaceted nature of the topic. An overview of different modelling paradigms based on the order of magnitude at interest is shown in Fig.1.3.

### 1.2.1 From the Single Nucleosome to Oligonucleosome Fibres

Nucleosomes have the ability to dissociate entirely in histones and DNA, upon unwrapping, and then reassemble [46]. The curvature of the DNA can either favour or disfavour histone-DNA contacts, and therefore the formation of nucleosomes [23]. Based on this premise, starting our analysis from the building blocks of chromatin, we encounter Partially Assembled Nucleosome States (PANS), which are interesting as they reveal the electrostatic and mechanical changes that occur when a nucleosome is forming or dissolving. [26] analysed three types of PANS (hexasomes, tetrasomes, and disomes) through MD simulations, visualising the structures with Atomic Force Microscopy (AFM) experiments. The nucleosome formation procedure was observed to occur as such: the two H3 and H4 dimers bind to the DNA first, forming a tetrasome, followed by the sequential addition of H2A and H2B dimers. The results were compared to Small Angle X-ray Scattering (SAXS), Förster Resonance Energy Transfer (FRET), and AFM data. Nucleosome disassembly follows the reverse order, and both assembly and disassembly were seen to be associated with DNA supercoiling, as a way to regulate torsional stresses on the fibre [47].

Linker DNA length is extremely important for chromatin compaction, not only for mechanical but also for electrostatic reasons. Determining how linker DNA influences chromatin topology, and how its length and sequence can affect compaction has been the subject of much study and speculation. In the work of Buckwalter et. al [33], for instance, the presence of so-called A-tracts, DNA segments where multiple A-T pairs are present in a row, and their influence on DNA rigidity, and therefore on chromatin fibre flexibility, are examined. It has been observed by comparison of Monte Carlo (MC) simulations and Electron Microscopy (EM) experiments on reconstituted oligonucleosome arrays that the presence of A-tracts causes DNA bending angles of up to 90°, and that these particular segments are often found in linker DNA [48]. The direction of bending of the linker DNA is also relevant for compaction: for example, when DNA bends inwards at the exit sites from the NCP the resulting structures are more compact compared to the opposite case, and produce to zig-zag configurations and closer overall nucleosome proximity. It is evident that linker DNA length is of great importance when it comes to chromatin topology; however, its role is not immediate; the really important parameters for packing are the DNA bending angles, which are influenced by linker DNA length through topological and persistence length constraints.

The presence of the linker histone H1 (or H5 in avian chromatin) is also a key for compaction 1.2. This histone is not always present in nucleosomes, and its position can vary on or off the nucleosome dyad axis (Fig.1.2), the axis of symmetry of the nucleosome [50]. The H1/H5 changes the orientation and flexibility of linker DNA, forming contacts with both entering and exiting strands. When two or more nucleosomes in sequence are bound to H1 histones, rigid structures termed DNA stems are formed, which present straighter linker DNA and reduced separation angle between the entering and exiting DNA; the latter effect is more pronounced in chromatin configurations with long linkers [36]. The increased rigidity of DNA because of the formation of DNA stems is mitigated by the dynamic nature of H1/H5 binding and unbinding on nucleosomes [51].

Most all-atom and coarse grained models dealing with chromatin simulations require the use of empirical force fields at some point, impacting on the simulation results. Even though an extensive critical comparison of force fields and force field modifications for nucleic acids is beyond the scope of this introduction, we suggest the works of [52] and [53].

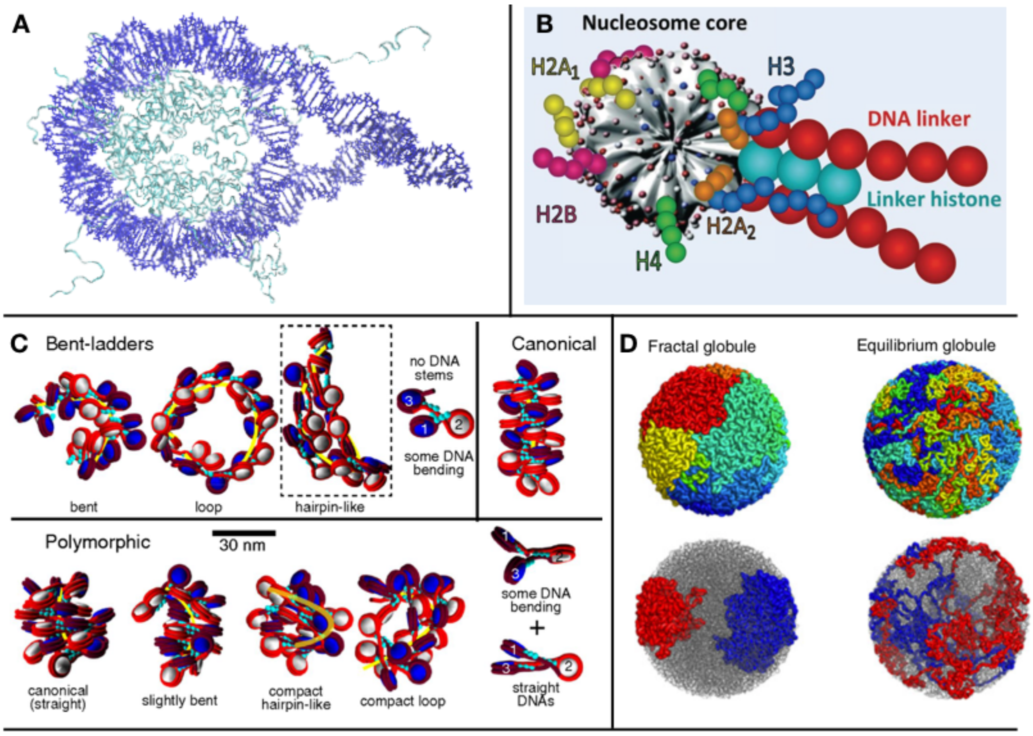


Figure 1.3: Treating different orders of magnitude in chromatin requires different levels of detail in the representation of nucleosomes: **A** for one to four nucleosomes, crystal structures can be used (structure used by [20], visualized with VMD); **B** for longer structures, a coarse graining model is required (such as Schlick's group model from [36], **C** which can be used to study the topology of oligonucleosome fibres. **D** In larger scales, where even the entire genome can be studied, fractal models are used [49].

### 1.2.2 Coarse-grained Oligonucleosome Models

According to the number of nucleosomes in the start of a fibre, different behaviours have been observed, the most prominent of which are the zig-zag and solenoid fibre models [33]. zig-zag models for chromatin propose what is commonly called a two-start fibre model (two nucleosomes at the start of the fibre), in which linker DNA crosses the main fibre axis. In two-start zig-zag models, nucleosomes are stacked in the periphery of the fibre and linker DNA occupies the central space of the structure. Solenoid models on the other hand propose compaction through coiling of the linker DNA along the superhelical path. In these models, fibres are one-start, and nucleosomes create frontal contacts, with 6 to 8 nucleosomes per turn of the fibre. It is thought that both models coexist in fibres, along with straight linker DNA and bent linker DNA. [54, 55] Contrary to the zig-zag fibre, where the dominant interactions are $n \pm 2$, in solenoid models they were found to be $n \pm 5$ or $n \pm 6$ [54, 56] where n represents the position of the reference nucleosome. Different oligonucleosome fibre topologies have also been studied in conjunction with *polar* and *antipolar* ordering of the nucleosomes' dyad axes using group theoretical analysis within the Landau–de Gennes framework [57].

Besides the number of nucleosomes at the start of the fibre, and taking into consideration the fact that linker DNA length is not always the same across the fibre, different Nucleosome Repeat Lengths (NRL) produce different fibre configurations, and alter the propensity of a fibre to unfold. In [36], Collepardo-Guevara and Schlick performed MC simulations on coarse-grained oligonucleosome fibres (Fig.1.1.C) to study these variations, and observed a variety of structures, reaching the – perhaps not surprising – conclusion that structures with highly varying NRL were more compact than uniform structures, a direct consequence of fewer topological constraints. In relation to gene expression, the study also found that transcriptionally active cells presented shorter NRLs, while in inactive cells the opposite has been observed [58]. In the coarse-grained model, shorter NRL fibres arranged in ladder-like forms, while medium fibres arranged in zig-zags and longer NRLs resulted in heteromorphic structures [54].

Nucleosomes bearing histone modifications, or even less histones than the canonic octamer [59] have also been studied as a factor influencing chromatin compaction. In this study by [60], a genome folding model was constructed using Monte Carlo (MC) simulations and introducing histone and nucleosome depletion. In a subsequent paper, the role of epigenetic modifications regarding nucleosome depletion was investigated, and MC data was compared to 5C and Fluorescence In Situ Hybridization (FISH) [61]. Even though full atom models are very instructive in the mononucleosome scale, in certain mesoscale chromatin models [46], DNA base pairs are represented as rigid bodies, with parameters that account for orientation and displacement. Oftentimes, in more coarse-grained models, nucleosomes are treated as rigid bodies with concentrated charge and the dynamics of the histone tails are modelled as Gaussian distributions or as series of beads. In some models [40, 62] chromatin is represented as an inextensible chain of beads, whose distance depends on the spatial scale of the desired simulations.

Works like [63, 64] aim to optimise chromatin morphology through studying its dependence on linker DNA elasticity and length, introducing the role of inter-NCP interaction potentials in the packing of the fibre. Such works often use MC or Brownian dynamics simulations [65, 66] and model electrostatic interactions based on potentials at various levels of sophistication. In [63], the chromatin fibre is constructed as a helical array by cyclically repeating a fundamental structure, defined as two nucleosomes and the linker DNA between them, in which nucleosomes are treated as rigid bodies and linker DNA as a series of beads. As we mentioned previously, histone modifications are also relevant factors for chromatin compaction, and are sometimes used as model parameters. An example of histone modifications as model parameters is [67], a polymer MC coarse-grained model using methylation as a parameter to study chromatin dynamics and conformation statistics.

In [68] so-called two-angle models were developed, using linker DNA entry and exit angles and NCP twist angles, generating ensembles of minimum energy conformations through MC and analysing their dynamics through Brownian dynamics. NCP geometry becomes itself a parameter in several works [46, 62, 69], in which internucleosomal interactions are specifically studied as triggers for compaction. When it comes to the representation of the NCP as a rigid body, shapes such as an oblate ellipsoid or an oblate spherocylinder are more accurate than simple spheres. In [62] Kepper et. al applied a coarse-grained computer model to a sample pool of 101 nucleosome arrays, using different chromatin models with and without the presence of linker DNA. It was shown that nucleosome spacing is relevant to chromatin stability, with the highest destabilisation occurring at a 2bp shift, by analysing energy landscapes. Energy variations were compared to values from chromatin stretching experiments [70]. After surpassing the 2bp energy barrier, nucleosome repositioning towards a new conformation, rather than returning to the original one, becomes more energetically favourable. Nucleosome orientation and tilting was also shown to be of importance [71], since, for example, it was observed that in cases where a nucleosome was oriented transversally it occupied more volume and caused its neighbours to be pushed further apart, hindering close packing.

### 1.2.3 Topological and Fractal Models

During the past decade, great progress has been made in the study of chromatin organisation due to the advent of Chromosome Capture technologies (3C). The field was particularly revolutionised by Hi-C, which provides the interaction frequencies between loci of an entire genome. 3D reconstructions of genomic regions and even entire genomes are possible, using Hi-C data, through structural inference and statistical methods [72, 73]. There are two main categories of techniques to generate 3D structures from Hi-C contacts: ensemble approaches and consensus approaches. In the latter case, the Hi-C data are considered as a single ensemble, while in the former models different categories of structures are created from the data. It has been suggested recently that it might be possible to reconstruct the diploid 3D chromatin structures [74].

It can be of interest to combine results from high throughput techniques, such as Hi-C, with computer simulations. Ohno et. al drew parallels between protein structure and chromatin [75]. Through a combination of Hi-C data at nucleosomal resolution obtained at several cell phases and coarse grained simulations, Ohno et al. observe two general secondary structure types in chromatin, which they call $\alpha$-tetrahedron and $\beta$-rhombus, as an analogy to the $\alpha$-helix and $\beta$-sheet structures in proteins, supporting the claim that fibres can alternate between these structures when nucleosome positioning changes. Information on nucleosome orientation was gleaned through analysis of the spatial proximity between DNA entry and exit points in individual nucleosomes across the genome and their 3D positioning. Solvation effects were not directly taken into account, as nucleosomes were modelled as space-filling objects, and linker DNA was also implicitly treated. Nucleosome positioning data has been shown to be a valid predictor of chromatin interaction patterns in simulations of micro-domain (0.5-10kbp) formation in yeast [76].

In the study of compaction and larger scale interactions within the chromatin fibre, for example for characterizing the Topologically Associating Domains (TADs), loop extrusion models are very significant. TADs are regions of the genome with enhanced contact frequency, identifiable on Hi-C maps as squares. During loop extrusion, Loop Extrusion Factors (LEFs), such as cohesin, interact with chromatin, inducing the formation of loops until they encounter a Border Element (BE), such as CTCF. Cohesin is a multi-protein

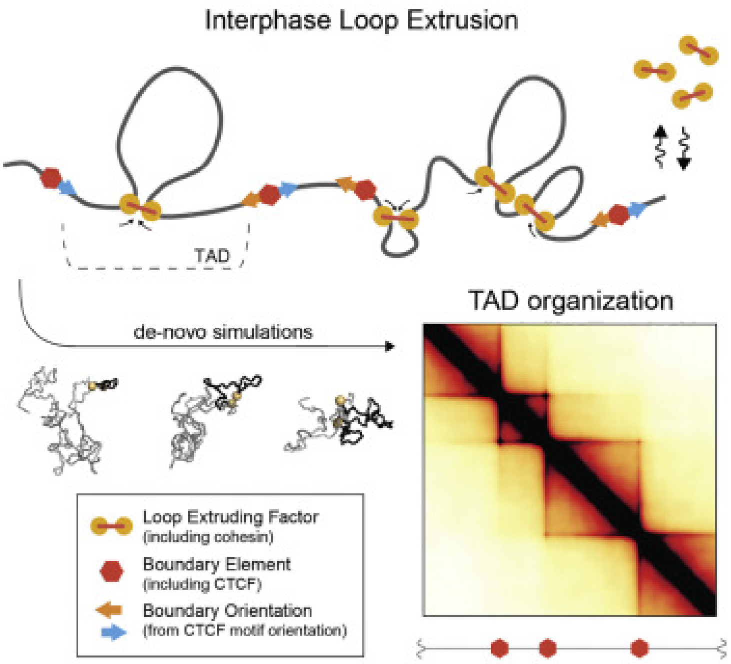


Figure 1.4: Loop extrusion as a mechanism of TAD formation in chromatin, through the combined action of cohesin and CTCF DNA-binding proteins. A single cohesin molecule embraces two different points in the fibre, creating loops that grow until they are halted by CTCF, which functions as a boundary element. Illustration from [43].

complex involved in establishment and maintenance of sister chromatid pairing during DNA replication and mitosis, and CTCF is a transcriptional repressor, which regulates 3D chromatin structure, anchors DNA to cellular structures, defines boundaries between euchromatin and heterochromatin, and promotes or represses gene expression. There are two main categories of loop extrusion models: embrace models, in which a cohesin ring traps two DNA strands with a closed (d~35 nm) or partially open (d~20 nm) centre (rod-like conformation), and handcuff models, where two cohesin rings interact with DNA and each other, capturing one DNA strand each.

It has been observed by [77] that macroscopic loop characteristics in loop extrusion models depend on the abundance of LEFs. Loop extrusion models provide explanations for experimental observations, such as the preferential orientation of CTCF, the enrichment of TAD boundaries in proteins with architectural functions, and TAD merging in LEF deletion experiments, and could provide insight on chromosome-level phenomena [43]. Polymer simulations are frequently used by loop extrusion models to make predictions and to validate analytical models. Loop formation has also been studied with mesoscopic models, where it was observed to depend on linker histone presence, ion concentration, and linker DNA length [78]. In addition to the "one-sided" loop extrusion mechanism described above, recent research indicates that "two-sided" loop extrusion might prove to be more robust in explaining experimental data [79]. For a review of loop extrusion models and a juxtaposition with other large-scale models of chromatin, particularly models based on the action of transcription factors, we point the reader to [80].

In the last decade, there has been growing interest on fractal models describing chromatin, and part of the chromatin modelling community, particularly emerging from polymer physics, has been focusing on the possibility that chromatin organises itself as a fractal, especially since a similar state has been proposed in the seminal paper of the Hi-C method by [81]. In this work, a distinct case of the previously theorised globular equilibrium model was proposed for the Mbp scale: the fractal globule – otherwise called crumpled globule [82], a polymer conformation that enables maximally dense packing while preserving the ability to easily fold and unfold any genomic locus [49, 81, 83] (Fig.1.1.D). In such models, as in polymer models for chromatin in general, chromatin is considered as a flexible polymer fibre, and the notion of the single nucleosome is lost. Because of their large scope, these kinds of models can be relevant for large scale systems or even the entire genome.

Distinct chromosomal regions can be modelled as equilibrium globules, structures used to describe polymers in poor solvent conditions [81]. The chromatin fibre could assume a Peano curve conformation, which represents a continuous fractal trajectory that densely fills space without crossing itself. In fractal globules, compaction is achieved through the collapse of the globule and it has been shown that the fractal globule has the ability to organise territorially, alluding to chromosome territories, [83] distinct regions in the nucleus occupied by certain chromosomes, in contrast with the previously proposed equilibrium globule, which does not present such organisation. In the fractal globule, the number of interactions as a function of volume shows a linear correlation, which leads to the interdigitation of different regions in the globule with each other, allowing for extensive genomic cross talk [49] (Fig.1.1.D). This is particularly interesting for two main reasons: cross talk has been observed in simulations between the regions, and fractal globules unfold in an optimal way, which is relevant in the study of transcription.

However, it needs to be noted that the fractal globule is a metastable state, unlike the equilibrium globule, and that its lifetime depends on topological constraints, which, in real cells, can be affected by enzymes and DNA-binding proteins. Fractal globules have been observed experimentally in Hi-C experiments [38, 81, 84] and Small Angle Neutron Scattering (SANS) experiments [19, 85]. The relationship between the physical environment of a fractal chromatin fibre and transcription has been studied in several works, such as [86], in which the analytical correspondence between changes in the fractal dimension of the chromatin fibre and increment of chromatin accessibility and compaction heterogeneity was studied. Furthermore, the authors speculated that differences in the transcription of a certain gene might be influenced by folding of neighbouring genomic regions. The findings were supported by microscopy measurements on cancer cells.

Fractal globule models have been criticised based on the argument that self-similarity cannot be assessed in only a couple of orders of magnitude. However, researches in the field, such as [87] claim that, even though mathematical fractals are self-similar ad infinitum, physical fractals are only self-similar within certain orders of magnitude, typically 2 or 3, while chromatin architecture spans 4 or more orders of magnitude, and a common fractal architecture would connect all of them under a single topological theme, without the need for separate structures in each order of magnitude.

## 1.3 Experimental Studies of Chromatin: from the Nucleosome to the Nucleus

Throughout this introduction, we have highlighted the main manifestations of the multiscale nature of chromatin, and we have explored the multitude of factors affecting its compaction. The interplay between simulations and experiments is crucial to reach a deep understanding of this complex system, and has given rise to breakthroughs that would have been impossible without the combination of the two approaches. Experimental investigations of chromatin can be carried out at different scales, similarly to computational approaches. Having already mentioned some experimental results validating computational models, we have specifically looked into some of the experimental techniques used in both small and large scales, from the NCP up to entire nucleus.

Starting from the nucleosome, experiments have been carried out to determine its crystal structure, with continuing endeavours starting from [88], in which a 2.8Å resolution structure of the NCP was obtained via X-ray crystallography, using reconstituted nucleosomes. In Luger's work, many of the structural elements of the nucleosome were uncovered, such as the the number of base pairs wrapped around the octamer, which were unknown despite the fact that the octamer histone structure had already been observed. The histone tails and their structural role have also been studied to great extent in [89]. Since then, further structures with 147 bp [90] and 146 bp [91] have been observed. The study of sub-structures such as the histone tails and of site-specific interactions [2] in more detail, required the use of NMR [90].

In latest years, there has been growing interest for the study of NCPs using Cryo-EM (Fig.1.5.A). The sample preparation protocols involved in this technique make it an interesting alternative to X-ray crystallography for structural studies. Cryo-EM provided information on custom-made NCPs in studies relevant to DNA binding protein-NCP interactions [92] and also on interactions of the NCP with components of the nuclear environment, such as the nuclear pore complex [93]. The orientation of NCPs has also been observed by Cryo-EM in a recent study, where it is stated that in the most common arrangement of a pair of NCPs they are placed in parallel, facing histone octamers [94].

X-ray crystallography provides structures with atomic resolution, which are key for atomic-level studies. However, this approach has some limitations; it fails to provide good information on the more mobile domains of the NCP, and it cannot be used for large oligonucleosomes (the largest structures that have been crystallised to date are tetranucleosomes [98, 99]. In order to circumvent these constraints, one can turn to scattering techniques. SAXS studies have looked into the issue of whether the histone tails protrude into the solvent surrounding the NCP or associate with DNA at physiological salt conditions. The histone tails are notoriously hard to resolve in crystallography because of their size and intrinsically mobile nature [31, 100, 101].

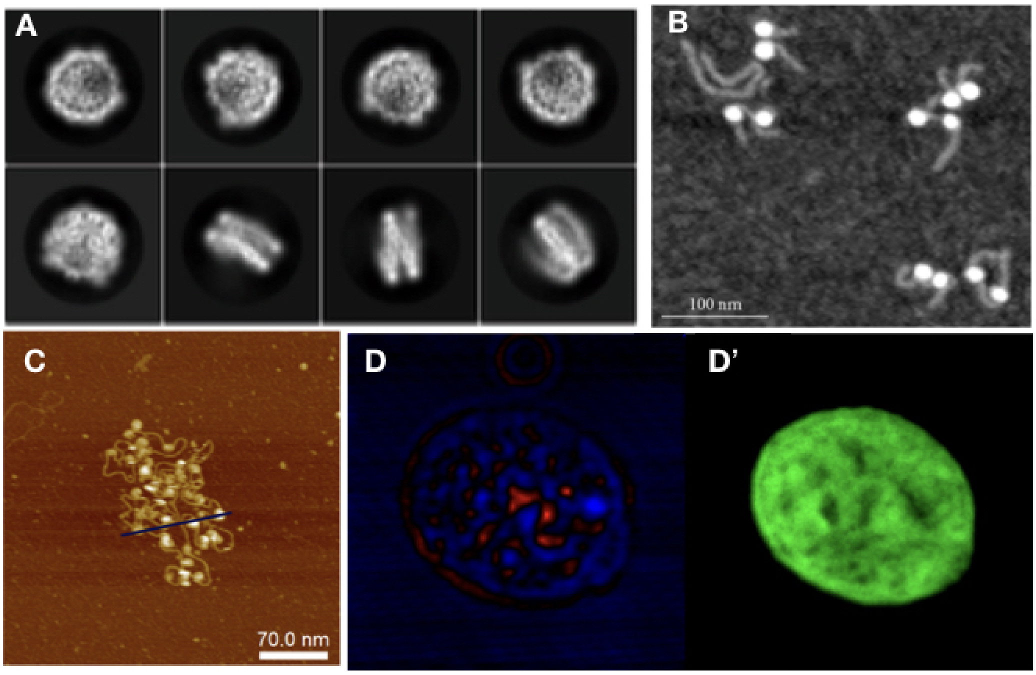


Figure 1.5: As with modelling approaches, in experiments different techniques are required to study different orders of magnitude in chromatin: **A** NCP imaged with Cryo-Em (adapted from [93], **B** NCPs with histone tails AFM image [95], **C** Nucleosome array, AFM image (adapted from [96], **D** Isolated Hek nucleus imaged with CIDS, labelled with Hoechst for chromatin-DNA organisation imaging. The fluorescence labelling (D') is used as a fingerprint of the chromatin to demonstrate the correlation with the label-free approach using circular polarisation excitation. [97].

Using SAXS however, it is possible to indirectly observe whether the histone tails are solvated or adherent to the DNA, by measuring changes in the overall structure size. [102] have applied SAXS to study histone tails as well, focusing on the structural details of inter-nucleosomal interactions and the effects that histone tails have on them. Often SAXS has been used in conjunction to other techniques to correlate structural to dynamical data. Mauney et. al used data from SAXS, FRET, and MD to dissect the sequence-dependent DNA unwrapping mechanism [103]. Fluorescence Correlation Spectroscopy (FCS) has been used in a work by [35] to estimate NCP stacking energy. In this combined experimental and theoretical work, model parameters were tuned based on comparison with single molecule FCS and SAXS data, which also showed that histone tails facilitate NCP stacking by acting as bridges between NCP surfaces. FCS data was also used by [104] to tune the parameters of an MC model of nucleosome arrays under the influence of external forces.

Moving on from NCPs to larger structures, nucleosome arrays are the next step; besides SAXS [105], AFM has also been used to study arrays of varying lengths (Fig.1.5.B). The advantage of using this technique for chromatin is twofold: there is the possibility of taking many measurements, making it good for statistical purposes; and it allows for the study of electrostatic and related interactions, such as differences in ionic concentration. The importance of ionic interactions with chromatin has naturally gained the attention of the experimental community. Studies such as [106] have shown that $Mn^{2+}$ ions bind to the major DNA groove near CG pairs. In [96] AFM was used to measure the changes in chromatin topological conformations depending on salt levels in the environment (Fig.1.5.C). Studying NCP arrays in varying salt concentration revealed that array compaction has a non-monotonic salt dependence. Increasing salt concentration induces partial screening of the charges of the DNA backbone, therefore reducing the electrostatic interactions between DNA and histones, directly impacting on compaction. The stability of mononucleosomes has also been investigated in correlation with salt concentration [107]: in low to intermediate salt regimes they observed some partially disassembled states (as also studied computationally by Rychkov et. al in [26], where H2A/H2B histone dimers partially dissociate from the NCP). Regarding the mechanical properties of chromatin, DNA stiffness was observed to be salt-dependent as well, in accordance with other experimental and computational studies [21, 108–110]; the persistence length was seen to increase at higher ionic concentrations.

Optical microscopy, a field traditionally tied to biological applications, is a natural candidate for chromatin studies, due to the advances in resolution obtained by super-resolution techniques, and to the fact that label-free optical microscopy methods have been on the rise for the past decade. Experiments using the single molecule super-resolution microscopy technique STORM [111] have observed units of chromatin organisation termed by the authors *clutches*, heterogeneous groups of various sizes. The size of the clutches has been speculated by Ricci et al. to be related to the pluripotency capacity of each cell, and the median number and nucleosome density in the nucleus was found to be cell-specific. From longer nucleosome arrays to chromatin fibre, other super-resolution techniques, such as Photoactivated Localization Microscopy (PALM) have been used to extrapolate chromatin topology in the nucleus from nucleosome dynamics. Label-free techniques are also used to study chromatin at the nuclear level, such as Circular Intensity Differential Scanning (CIDS) by Le Gratiet et. al [97]. In this work, it is shown that the main advantage of this polarimetric method compared to standard fluorescence microscopy is the capability to obtain specific contrast mechanisms due to the chiral organisation of the DNA in a label-free approach without a priori knowledge of the sample. Indeed, it is shown that the stronger signal region corresponds to more compacted DNA region, i.e. heterochromatin, while the weaker signal, such as for the nucleoli, corresponds to a lower compaction, i.e. euchromatin region (Fig.1.5.D).

Experimental validation has been attempted also for some among the most exotic theoretical models proposed for chromatin, namely those hypothesising fractal globules. Fractal globules have been observed experimentally in Hi-C experiments [81] and Small-Angle Neutron Scattering (SANS) experiments [19, 85]. The important question tackled by works on this topic is the way in which fractal states with stable long-lived properties are formed. SANS has been considered a good technique for experiments looking for fractal structures in the nucleus because of its extended spatial range, from approximately 15 nm to 10 μm. The use of Cryo-Electron Tomography (Cryo-ET) has provided insight on the structure of mitotic chromosomes in fission yeast [112]. SAXS and Cryo-EM have also been used in structural analysis of the fibre up to the chromosome level (Fig.1.5.A) [11, 113–115].

Finally, we provide a table (Table 1.1) summarising the main nucleosome crystal structures that will be mentioned in this work and used in our analysis. We indicate the PDB code of the structure, its main structural features (e.g. the presence of linker DNA, a twist defect, histone tails etc.), and the organism from which the DNA and histones of each structure originate. In this work we perform full-atom electrostatics calculations on nucleosomes, so in this table relative information (the number of atoms and electric charge) is provided. The electric charge is provided as a multiple of the charge of an electron ($q_e \simeq 1.602 \times 10^{-19}$C). Finally, we provide the main reference for each structure. Unless explicitly mentioned, the structures listed do not include the histone tails.

| Structure | Features | DNA Organism | Histone Organism | Number of Atoms | Electric Charge | Publication |
|---|---|---|---|---|---|---|
| 1KX5 | histone tails | *Homo Sapiens* | *Xenopus Laevis* | 25086 | -144 | [116] |
| 5NL0 | linker DNA, linker histone | Synthetic | *Xenopus Laevis* | 38190 | -461 | [117] |
| 3AFA | twist defect | *Homo Sapiens* | *Homo Sapiens* | 21717 | -216 | [91] |
| 2PYO | | *Homo Sapiens* | *Drosophila Melanogaster* | 22014 | -216 | [118] |

Table 1.1: Summary reference table of nucleosome structures used in this work. Information on the particular structural features of each structure, the organisms from which the DNA and histones originate, the number of atoms, and the electric charge is included, along with the relative publications.

To conclude this Chapter, we present a summary table of the models mentioned, categorised by the final order of magnitude that they treat (e.g. single nucleosomes, oligonucleosome arrays, entire genome). We include information on the computational methods used, and, when available, the type of experimental data used for result validation.

| Scale | Publication | Subject | Computational Techniques | Experimental Data |
|---|---|---|---|---|
| NCP | [35] | Ionic dependence of aggregation | Langevin MD | FCS |
| | [119] | Ion condensation, NCP Solvation | MD, PBE | PDB structure |
| | [33] | Sequence dependence of DNA curvature | MC | EM |
| | [110] | Sequence dependence of DNA electrostatics | PBE | PDB structures |
| | [20] | Histone tail interaction | MD | |
| | [26] | NCP Assembly | MD | SANS, FRET, AFM |
| | [90] | NCP Solvation | | NMR |
| | [88] | Nucleosome structure | | X-ray crystallography |
| | [102] | Histone tail interaction | | SAXS |
| Nucleosome arrays | [60, 61] | Histone and nucleosome depletion | MC | FISH |
| | [36] | NRL-produced patterns | MC | |
| | [62] | NRL-produced patterns | MC | Stretching experiments |
| | [104] | Nucleosome array unwrapping | MC | FCS |
| | [69] | Inter-NCP interactions | MC | Data from various techniques |
| | [106] | Ionic dependence of aggregation | PBE, DiSCO | Data on DNA bending |
| | [54] | Linker histones, ionic dependence | MC | EM |
| | [37] | Electrostatics, histone tails, linker DNA | PBE | Cryo-EM |
| | [63] | Linker DNA | Energy optimisation | EM, FCS |
| | [78] | Loop formation | MC | 3C |
| Entire Genome | [75] | 3D genome architecture | Hi-CO method | Hi-C |
| | [74] | Hi-C 3D reconstruction | Mathematical modelling | Hi-C |
| | [81] | Fractal globule | Polymer simulations | Hi-C |
| | [19] | Fractal globule | Mathematical analysis | SANS |
| | [38] | Chromosome modelling | Coarse-grained Polymer modelling | Hi-C |
| | [43] | TADs | Polymer model | Hi-C |
| | [40] | TADs | Polymer model | 3C, FISH |
| | [41] | TADs | Polymer model | 5C |
| | [111] | Nucleosome aggregation | | STORM |
| | [97] | Chromatin organisation in the nucleus | | CIDS, Fluorescence |

Table 1.2: Computational and experimental works mentioned in this chapter (partial account), listed under the scale of interest. Computational techniques and experimental data used are listed, when applicable.

CHAPTER 2

# Electrostatic Interactions in Chromatin

In this Chapter we focus on the role of electrostatics and solvation as the driving mechanisms of chromatin conformational changes and as determinants of the relative equilibria. We examine the fundamental importance of electrostatic interactions in chromatin, and their impact on fibre compaction and polymorphism. This brings us to an exploration of the often underrated role of solvation in chromatin compaction. We also introduce and discuss the Poisson Boltzmann Equation (PBE), the equation that describes the interactions of charged particles in solution. This description is widely used in computational biophysics to describe the behaviour of biomolecules. We discuss some of the situations in which the PBE can be solved analytically, and the DelPhi PBE solver.

Electrostatics in chromatin encompasses an intricate combination of different mechanisms and the importance of its role in compaction and chromatin remodelling is paramount. The high negative charge of the DNA is partially neutralised by the direct interaction of the latter with histones (including the effects of histone tails and the linker histone), but electrostatic stabilisation of the chromatin fibre is achieved through a combination of this effect with long-range electrostatics and solvent screening. Simulations in which ionic interactions with chromatin at the NCP level are treated more accurately would be a great improvement to existing approaches.

In addition, a more accurate representation of the nucleosome core is crucial when performing these analyses, since solvation has proved to be a very important factor in nucleosome behaviour, whereas neglecting these effects would hamper a correct understanding of chromatin compaction. To complete our analysis we provide a description of the electrostatic features of nucleosomes, focusing on the acidic patch, a region on the surface of the histone core that possesses negative surface charge density, and the channel that traverses the nucleosome core, which is an important avenue for ion permeation.

Sections 2.2 and 2.3 of this Chapter have been published in [1].

## 2.1 Electrostatic Interactions of Charged Systems in Solution

In electrostatics, the relation between the electric potential $\psi(r)$ and the charge distribution $\rho(\vec{r})$ at any point in space $\vec{r}$ is given by the Poisson equation:

$$\nabla^2 \psi = -\frac{4\pi}{\epsilon_0 \epsilon_r} \rho(\vec{r}), \tag{2.1}$$

where $\epsilon_0$ is the dielectric constant of vacuum $\epsilon_r$ is the relative dielectric constant of the medium. Using the Dirichlet or Neumann boundary conditions, one can calculate the potential or electric field on a closed surface, which corresponds to a charge density distribution.

The Poisson equation (2.1) determines the electric potential for a fixed spatial charge distribution. However, it does not take into account the mobility of ions in the solution. Let us assume a system in which there are two monovalent ion types, positive and negative, and that the system is in thermal equilibrium and the continuum hypothesis is valid. The local density distributions of the two ion types $n_\pm(\vec{r})$ (given in units of number per unit volume) adjust to the values of the electric potential. In thermal equilibrium, the electro-chemical potential $\mu_i$, defined as $\mu_i = ez_i\psi + T \ln n_i$, is a constant. Consequently, the ion densities in the solution follow a Boltzmann distribution which is a function of the value of electric potential:

$$n_i = n_0^{(i)} e^{ez_i\psi/T}. \tag{2.2}$$

### 2.1.1 Helmholtz, Gouy-Chapman & Stern Theories

In situations where the surface of a charged object is exposed to an electrolyte, an electrical double layer forms around the object. In the Helmholtz theory (1853), the interface between a conductor and an electrolyte is composed of two layers of opposite polarity but equal charge. The term "double layer" is used to indicate the layer of charges in the surface of a conductor, therefore anions, and the rigid layer formed by cations already present in the solvent, attracted by the charged surface. The equivalent of this double layer configuration is a parallel plate capacitor. However, the constant capacity of the interface that this model predicted failed to describe real systems.

In the Gouy-Chapman theory (1910), the double layer is represented by a diffuse model: the charge distribution of the ions as a function of the distance from the charged surface follows Maxwell-Boltzmann statistics, and the electrostatic potential decreases exponentially as the distance from the charged surface increases. The net electric charge in this diffuse layer is equal in magnitude to the surface charge but opposite in polarity, so that the sum of the charge in the interior of the delimiting Gouy plane is zero. In this model, therefore, the distribution of the ions is governed by both electrostatics and statistical mechanics. However, ions are considered to be point charges, and would therefore be theoretically able to approach arbitrarily close to the charged surface, therefore causing the capacity of the interface to increase without limit.

The Stern theory (1924) for the electrical double layer combines the Helmholtz and

Gouy-Chapman models, incorporating a layer of immobile atoms around the surface, the Stern layer, and a diffuse layer where ions are progressively more influenced by thermal motions the farther they are from the charged surface. In the first layer the potential decreases linearly, and then in the second layer it decreases exponentially as the distance from the charged surface increases. The Stern layer accounts for the finite size of the ions, and its width corresponds to their mean radius. This separation of layers also implies a separation of capacities, as the two layers can be represented as two parallel plate capacitors in series, addressing the issues of the Helmholtz and Gouy-Chapman models. However, the behaviour of the adsorbed ions cannot be described purely through electrostatics, as their positioning is governed by chemical effects of specific adsorption, which need to be taken into account as well. Due to these specific effects, not all adsorbed ions are going to be necessarily of opposite charge to the surface charge of the solvated object.

### 2.1.2 Debye-Hückel Theory

Debye and Hückel (DH) proposed a continuum method for the calculation of the electrostatic free energy of small spherical ions in solution in 1923 [120], treating the ionic solution as a continuum characterised by a uniform dielectric constant. Limiting their analysis to monovalent symmetric electrolytes (such as NaCl and KCl), DH calculated the electrostatic energy of a single ion in solution using the mean-field potential of all other ions present in the solution. They proposed an equation that describes the electrostatic potential inside the ionic solution based on Gauss' law for the electrostatic potential and the Boltzmann distribution for the ions in the solvent.

The fundamental equation of DH theory is a three-dimensional second order nonlinear partial differential equation describing the electrostatic potential $\psi$ at a position r. Let us study a system composed of an electrolyte solution and a continuous charge distribution, a molecule, immersed in the solution. In the special case of a 1:1 electrolyte, for the dimensionless potential $u(r) = e_c k_B^{-1} T^{-1} \psi(\vec{r})$ this equation can be written as follows:

$$- \nabla \cdot (\epsilon(\vec{r}) \nabla u(\vec{r})) + \kappa^2(\vec{r}) \sinh\big(u(\vec{r})\big) = \frac{4\pi e_c^2}{k_B T} \sum_{i=1}^{Nm} z_i \delta(\vec{r} - \vec{r}_i), \tag{2.3}$$

where $\kappa = \lambda_D^{-1}$ is the DH parameter. The molecule is represented as $N_m$ point charges $q_i = z_i e_c$ at positions $\vec{r}_i$. The permittivity $\epsilon(\vec{r})$ is $\epsilon_m$ inside the molecule, and $\epsilon_w$ in both the solution and the ion-exclusion layer surrounding the molecule, the Stern layer. We define the modified DH parameter $\overline{\kappa}(\vec{r})$, which is proportional to the ionic concentration and dielectric independent, and takes the values $\overline{\kappa}(\vec{r}) = \sqrt{\epsilon_w}\kappa$ in the solution and $\overline{\kappa}(\vec{r}) = 0$ in the molecule. Eq.(2.3), the full nonlinear PBE, is often approximated by the linearised PBE, valid in low electrostatic potential regimes:

$$- \nabla \cdot (\epsilon(\vec{r}) \nabla u(\vec{r}) + \kappa^2(\vec{r}) u(\vec{r})) = \frac{4\pi e_c^2}{k_B T} \sum_{i=1}^{Nm} z_i \delta(\vec{r} - \vec{r}_i). \tag{2.4}$$

### The Debye screening length

Let us discuss the physical properties of the Debye screening length $\lambda_D$. Let there be an electrode of potential $\psi = 0$ immersed in an electrolyte solution at potential $\psi$. Using eq.(2.4) and letting z denote the axis perpendicular to the electrode, the potential can be described as:

$$\psi(z) = \psi(0)e^{-\kappa z}. \tag{2.5}$$

We see that the electrostatic potential of the single ion decreases in an exponential manner with a characteristic length $\lambda_D = 1/\kappa$. As the right hand side of eq.(2.4) denotes the source term of Poisson equation (2.1) we can now calculate the ion density

$$\rho(z) = -4\pi l_B \kappa^2 \psi(z)e^{-\kappa z}. \tag{2.6}$$

The Debye length $\lambda_D$ is therefore the distance at which the ion density drops to 1/e of its original value. Assuming a spherical particle, the potential decreases linearly from the surface of the particle up to the Stern plane and then decays exponentially in the diffuse layer in the Stern model, as $\psi(x) = \psi_\delta e^{-\kappa x}$, where $\psi_\delta$ is the potential on the Stern plane, $\kappa$ is the DH parameter and $x$ is the radial distance from the particle surface. Since the DH parameter is directly proportional to the ionic concentration, and the value of $1/\kappa$ gives us the thickness of the double-layer, we see that when the ionic concentration increases the double layer decreases in size, and the electrostatic potential should decrease as well.

## 2.1.3 Mathematical Derivation of the Poisson Boltzmann Equation

The model assumed by DH is given in Fig.2.1 and the derivation follows [121]. Divide space in three regions: $\Omega_1$, $\Omega_2$, and $\Omega_3$. The continuous charge distribution that generates the electrostatic potential defines region $\Omega_1$. Region $\Omega_3$ consists of the solvent, characterised by a dielectric constant $\epsilon_3$, assumed to contain mobile ions. Region $\Omega_2$ represents the Stern layer around the macromolecule in which no mobile ions are present, in which $\epsilon_2 = \epsilon_3$. Assuming that all mobile ions in the solution are monovalent, they can be treated as positive and negative charges $+e_c$ and $-e_c$, where $e_c$ is the charge of an electron. The electrostatic potential satisfies Gauss' law in each of the three regions.

Define a coordinate system in $\mathbb{R}^3$, and define a translation vector $\vec{r} = (x, y, z)$. Represent a continuous dielectric distribution as a sum of $N_m$ discrete charges $q_i$ at positions $\vec{r}_i$, where $q_i = z_i e_c$, $zi \in R$, $i = 1, ..., Nm$. The potential in region $\Omega_i$ is given by

$$\psi_1(\vec{r}) = \sum_{i=1}^{Nm} \frac{q_i}{\epsilon_0 \epsilon_1 |\vec{r} - \vec{r}_i|}. \tag{2.7}$$

Taking into account the form of the Green's function for the 3D Laplacian $(-4\pi|\vec{r} - \vec{r}_0|)^{-1}$, we apply the Laplacian to both sides of eq.(2.7) and obtain

$$\nabla^2 \psi_1(\vec{r}) = \sum_{i=1}^{Nm} -\frac{4\pi q_i}{\epsilon_0 \epsilon_1} \delta(\vec{r} - \vec{r}_i), \tag{2.8}$$

where $\delta(\vec{r} - \vec{r}_i)$ is the Dirac delta function.

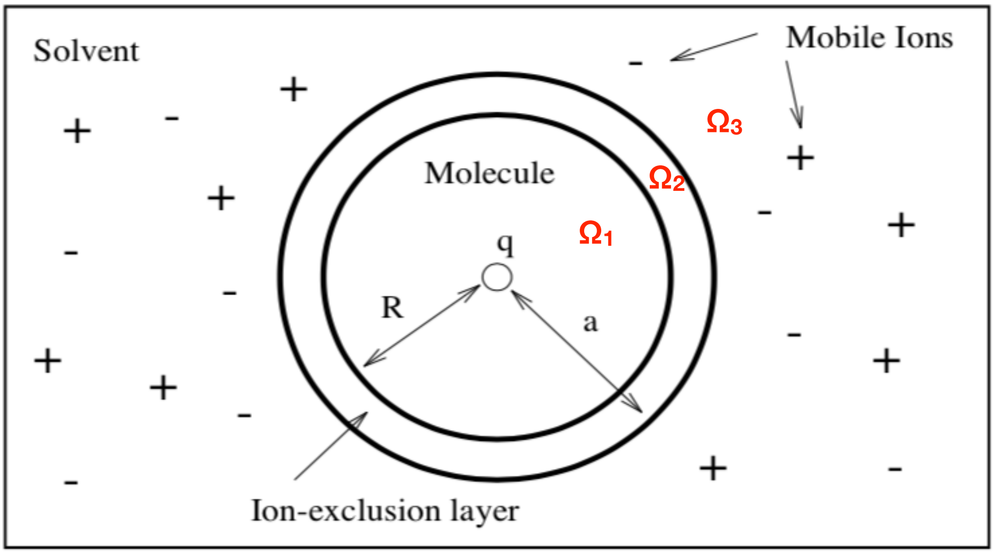


Figure 2.1: Graphical depiction of the DH theory model: a spherical dielectric particle $\Omega_1$ immersed in a dielectric solution of monovalent ions $\Omega_3$. An ion exclusion layer $\Omega_2$ termed the Stern layer is formed around the particle. Illustration adapted from [121].

In region $\Omega_2$, there are no mobile ions, and the charge density function is given by $\rho_2(\vec{r}) = 0$. Gauss' law for the electrostatic potential in region $\Omega_2$ becomes:

$$\nabla^2 \psi_1(\vec{r}) = -\frac{4\pi\rho_2(\vec{r})}{\epsilon_0 \epsilon_2} = 0, \tag{2.9}$$

In region $\Omega_3$, assume the presence of monovalent ions of charges $+e_c$ and $-e_c$. Let the bulk concentration be M per cubic metre for each of the two ions types. The number $M^+$ of positive ions and $M^-$ of negative ions per cubic metre will differ near the molecule. The fundamental assumption in DH theory is that the ratio of the concentration of one type of ion near the molecule in $\Omega_1$ to its concentration far from $\Omega_1$ is described by the Boltzmann distribution $e^{-W_i(\vec{r})/k_B T}$, where $W_i(\vec{r})$ is the work required to move the ion of type i from $|r| = \infty \longrightarrow (\psi(\vec{r}) = 0)$ to the point $\vec{r}$. In the case of two ion types, the necessary work is:

$$W_1(\vec{r}) = +e_c \psi_3(\vec{r}), \qquad W_2(\vec{r}) = -e_c \psi_3(\vec{r}), \tag{2.10}$$

for the positive and negative ions, respectively. Applying the Boltzmann distribution we then obtain:

$$M_+ = -M e^{e_c \psi_3(\vec{r})/k_B T} \qquad M_- = M e^{+e_c \psi_3(\vec{r})/k_B T}, \tag{2.11}$$

where we assume that $M_+ = M_- = M$ far from $\Omega_1$. The charge density at any point in $\Omega_3$ is given by:

$$\rho_3(\vec{r}) = M_+ e_c - M_- e_c = 2M e_c e^{-e_c \psi_3(\vec{r})/k_B T} = -2M e_c \sinh\left(\frac{e_c \psi_3(\vec{r})}{k_B T}\right), \tag{2.12}$$

The differential form of Gauss' law for $\Omega_3$ yields:

$$\nabla^2 \psi_3(\vec{r}) = -\frac{4\pi\rho_3(\vec{r})}{\epsilon_0\epsilon_3} = \left(\frac{8\pi M e_c}{\epsilon_0\epsilon_3}\right)\sinh\left(\frac{e_c\psi_3(\vec{r})}{k_B T}\right). \tag{2.13}$$

### Boundary Conditions

The product of the dielectric constant times the normal derivative of the potential $\epsilon\nabla\psi(\vec{r})\cdot\vec{n}$, where n is the unit outward normal vector, is also continuous at the interfaces. The discontinuity between the dielectric constants of the different regions is reflected on a discontinuity of the normal derivative of $\psi(\vec{r})$ at the boundaries. These considerations are summarised as follows:

$$\psi_1(\vec{r}) = \psi_2(\vec{r}), \qquad \epsilon_1\nabla\psi_1(\vec{r})\cdot n = \epsilon_2\nabla\psi_2(\vec{r}), \tag{2.14}$$

$$\psi_2(\vec{r}) = \psi_3(\vec{r}), \qquad \epsilon_2\nabla\psi_2(\vec{r})\cdot n = \epsilon_3\nabla\psi_3(\vec{r}). \tag{2.15}$$

To these boundary conditions we add the condition for the value of the electrostatic potential at infinity $\psi(\infty) = 0$.

### The non-linear and linearized forms of the PBE

We represent the dielectric constant as a piecewise function $\epsilon(\vec{r})$ on all $\Omega_i$ regions, and define the dielectric independent modified DH parameter $\overline{\kappa}(\vec{r})$:

$$\epsilon(\vec{r}) = \begin{cases} \epsilon_1 & \text{if } \vec{r} \in \Omega_1 \\ \epsilon_2 (= \epsilon_3) & \text{if } \vec{r} \in \Omega_2 \text{ or } \Omega_3 \end{cases}, \qquad \overline{\kappa}(\vec{r}) = \begin{cases} 0 & \text{if } s\vec{r} \in \Omega_1 \text{or } \Omega_2 \\ \sqrt{\epsilon_3}\kappa & \text{if } \vec{r} \in \Omega_3 \end{cases}. \tag{2.16}$$

We can now unify our considerations for the behaviour of the electrostatic potential in all three $\Omega_i$ regions in a single field equation, the nonlinear PBE (2.3):

$$-\nabla\cdot(\epsilon_0\epsilon(\vec{r})\nabla\psi(\vec{r})) + \overline{\kappa}^2(\vec{r})\left(\frac{k_B T}{e_c}\right)\sinh\left(\frac{e_c\psi(\vec{r})}{k_B T}\right) = 4\pi\sum_{i=1}^{Nm} q_i\delta(\vec{r}-\vec{r}_i). \tag{2.17}$$

In regions where the potential is low, we can use the series expansion $\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + ...$ as a linear approximation to $\sinh x$, obtaining the following linearized PBE (2.4):

$$-\nabla\cdot(\epsilon_0\epsilon(\vec{r})\nabla\psi(\vec{r}) + \overline{\kappa}^2(\vec{r})\psi(\vec{r})) = 4\pi\sum_{i=1}^{Nm} q_i\delta(\vec{r}-\vec{r}_i). \tag{2.18}$$

This approximation is used in situations where surface potentials are smaller than 25 mV at room temperature when treating biomolecules, as the value of the potential $\psi(\vec{r})$ has to be small with respect to $k_B T$.

## 2.1.4 Solutions to the Poisson Boltzmann Equation

The PBE is, broadly speaking, a good descriptor of the electrostatic potential in biomolecules, as its model is valid in physiological conditions (ionic concentration 0.145 M NaCl), in presence of Sodium Chloride salt in solution, and situations in which the surface potentials involved are not too large. However, being based on a mean field potential approach, if a greater level of detail and "granularity" is required, for example the treatment of very elevated surface potentials, or the explicit treatment of ions and solvent molecules, more detailed approaches are required.

In special cases, analytical solutions to the linearised PBE can be derived, in systems with particular geometries and with particular boundary conditions [122]. These cases mainly regard systems presenting spherical symmetry, such as a dielectric sphere immersed in an electrolyte [123]. Analytical solutions to the linearised and nonlinear PBE are quite complex, even in the few simple situations for which they exist [124–128]. Even in the cases where a solution of the linearised PBE is attempted, in a system possessing spherical symmetry, the Green's function is very convoluted, involving spherical harmonics, Bessel functions, Gegenbauer polynomials, and other such mathematical horrors (see Appendix B). In the simplest case of an infinite planar charge distribution, we can use the Guy-Chapman theory [129, 130]. However, in the more general case without assuming specific symmetries or boundary conditions, even the existence of a solution is debated, especially for the nonlinear PBE. Even if a solution exists in a particular system configuration, it might be extremely hard to derive it. Therefore, given the relevance of the results provided by the solution of the PBE and the difficulties in obtaining this solution, several numerical approaches have been developed in order to solve the PBE. The most commonly used are DelPhi [131], developed by Prof. Barry Honig's lab in Columbia University (NY, USA), and APBS [132]. The solver we have used for our analysis in this work is DelPhi.

Interactions between molecules are often described by empirical force fields, where the electrostatic terms treat the coulombic and van der Waals interactions between two charges, for all the charges contained in the system in question. In such approaches, for example in full-atom molecular dynamics simulations, solvent molecules are treated explicitly. However, these explicit solvent methods are often not applicable to large systems, because of limitations to computational resources. This leads to the development of implicit solvent methods, making use of the continuum approximation. In these methods, the solvent molecules and solvated ions are not taken in explicit consideration, but are regarded as a continuous dielectric distribution. A solute molecule immersed in a solvent of different dielectric constant can be described as a non-uniform medium. Although analytical solutions to the PBE, or even to the Poisson equation, are very challenging in systems with non-symmetric geometries, as we discussed previously, in principle it is possible for numerical approximations to describe the shape of the solute in atomic detail while imposing a simplified "continuum" description of the solvent.

The dielectric constant in continuum models practically accounts for the charge rearrangements that are not accounted explicitly in the model [133]. In this framework, biomolecules are usually, but not always, assigned a low dielectric constant value which accounts for electronic polarisability. When using DelPhi, the relative dielectric constant of proteins is usually assigned a value of 2. Water, on the other hand, has a very high polarisability ($\epsilon_w = 80$), causing a large upsurge between relative dielectric constants at the protein-water interface. Furthermore, this large difference in polarisability introduces a strong dependence on the shape of the molecule generating the electric field. As we have already seen, another complication to the determination of the electrostatic potential is the presence of mobile ions in the solvent, and it does not allow for a straightforward application of Coulomb's law. The basis of continuum solvent models is the reaction field, the favourable interaction between a fixed and an induced charge of opposite sign.

### 2.1.5 The DelPhi PBE Solver

DelPhi applies the finite volumes method to solve the PBE for molecules of arbitrary shape and charge distribution. The calculation is conducted according to the following procedure: first, DelPhi maps the atomic coordinates of a charge distribution onto a cubic grid. This charge distribution could be for example originated by the partial charges located in the atom centres on a biomolecule from the Protein Data Bank, or a geometric shape, such as a sphere containing a point charge. The solute is represented as a cavity in the solvent, containing fixed charges and dipoles. The shape of the cavity and the positions of the charges, depend on the charge distribution that the user has provided as input. The input for the case of a biomolecule will be for example the crystal structure. The solvent contains counterions to the fixed charges of the solute distributed according to the DH model.

The protein-solvent boundary is defined as the solvent accessible surface and it is determined by the Connolly algorithm [134]. In fact, the newest versions of DelPhi are interfaced with NanoShaper [135] and provide increased accuracy in the calculation of the potential with decreased grid resolution. NanoShaper is a program that constructs and analyses the molecular surface of nanoscopic systems. It provides volume and surface area for any considered molecular system, including that of internal cavities, and identifies pockets. NanoShaper is also interfaced with the widely used molecular visualisation software VMD. After the determination of the boundary between the two dielectric distributions, the dielectric constant is given an appropriate value at each grid point, depending on whether the point is found inside or outside the solute. The Debye parameter is set to zero for grid points in the interior of the solute, and it assumes the bulk value in the solvent, which is determined according to DH theory.

Let us look at how DelPhi assigns the charges to the grid points. Let $h$ be the spacing between the grid points. The centre of the coordinate system of the solute is placed at the geometric centre of the box. The coordinates of the molecule are then scaled such that the maximum linear dimension of the molecule corresponds to a predetermined fraction of the side of the box. In other words, the user can predetermine the percentage of the simulation box that is occupied by the solute. The charge on each charged atom is mapped to the grid by assigning fractional charges to the nearest six grid points, using a bilinear weighting function which preserves the nullity of the dipole moment of the mapped charge distribution. The fractional charge $q_f$ assigned to a certain grid point is given by:

$$q_f = (1 - a)(1 - b)(1 - c) \tag{2.19}$$

where $a$, $b$, an $c$ are the distances of the respective grid point in the three Cartesian directions, expressed as fractions of $h$. Mapping to a grid unavoidably causes computational artefacts, depending on the grid resolution and the quality of the construction of the interface between the two areas of different dielectric constant. Boundary grid points are assigned values of the potential in the beginning of the calculation, selected by the user.

After the determination of the system geometry and the assignment of charges and dielectric constants on the grid, DelPhi makes an initial estimate of the potential and then iteratively improves this estimate, updating the potential on each grid point by applying a stencil derived from the discretisation of the PBE, and according to a Gauss-Seidel scheme. The convergence criterion is that the mean square change in successive iterations is less than a preset value, typically 0.0001. For more information on the iterative convergence procedure, and on the features of DelPhi see [131]. Naturally, the PBE must be satisfied on each grid point. By calculating the electrostatic potential on every grid point, DelPhi can also derive physical values such as the electric field, the reaction field energy, and the surface charge at the boundaries between the solute and the solvent.

In DelPhi, the total electrostatic energy is partitioned into the following terms: Coulombic, Reaction Field, Self-Reaction Field, Ionic, Osmotic Pressure, and Electrostatic Stress [131]. Naturally, the last three terms only appear when the ionic strength is non-zero. Finally, the Osmotic Pressure and Electrostatic Stress terms cancel out in the linear PB equation. The Coulombic Interaction Energy is calculated analytically using Coulomb's law and its definition is rather intuitive: it is defined as the energy required to bring charges from infinite distance to their resting positions within the dielectric specified for the molecule. The Reaction Field Energy, or Solvation Energy, is calculated as follows: using Gauss's theorem, the polarisation charge on each volume at the boundary of the solute is calculated; then, the interaction energy of these polarisation charges with the charges in the solute is calculated. Usually in this sort of finite volume scheme the calculated charge is then mapped to the centre of the volume; in DelPhi however the charges are projected on the Connolly surface. Interfacing DelPhi with NanoShaper provides the analytical Connolly surface, and therefore an improvement in accuracy without requiring very small grid spacing, and therefore enhancing performance. The Reaction Field Energy is therefore defined as the product of the potential due to induced surface charges and the values of all fixed charges of the solute. We note that, depending on the physical process that is being studied, this term may coincide with the actual solvation energy, but in general the latter is obtained by taking the difference in reaction field energies between suitable reference states.

For systems with several dielectric regions, the reference medium is vacuum, with a relative dielectric value of $\epsilon_v = 1$. In such systems, we need to take into account the transfer free energy of a charge from vacuum to a medium with relative dielectric constant $\epsilon_r$. This free energy corresponds to the interaction between the charge and the volumetric polarisation charge of the medium. This contribution is the Self-Reaction Energy. This energy is always negative because transferring a charge from vacuum to another medium is an energetically favoured process. Regarding the ionic contribution to the interaction energy, it can be calculated in two ways using DelPhi: either performing an integration on the entire volume of the solvent where the potential is non-zero to calculate the interactions between the ions, or by two runs on the same system with and without ions in the solvent, and a subsequent calculation of the ionic component of the interaction energy through the difference in the grid energy contribution in the two cases.

## 2.2 Electrostatic Interactions in the Chromatin Fibre

At large scales in the chromatin fibre, structures are approximately electrostatically neutral, allowing for an average treatment of electrostatics and solvation in polymer models for chromatin. However, at the NCP and oligonucleosome scale, electrostatics and solvation become extremely important, due to the high charge of the DNA. The charges present on the DNA backbone are partly neutralised by the winding of DNA around the histone core, especially through the effect of the histone tails, and partly through counter-ions present in the nuclear environment. The modelling of internucleosomal interactions using reductionist analytical potentials, which omit the explicit role of histone tails, can cause secondary, but still relevant, electrostatic effects to be overlooked 2.2.

Considering the biological importance of different ionic types, $Mg^{2+}$ is particularly significant, as it has been found to promote nucleosome condensation and aggregation and could promote linker DNA bending, because in its presence interactions of first and third neighbouring nucleosomes are boosted [54]. Tetravalent cations on the other hand require lower concentrations to induce compaction [136]. In [35], systems of 1-10 nucleosome core particles (NCPs) were studied using a coarse-grained model in order to study the effects of monovalent, divalent, and trivalent cations on these structures, reproducing experimental data. It was observed that an increase in $K^{+}$ ions amplified the repulsive internucleosomal electrostatic interaction; increasing $Mg^{2+}$ concentration caused partial aggregation, and an increase in $COHex^{3+}$ ions triggered a strong mutual internucleosomal attraction in 10 NCP systems, therefore showing that the aggregation of NCPs is different under the effect of different types and concentrations of counterions.

Multivalent ions and the effect of their distribution around NCPs on chromatin conformation were also studied in [106], using a mean-field PBE approach, with an emphasis on shielding charges, which aggregate particularly around DNA and the exposed parts of the histone tails. The fact that a surface needs to be exposed to solvent in order for ions to bind on it makes ion-caused electrostatic screening (a change in the effective electric charge) and ion-chromatin interactions in general directly dependent on compaction. Calculations showed that the enhanced screening due to divalent ions might not only be because of their higher charge, but also because they form a denser layer of counterions around the NCP and fluctuations in this layer are correlated to different fibre conformations. This makes even more evident the fact that the topology of compaction is a key determinant for chromatin-ion interaction. It was observed in these simulations that the shielding charge arising from both monovalent and divalent ions was linearly correlated with the ionic strength of the solution.

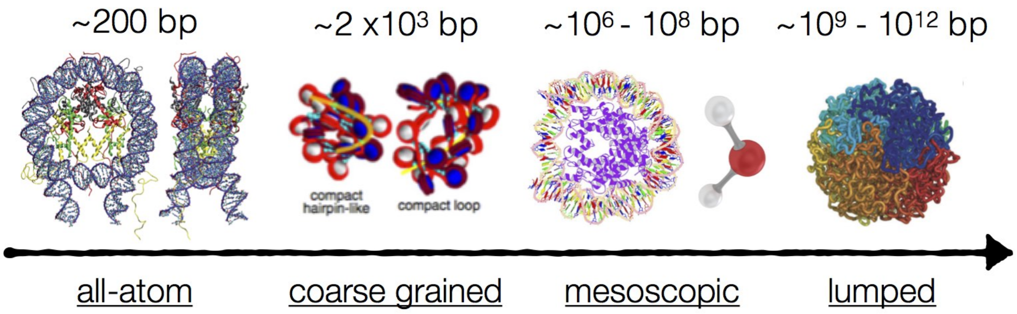


Figure 2.2: Depending on the order of magnitude in terms of bp that is of interest, electrostatic interactions require different modelling approaches, which in turn provide different kinds of data. In the order of magnitude on nucleosomes, approximately 200 bp, we have structures of atomic resolution available (nucleosome crystal structure [116]). In fact, the largest crystal structure available is that of the tetra-nucleosome (PDB code 1ZBB [137]). Electrostatic studies using full-atom structures can provide information on local properties, but cannot extrapolate on properties of the larger chromatin structure. In the kbp order of magnitude, coarse grained models are used, where nucleosomes and linker DNA are represented as beads and flexible rods (figure adapted from [51]). These models provide local informations, for short fibres containing tens of nucleosomes. In cases in which one wishes to study properties of the chromatin fibre in the Mbp scale without losing the level of detail that information from atomistic models provides, so-called mesoscopic models can be used, where nucleosomes and linker DNA are parametrised in an ultra-coarse-grained regime. Finally, when properties of the entire fibre are of interest, such as chromosomes or the entire genome even (Gbp - Tbp order of magnitude) lumped models are used, where chromatin is represented as a continuous fibre and therefore the notion of the single nucleosome is lost. The study of electrostatic interactions in these models is implemented using analytical potentials and notions from polymer physics to study the global properties of the fibre (figure adapted from [47]).

In the study of structures as large and complex as chromatin, it has been proposed in [37] that implicit solvent Generalised Born (GB) simulations would be preferable to traditional fully explicit MD, in order to circumvent computational limitations. However, standard GB scales poorly with the number of solute atoms and, in this work, a multiscale atomistic GB model that incorporates improvements in the electrostatic calculations is presented, the accuracy of which was evaluated through point-by-point comparison with PBE calculations. Taking advantage of the natural hierarchical organisation and charge distribution of chromatin, Izadi et al. used approximate point charges to calculate electrostatic interactions between distant points in a 40-nucleosome structure, containing approximately 1 million atoms, focusing particularly on the behaviour of the histone tails. They were able to reproduce experimental findings of the interaction of the H3 histone tail and the linker DNA. The GB approach proved the existence of viable alternatives that drastically reduce the cost of conformational sampling in very large structures.

One could not conclude a discourse on chromatin electrostatics without mentioning the effect of the histone tails, which have been found to promote stability of the linker histone on the NCP. In some models, histone tails are modelled as a series of beads with one positive charge per bead [35, 106, 138]. It was seen by [20] that certain histone tail configurations promote DNA bulging at entry and exit sites 1.2, possibly contributing to the formation of twist defects in the nucleosomal DNA. Twist defects are DNA deformations that allow for one more or less DNA bp in positions where DNA interacts closely with histones [24]. They are important, among other reasons, because their presence causes the formation of nucleosomes with 146 bp instead of the usual 147 [21], due to overwinding and stretching of the DNA [90]. They also speculated that the presence of arginines and lysines might impose constraints on histone tail motion because of attractive electrostatic interactions. Contacts between DNA and histones were seen to be dominated by the histone tails, making up 60% of protein-DNA interactions in the nucleosome, rapidly wrapping around the DNA (in [20], it was observed that they do so in the first 20ns of the simulation).

In another study, the N-terminal of the H4 histone tail was observed to interact with the “acidic patch” present on the surface of adjacent nucleosomes, a small groove formed by eight residues, six belonging to H2A and the remaining to H2B, which constitutes a region of highly negative charge density on the nucleosome surface, serving as a hot-spot for DNA-binding proteins and histone tails [15, 17, 139]. Throughout 1 µs-long MD simulations in [20], the NCP is seen to be very stable in dynamics, in contrast to histone tails and linker DNA: large scale unwrapping or opening of NCP DNA were not observed, even when simulations were performed in 1M salt concentration, under which conditions they are known to occur [140]. This indicates that such phenomena might take place on longer time scales. Of particular interest are the histone H3 tails, which have been suggested by experiments [31] to form stable folded structures, and even to potentially compete with other DNA-binding proteins, affecting accessibility of epigenetically modified sites in the minor grooves.

It has already been mentioned that the presence of A-tracts can change the curvature of DNA, causing the minor grooves to be narrower than those in segments with lower curvature, and locally enhancing negative electrostatic potentials. In [110], PBE calculations were performed on DNA, showing that the electrostatic potential caused by the DNA backbone had intensity peaks inside the major and minor grooves. The position of these peaks correlates with the positions of arginine residues on the histone core. Previously observed binding preference for arginines over lysines in minor grooves, and especially in narrower ones, was partly explained via a combination of electrostatic and desolvation effects. For the study of minor groove geometry, all the crystal structures of protein-DNA complexes containing at least one base atom - aminoacid contact were analysed. Analysis of nucleosomal DNA was based on the nucleosome structures available on the Protein Data Bank (PDB) at the time.

## 2.3 "The nucleosome as a sponge": Solvation in Chromatin

The role of the solvent in biomolecular interactions is known to be crucial. In part, this is because of solvent-mediated electrostatic effects - the screening of the water molecules and that of the ions in solution. In addition, there is the so-called cavity formation phenomenon, which penalises the occurrence of solvent-excluded regions. Chromatin spatial arrangement, due to NCP charge, size and porosity, is expected to be particularly affected by these phenomena, which must be accurately considered. It has already been described that the formation of the fundamental unit of chromatin, the nucleosome, occurs by the complexation of the negatively-charged DNA polymer with the positively-charged histone protein octamer. If investigated at the molecular level, this process is governed by a number of interactions such as hydrogen-bonds, salt-bridges, and water-mediated interactions occurring along the positively-charged arginine anchors that intercalate deep inside the minor grooves of DNA facing the histone core [17, 141]. When it comes to histone core-DNA electrostatic interactions, it is known that every nucleosome presents 14 non-covalent histone-DNA contacts, at the sites of arginine residues [23].

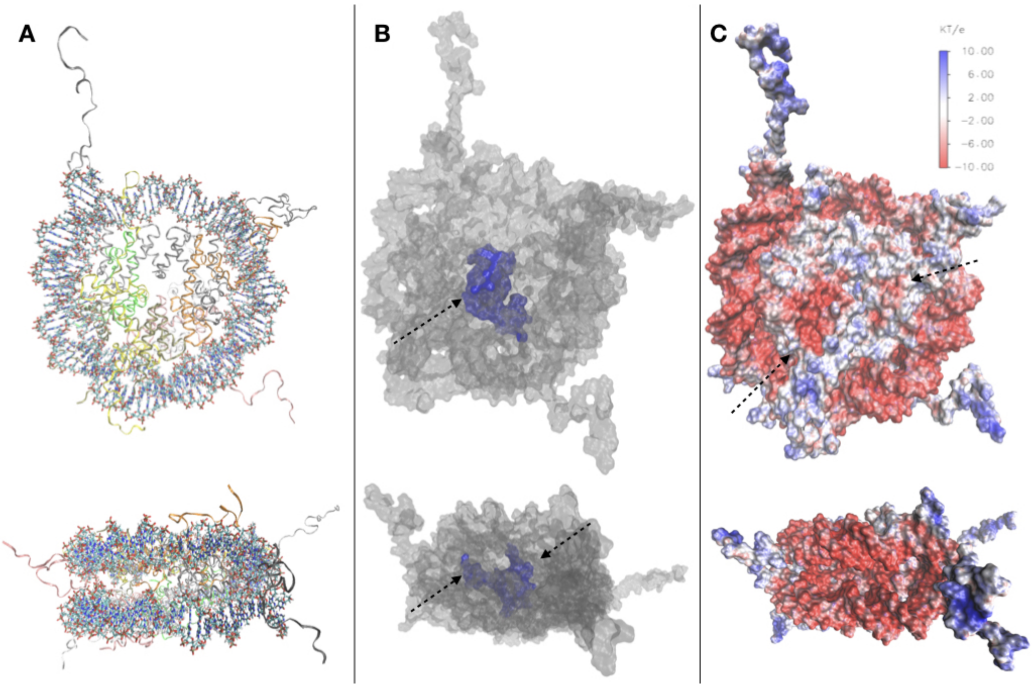


Figure 2.3: **A.** Top and side view of the 1KX5 crystal structure. **B.** Top and side view of the SES of 1KX5, constructed with NanoShaper [135] and visualized via VMD. The channel traversing the histone core is represented in blue together with an adjacent open cavity and is indicated by an arrow. On the side view, the entrance and exit of the channel can be seen, indicated by arrows. **C.** Electrostatic map of the SES of 1KX5. Areas of negative surface potential are indicated in red and areas of positive surface potential in blue. The acidic patch is indicated by an arrow on the histone core. Another minor acidic region, composed by fewer residues on the surface of histone H4, is also highlighted. Most of the exposed regions of the histone core are electrically neutral, with the acidic patch representing the main exception. Remaining positive charges of the histone core are buried, due to the binding of encircling DNA. We also note positive charges on the histone tails, and strong negative charges on the DNA backbone. For more information on the 1KX5 nucleosome structure see Table 3.3.

Solvent exposure affects electrostatic interactions at the nucleosome level: compared to H3 and H4 histones, the two H2 variants are more solvent exposed, making them more accessible to chromatin-binding proteins as well [37]. Specific ion binding sites and their location on the nucleosome are also of particular interest, and they can be studied using electron density maps in combination with chemical information [90]. It has been observed that sodium preferentially condenses around regions rich in solvent accessible acidic residues, especially in areas with two or more acidic residues in close proximity [119]. It is also speculated that, in chromatin fibres exhibiting high compaction, internucleosomal electrostatic repulsion could be reduced in intensity because of an increased neutralisation of the DNA backbone charge by the neighbouring histone cores and counterion screening. The effect of the solvent and of ionic interactions on DNA arrays has been studied extensively by Podgornik, Rau, and Parsegian [142, 143]. They showed that the repulsive force between parallel DNA double helices is sensitive to ionic species and ionic strength, and the dependence of this force on distance was studied as well. In a more recent work, Zavadlav, Podgornik, and Praprotnik [144] performed multiscale MD simulations to study the effect of varying osmotic stress on densely packed DNA arrays, identifying the hydration force as the primary interaction mechanism at high DNA densities.

The idea that the nucleosome is an impermeable object has been proven erroneous [119]; in this work, it was seen that mobile ions are able to reach the NCP inner core because of high levels of local solvation (more than 1000 water molecules). This led to the conclusion that the local value of dielectric constant in the region facing the histone core is larger than expected. The authors also looked into the mobility of water molecules on the first hydration layer of the nucleosome and, as expected, found them to be less mobile than bulk water molecules. Through detailed visualisation of structured water at the protein-DNA interface, they also found that water molecules not only contribute significantly to the stability of DNA binding but also adapt histone surfaces to conformational variations of DNA, facilitating nucleosome dynamics. All-atom electrostatics calculations were conducted and compared to PBE calculations, observing a slight inconsistency between the two. PBE predicts that the most significant contribution to DNA charge neutralisation comes from the enhancement of the electric field and that it is a result of the tight wrapping of the DNA around the histone core. These results indicate that close condensation of ions around the nucleosome can significantly reduce the short range effect of the nucleosomal charge, having as a natural consequence the facilitation of chromatin close packing. For an in-depth study of the electrostatic constant of DNA, we point the reader to [145].

In another work concerning NCP solvation [90], the solvent-accessible surface area (SASA) of nucleosome crystals with 147 bp and 146 bp was investigated. NCPs with 147 bp were found to possess a SASA of approximately 74 $\text{Å}^2$, which is distributed mostly in the cavities within the histone octamer and in the space between it and the DNA. The primary hydration layer of the NCP was found to contain slightly more than 2000 water molecules, the positions of which were found to largely correspond to the positions of A-tracts, especially in the vicinity of the minor groove. Water was shown to be important in the two main mechanisms of protein-DNA recognition: direct readout (nucleotide chemically specific bonds) and indirect readout (sequence-dependent conformational features of DNA recognised by sterically complementary protein contacts). Structures termed "spines of hydration" were also observed, in which water molecules bind regularly to adenine N3 and thymine O2 atoms [146]. Structural analyses have shown that the phosphate groups are the most strongly solvated components of the DNA [147, 148].

### 2.3.1 Qualitative Study of Nucleosome Porosity

In order to illustrate the porosity of the nucleosome, particularly described in [119], we have conducted a study on the nucleosome crystal structure (PDB code 1KX5 [90], Fig.2.3.A) using NanoShaper interfaced with VMD [135, 149], providing the values of the Surface to Volume Ratio (SVR), the number of cavities and pockets. We measure an SVR of $0.387\ \text{Å}^{-1}$, which reflects a quite high porosity [150], and a number of cavities and pockets. In Fig.2.3.C, we visualise the channel traversing the nucleosome core, which significantly impacts on NCP accessibility to water and ions. Our results are consistent with previous qualitative analyses mentioned in this section, and indeed indicate that the nucleosome is highly solvated and porous. We have also constructed an electrostatic map of the nucleosome, using data from the DelPhi Poisson Boltzmann solver [131] on the potential and constructing the SASA of the nucleosome with NanoShaper, as seen in Fig.2.3.B, where it is possible to clearly see, among other features, the position of the acidic patch (residues E56, E61, E64, D90, E91, E92 of H2A and E102, E110 on histone H2B [15], and the highly charged histone tails, both key elements in chromatin compaction and chromatin interaction with DNA-binding proteins. This analysis showed a minor acidic region, on the surface of histone H4. Further quantitative information regarding solvation in nucleosomes can be found in Chapter 4.

## CHAPTER 3

# The Role of Histone Tails in Nucleosome Stability: an Electrostatic Perspective

The aim of this Chapter is to provide information on the effect of the histone tails by studying the electrostatic interactions occurring in the nucleosome core particle, and to propose a methodology to connect electrostatic calculations to the structural and functional features of protein-DNA systems. To that end, we connect observations made through the analysis of electrostatic calculations to structural features of the nucleosome. We infer the electrostatic forces acting on the phosphate atoms of the DNA backbone by calculation of the electric field on the former.

We investigate the electrostatic origins of effects such as different stages in DNA unwrapping, nucleosome destabilisation upon histone tail truncation, and the role of specific arginines and lysines undergoing Post-Translational Modifications. We find that the positioning of the histone tails can oppose the attractive pull of the histone core, locally deform the DNA, and tune DNA unwrapping. Small conformational variations in the often overlooked H2A C-terminal tails had significant electrostatic repercussions near the DNA entry and exit sites. The H2A N-terminal tail exerts attractive electrostatic forces towards the histone core in positions where Polymerase II halts its progress.

We therefore propose a methodology for the study of protein-DNA electrostatic interactions and apply it to clarify the effect of histone tails in nucleosomes. This method can be used to correlate electrostatic interactions to structural and functional features of protein-DNA systems, and can be combined with coarse-grained representations. We validate our results with comparisons to previous experimental and computational observations. We conclude that the presence of the histone tails leads to non-trivial electrostatic effects of great significance for NCP stability and DNA unwrapping.

The work presented in this Chapter has been published in [151].

## 3.1 Introduction

In general, protein-DNA interactions are dominated by three types of phenomena: electrostatic interactions, hydrogen bonds, and hydrophobic forces. Direct electrostatic interactions and solvation are of particular importance in nucleosomes, because of the elevated charge of the interacting entities. Solvation effects and electrostatic interactions are described by the PBE, which uses a continuum mean-field description of the solvent, including dissociated salts, while solvated molecules are considered as continuous regions of lower dielectric value [131]. PBE theory has been used in both coarse-grained (CG) [36, 106] and full atom approaches [37, 119], directly studying data acquired by solvers such as DelPhi [131] and APBS [132] or indirectly at a model-building level [152].

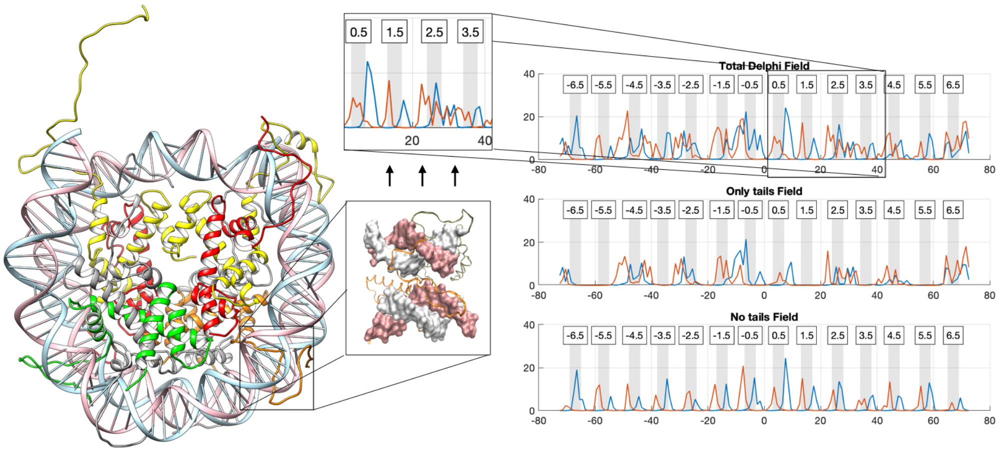


Figure 3.1: Graphical Representation of our methodology connecting structural and electrostatic features of protein-DNA systems, presented in this Chapter. We calculate the values of the electric field and the force exercised on the positions of the phosphate atoms on the nucleosomal DNA backbone, and study the effects of that the structural changes of the histone tails have on nucleosome stability. On the right, the electric field exercised on the phosphates of the DNA backbone of each chain of nucleosomal DNA is presented, indicated in different colours. We have measured the electric field on three types of structures: structures in which the contribution of both the histone tails and the histone core was taken into account, structures in which only the histone tails where taken into account, and structures where the histone tails were truncated so as to study the electrostatic interactions solely between the histone core and the nucleosomal DNA. The values of the electric field are correlated to the number of histone tail atoms in proximity of the DNA backbone.

Studies elucidating the separate effects of histone core and histone tails, and on the sensitivity to the conformational changes of the latter, can prove very useful to construct improved models and to help explain observed phenomena that might otherwise be puzzling. Interestingly, the forces acting on each phosphate group, which are mainly of electrostatic origin and have been derived from DelPhi calculations, can be used in combination with the sequence–dependent cgDNA+ [153] model, which includes the mechanical forces acting on the DNA backbone and explicitly treats each base and each phosphate group. It is possible to minimise the CG energy predicted by cgDNA+ with external constraining forces applied to each phosphate group. This approach would allow the exploration of the CG energy landscape as a function of the DNA sequence with given fixed forces separating them in those caused by the histone core ("no tails" structure), and by the histone tails ("only tails" structure).

In this Chapter, we provide results that we obtained in our study of the electric field, electrostatic potential, axial and radial forces acting on each bp by performing PBE calculations on full atom structures extracted from the trajectory provided by Shaytan et al. [137] In order to investigate the effect of the histone tails on said interactions, we analyse and compare our results on three sets of structures: structures in which the histone tails are included in the NCP, structures in which the histone tails have been truncated (referred to as "no tails" structures), allowing us to observe the electrostatic effects exclusively due to interactions with the histone core, and structures in which all charges except those of the histone tail residues have been put to zero (referred to as "only tails" structures), in order for the pure histone tail contributions to the electric field to emerge. We begin our analysis by discussing the effect of the histone tails on the electrostatic potential in correspondence of DNA, along with a more general structural characterisation of the NCP in terms of electrostatic interactions and DNA-histone contacts. We then move on to a study of the radial and axial components of the electrostatic force acting on the phosphates and their implications for the NCP structure, focusing on DNA unwrapping, the role of histone tail truncation, and the particular importance of the positively charged arginine and lysine residues.

The histone tails and their structural role have been studied to great extent in experimental works using a variety of methods, such as X-ray crystallography [89] to SAXS [102, 154], FRET [59], and FCS. Their mobile and intrinsically disordered nature makes studies using X-ray crystallography particularly challenging [31, 100, 101], even though it provides higher resolution data than fluorescence and scattering techniques, which naturally do not allow for atomic level resolution. SAXS is often used in conjunction to other techniques, such as FRET and FCS to correlate different kinds of data [35, 103, 104]. However, structures in which the histone tails are completely lacking or badly resolved are often used.

Becker and Everaers, in their study of DNA nanomechanics, have developed a methodology to infer external forces and torques applied to a protein-DNA complex, in cases where the static shape of a structure is provided. This procedure is made possible by the fact that the elastic and mechanical properties of DNA are known [22]. Using a bp-level CG representation of the DNA, they were able to infer the values of the forces and torques acting on each individual bp. The amount of deformation that the DNA is subject to when in complex with proteins is deduced from comparison with free DNA segments, and free energy landscapes with respect to different degrees of freedom were used to quantify and describe this deformation. The starting points for their calculations were mean CG conformations of full-atom structures found in the Protein Data Bank. Their analysis allowed for the discrimination between force transmitting and non-transmitting protein-DNA contacts. A CG approach is useful in this case due to the large number of degrees of freedom involved. However, the analysis conducted by Becker and Everaers is indirect, inferring the forces, as we mentioned above, from the deformation of the DNA. Protein-DNA interactions are not explicitly taken into account. This methodology was subsequently applied to the nucleosome [155], including studies of particularly interesting features such as the twist defect, or the torsion of linker DNA connecting two adjacent NCPs.

## 3.2 Computational Methods

### 3.2.1 Input structure selection and generation

The original PDB files and MD trajectory were taken from publicly available data [137] from the work of Shaytan et al. [20] The structure (`FN model` in [137]) is composed of the PDB code 1KX5 crystal structure [90] and two 20bp-long B-DNA segments of linker DNA. In order to identify the most stable histone tail conformations visited during the MD trajectory, we customised a clustering protocol, a variant of the well-established k-means algorithm, using the BiKi Life Sciences software [149]. For more information on the 1KX5 nucleosome structure and other nucleosome crystal structures used in this work see Table 3.3.

Consistent with the fact that a linear combination of the atomic coordinates of two conformations does not normally correspond to a viable conformation, in this clustering algorithm no underlying vector space is assumed, and only the pairwise distance, i.e. the RMSD of the heavy atoms of the tails, was used. This has the further advantages that no reference structure is needed for the calculation and that the representative structure of each cluster, the medoid, is by construction a conformation observed during the MD. It is worth noting that the reliability of these structures in terms of being representative of equilibrium states depends on the overall exhaustiveness of the sampling performed in the trajectory. Shaytan et al. specify that their simulation explored conformational ensemble starting from a local quasi-equilibrium state.

We selected four medoids representing the largest clusters, in order to have the most diverse and representative structures in terms of the conformations assumed by the histone tails. The medoids are named medoid13, medoid15, medoid16, and medoid17, indicating the cluster that they represent. Cluster labelling does not reflect the cardinality of the clusters. The number of structures in each cluster (population) and the percentage with respect to the total number of structures are provided in Appendix A (Fig 2). We rely on the 1μs-long trajectory from Shaytan et al. [20]. It is highly likely that the this trajectory is not ergodic, given the size of the system, impacting on the statistical representativity of the medoids.

We then generated PQR files, i.e. PDB format files where atomic charge and radius replace beta-factor and occupancy, from the medoid PDB files via the PDB2PQR web server [156]. PQR files were used as input files for the DelPhi PB solver, which requires atomic radii and charges. We used the CHARMM force field setting for the PDB2PQR conversion (PDB2PQR uses CHARMM27), as it was the one used by Shaytan et al. in the MD simulations. After file conversion with PDB2PQR, we had to perform some manual changes to the DNA charge parameters due to nomenclature issues causing only partial recognition of the DNA structure. While the linker DNA was present in the original trajectory, and therefore in the medoids, it was truncated in order to conduct electrostatic analysis only on the NCP.

### 3.2.2 Electrostatic field and potential calculations

We calculated the values of the electrostatic potential and the electrostatic field on the positions of the phosphate atoms belonging to the DNA backbone taking into account only protein charges in three distinct cases: intact structures, in which both the charges of the histone core residues and of the histone tail residues are taken into account, structures in which the histone tails have been truncated, and therefore only the charges of the histone core are taken into account, and finally structures in which only the charges of the histone tails are present. The charges of the DNA were switched to zero in all our calculations, so as to study the electrostatic contributions exclusively due to histone-DNA interactions. We denote with SHL the Superhelical Locations on the DNA. Electrostatic energies were computed using the DelPhi PBE solver, interfaced with NanoShaper [135]. The full non-linear PBE was used for the calculations (DelPhi parameter ni = 12000). The solute was assigned a dielectric value of 2, and the solvent a value of 80. Salt concentration was set at 145mM. Atomic partial charges and radii were taken from the PQR files. Other parameters assumed default values.

The results obtained in this way are semi-quantitative. Indeed, the high charge on the NCP requires the adoption of the full non-linear PBE formalism and this would not, in principle, allow to calculate the effect due to one part of a system by just switching off the charges of the remainder. In order to assess the significance of this approximation, we followed the conceptual partitioning described in [131] and used the ionic potential at phosphorous sites in DNA as an estimate of the intensity of the entire non-linear effect, which also includes osmotic pressure and electrostatic stress terms. Ionic potential was calculated by subtracting the grid potential at the phosphorous' sites obtained with two PBE runs, one with physiological and the other with null ionic strength.

According to the mentioned partitioning, in a fixed configuration the local electrostatic potential is the sum of the coulombic, reaction field (arising from the difference in polarisability between the protein and the solvent) and ionic contributions. Non-linear PBE calculations of the fully charged system show that the reaction field contribution is practically unaffected (absolute relative error $< 0.1\%$) by the introduction of the non-linearity and that the ionic potential amounts to the $2\%$ of the former term, which, in turn, is of the same order of magnitude of the coulombic contribution generated by the histones' charges alone. This supports the correctness of the reported trends and patterns of electric field and potential and effects thereof.

### 3.2.3 DNA-Histone tail contact calculation

From the PDB files of each medoid we extracted the cartesian coordinates of all histone tail atoms and the cartesian coordinates of phosphate atoms. The parsing of the PDB files was done using the class PDB of Biopython [157]. We generated DNA-Histone tail contact tables by selecting for each phosphate atom the histone tail atoms that are within a cut-off distance in the range of 10 to 25Å. For each medoid we have computed the total number of histone tail atoms that come into contact with each phosphate as a function of the cut-off distance. Moreover, using the CHARMM36 force field parameters we computed the electrostatic field generated by a phosphate and the neighbouring histone tail atoms as function of the cut-off. In this work, we do not consider values of the cutoff higher than 25Å, because we observed that, after this threshold, the contribution of additional contacts with histone tail atoms to the electric field was not significant. The naming scheme of the chains present in the NCP and the residues belonging to each histone tail is summarised in Table 3.1. H-bond formation between each histone tail and the DNA in medoid13 were analysed using the dedicated tool of VMD software (Appendix A Figs. 80-89).

| Molecule | Chain ID | Residues | Histone Tails (res. #) |
|---|---|---|---|
| Histone H3 (H3') | A (E) | 1-135 | N-terminal (1-44) |
| Histone H4 (H4') | B (F) | 1-102 | N-terminal (1-24) |
| Histone H2A (H2A') | C (G) | 1-128 | N-terminal (1-17)<br>C-terminal (99-128) |
| Histone H2B (H2B') | D (H) | 1-122 | N-terminal (1-34) |
| DNA | I | -73 - 73 | |
| DNA | J | -73 - 73 | |

Table 3.1: Naming scheme used in this work for chains in medoids, and residues of each histone tail.

### 3.2.4 Axial and radial electric field and force component calculation

For each medoid we computed the average, i.e. the geometric centre, and the covariance of the phosphate cartesian coordinates. We then defined the axial direction as the eigenvector corresponding to the smallest eigenvalue of the covariance matrix, due to the shape of the system. The axial direction is computed once for the full medoid. First, we computed each phosphate atom position with respect to the centre and we projected the relative vectors on the axial direction. Finally, we converted the electric field [kT/Å/e] into the force acting on phosphates [kT/Å] by multiplying the DelPhi vector field by -e. In order to observe the trend of the spatial distribution in the electrostatic force, we considered the axial and radial components of the unitary vector corresponding to the electric field on each phosphate as reported by DelPhi. The resulting plots for the normalised and non-normalised radial and axial components of the electrostatic force are found in Appendix A (Figs.12-23, Figs.36-71).

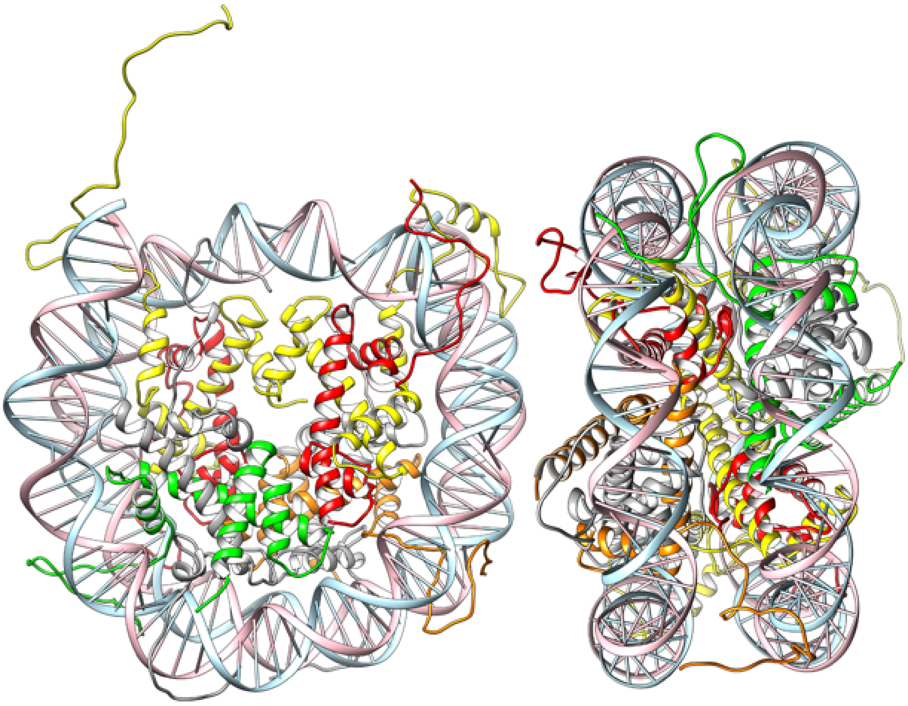


Figure 3.2: Front and side view of the nucleosome. The histones and DNA are coloured by chain: histones H4/H4' in red, histones H3/H3' in yellow, histones H2A/H2A' in silver, and histone H2B in orange and H2B in green.

## 3.3 Structural Characterisation of Medoids

The structures we selected to conduct our analysis of electrostatic interactions between histone tails and DNA are the representatives of the four largest clusters we obtained after performing a k-medoids clustering on a full-atom MD trajectory of the NCP. Further details on this process are found in Subsection 3.2.1. To quantify the difference between the histone tail conformations in each medoid, we used VMD to calculate the Root Mean Square Deviation (RMSD) of the residues belonging to the histone tails in each chain for all medoids, taking medoid13 (Fig. 3.2), that is the representative of the most populated cluster, as reference. The results of this comparison are seen in Table 3.2. The most diverse conformations are those assumed by the H3', the H2B, and the H2B' tails while, at the opposite end of the spectrum, the H2A' C-terminal tail presents similar conformations in medoids 15, 16, and 17. Medoids 15, 16, and 17 are depicted in Appendix A (Fig. 1). Taking into account the electrostatic and mechanical effects that the plurality of interactions between the DNA and the histone tails produce, the use of structures obtained from the analysis of an MD trajectory emerges as almost a necessity. The use of the crystallographic structure in our analysis would overlook the adsorption of the histone tails on the DNA, their dynamics, and the changes they cause on the local curvature of the double helix. A variety of twist-defect nucleosome variants have been experimentally observed to exist in solution [158], while only a few have been captured in crystal structures, such as 3AFA (see Table 3.3) [91].

| Histone Tail (res. #) | medoid15 | medoid16 | medoid17 |
|---|---|---|---|
| H3 N-terminal (1-44) | 7.743 | 8.816 | 4.885 |
| H3' N-terminal (1-44) | **11.014** | **18.034** | 13.593 |
| H4 N-terminal (1-24) | 5.163 | 6.576 | 6.038 |
| H4' N-terminal (1-24) | 8.728 | 8.266 | 12.033 |
| H2A N-terminal (1-17) | 5.089 | 6.555 | 4.548 |
| H2A C-terminal (99-128) | 3.556 | 5.517 | 2.376 |
| H2A' N-terminal (1-17) | 3.765 | 7.439 | 3.102 |
| H2A' C-terminal (99-128) | 4.833 | 4.437 | 4.489 |
| H2B N-terminal (1-34) | 5.489 | 9.022 | **18.974** |
| H2B' N-terminal (1-34) | 7.790 | 11.495 | 2.551 |

Table 3.2: Root Mean Square Deviation (RMSD) of histone tail residues among different medoids, compared to their positions in medoid 13. These values provide a quantitative indicator of the difference in histone tail conformations. Maximum RMSD for each medoid are indicated in bold.

In order to associate the conformational changes of the histone tails to electrostatic effects in NCPs, we calculated the number of histone tail atoms that come into contact with the phosphate atoms in the DNA backbone, within a range of distances from 10 to 25Å. The resulting 2D plots are provided in Appendix A (Figs.8-11). We observed hotspots of phosphate/histone tail interactions in the DNA entry and exit points, particularly at positive SHL. Numerous DNA-histone tail interactions in this region are known to be very important for DNA unwrapping, as we will discuss further in our analysis. Furthermore, there are regions of frequent DNA-histone interactions in chain J between bp 27 and 57, in which DNA interacts with both copies of H2A and H2B histones, as well as H4'. This is a key area for transcription, where Polymerase II has been observed to pause its progress [159]. We also observe marked differences in chain J between medoids 16 and 17: in the former, the histone tails are overall farther away, and are only closer to the DNA in the region of bp 33-35, where the DNA interacts with the H2B' tail. Overall, in medoid16 we have more localised tail-DNA interactions. Finally, in medoid17 we observe overall less contacts in negative SHL compared to other medoids.

The dynamics of the histone tails influence DNA geometry, causing, for example, DNA bulging in the entry/exit sites, and twist defects. In addition, histone tails form the majority of protein-DNA contacts in the NCP entry/exit sites. In terms of contacts, the histone tails make up 60% of all histone-DNA contacts. During the 1µs simulation, up to 90% of histone tail amino acids formed direct or water-mediated contacts with the DNA, with no particular clustering of contacts toward the beginning or end of the tails. Arginine and lysine residues are found inserted in the DNA minor grooves, particularly the following residues: Arg8 and Arg26 of histone H3; Lys16 and Arg17 of histone H4; Arg11, Lys13 and Lys126 of histone H2A; and Arg29 and Arg30 of histone H2B.

It is worth noting that, in the initial crystal structure, no interactions with the DNA were observed for the H3 and the H4 tails, and protein-DNA interactions through lysine residues were severely underestimated, indicating the pitfalls of only using a single structure in NCP analysis. In Shaytan et al, the histone tails exhibited large fluctuations, with an RMSD of more than 6Å, and were rearranged in the minor grooves, rapidly adsorbing on the DNA. The dynamics of the histone tails also depend on whether they adsorb on linker DNA or NCP DNA, the latter being much less flexible. The rapid adsorption of the histone tails on the nucleosomal DNA is also shown by the creation of H-bonds along the MD trajectory. The relevant plots are provided in Appendix A (Figs.80-89). The tail forming the most H-bonds is H3 (chain A), from initial 7 bonds stabilising to a mean of 20 bonds. The smallest number of bonds was found in the N-terminal tail of H2A (chain C) and H3', which mostly interacts with linker DNA (less than 6 bonds). Finally, the tail of H2B (chain D) was already forming H-bonds with the DNA in the original structure, and their number oscillated around this value.

## 3.4 Electrostatic Interactions between Nucleosomal DNA and Histone Tails

As a first marker of electrostatic interactions, we calculate the electrostatic potential (plots found in Appendix A Figs.4-7) and the electric field exerted by the protein system on the phosphates, shown in Figs. 3.3-3.6. As expected, in the structures where the histone tails have been truncated ("no tails") the coulombic potential follows the periodicity of the DNA double helix: it is higher in the regions where the DNA backbone faces towards the histone core, and lower when it is facing away, presenting qualitative similarities to the trend of the forces described by Becker and Everaers [155]. We remind that, in our simulations, the charge of the atoms belonging to the DNA residues has been "switched off", indicating that the observed pattern is due to the histone core and the solvent.

The variations we observe in the “no tails” potential across medoids are attributable to the slight variations in DNA bp positioning along the MD trajectory, and to differences in core-DNA distances. We observe the same behaviour in the electric field. When it comes to the contributions to the potential solely due to the histone tail residues, we observe a non-trivial dependence on their positioning. This behaviour "breaks the symmetry" presented in CG models that only consider the deformation of the DNA and interactions with the histone core, and points to more complex interactions taking place in nucleosomal DNA. Looking at the values of the electric field for the “only tails” structures, we observe an obvious correlation between the vicinity of histone tail and phosphate atoms and an increase in magnitude in the electric field (Figs. 3.3-3.6 "only tails field").

Let us look at some examples illustrating how conformational changes on the histone tails reflect in changes in the electric field: we observe four regions of the DNA in which the electric field contribution of the histone tails is zero, indicating a lack of interaction with the histone tails in any of the four medoids: from bp -70 to -53, from -40 to -32 bp (in medoid16 the interactions on chain J are shifted away from the dyad by a couple of bp), bp 10 to 20 and 50 to 63. Conducting a proximity analysis between the phosphates of each DNA chain and the atoms of the histone tails, the results of which can be seen in Figs.8-11 in Appendix A, we observe that in these regions there are very few (less than 100) or no atoms in a range up to 25Å.

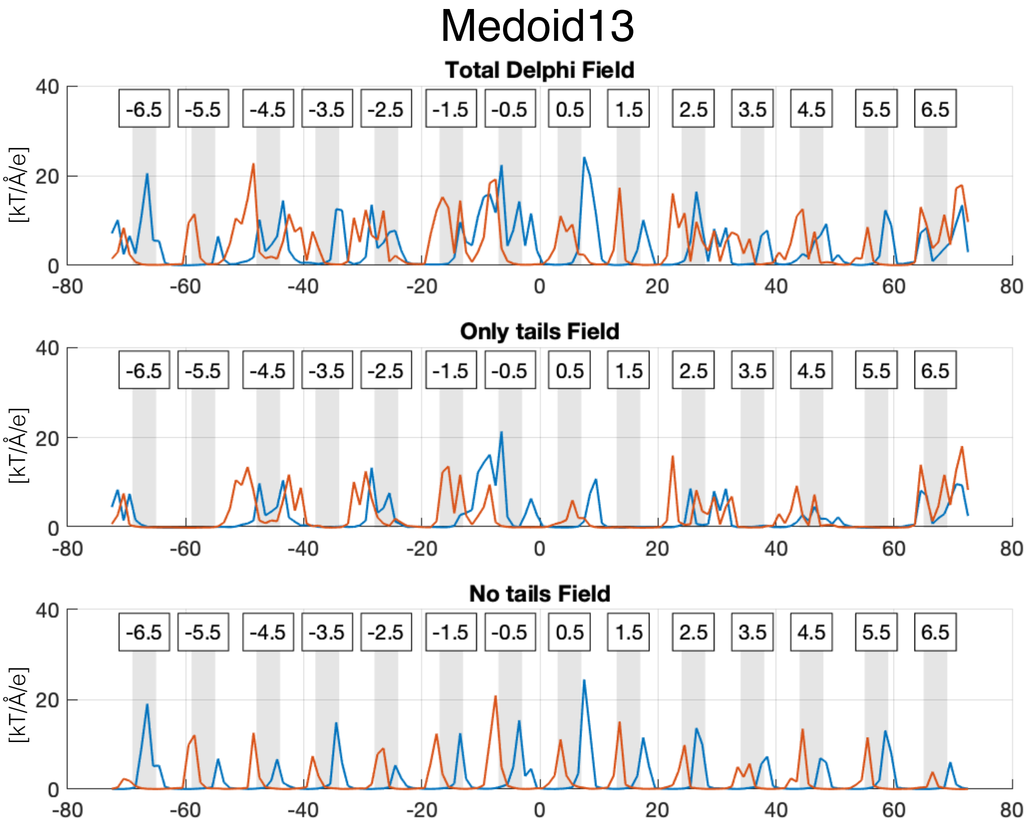


Figure 3.3: Electrostatic field evaluated on the phosphates of the DNA backbone for each DNA chain in medoid13 vs bp number; chain I in blue and chain J in orange. The calculations were performed on three versions of each medoid: the full structure ("total"), a version in which only the charges of the histone tails are non-zero ("only tails"), and a version in which the histone tails are truncated ("no tails"). The charges of the DNA have been put to zero in all structures, in order to consider only the electric field from DNA-histone interactions. SHL regions are highlighted in grey. In the "no tails" structures, the electric field clearly follows the periodicity of the DNA, while in the "only tails" structures the field shows a non-trivial dependence on the positions of the histone tails.

### 3.4.1 DNA - histone tail interactions at the DNA entry/exit site

An area in which the contribution to the electric field owing to the histone tails is particularly marked can be consistently found between SHL -1.5 and -0.5, due to interactions with the tails of H3 and the C-terminal tail of H2A'. These tails appear to have approximately the same positioning across medoids, with the exception of medoid17, in which the H2A' C-terminal tail is found to be inserted in the minor groove (Fig. 3.7D), a conformational change that is strikingly reflected on a more localised and higher in intensity peak in the electric field in SHL -0.5. We remind that SHL regions are highlighted in grey in Figs. 3.3-3.6 and it is defined in Subsection 3.2.2. Medoids 15 and 16 present more similar interactions in this area, compared to medoids 13 and 17. The electrostatic field in the former is characterised by a more diffuse signal, but lower in intensity compared to the more pronounced and localised peaks in the latter. Observing the conformational

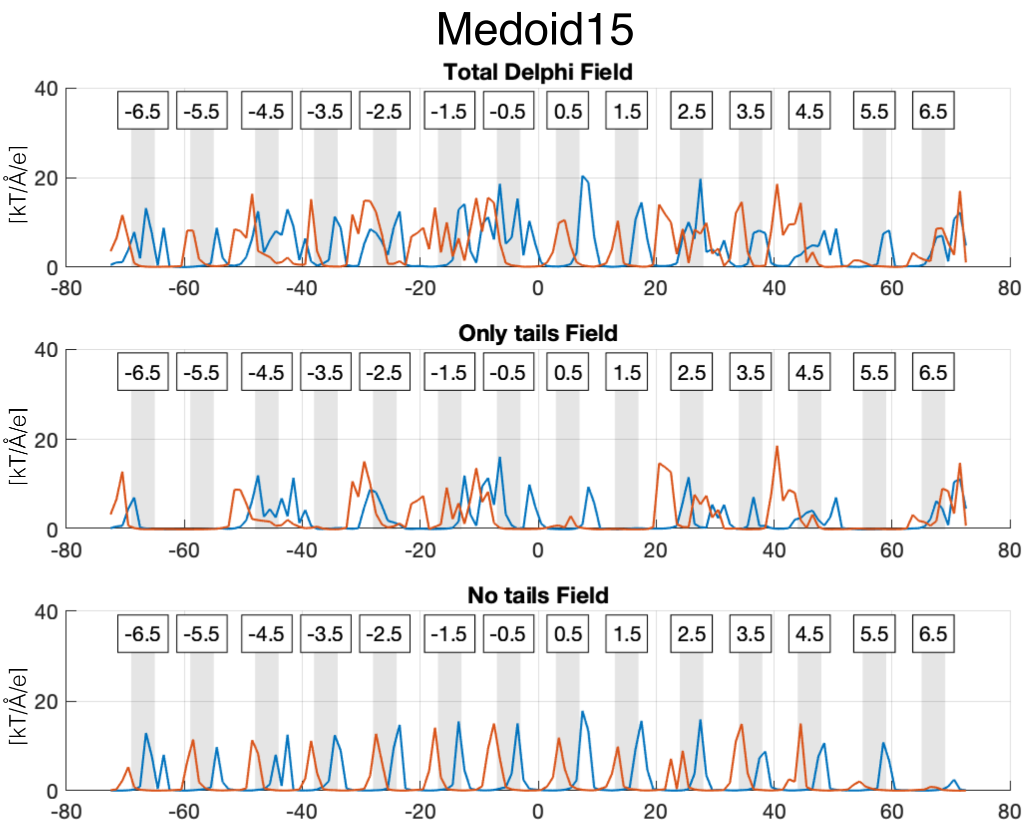


Figure 3.4: Electrostatic field evaluated on the phosphates of the DNA backbone for each DNA chain in medoid15 vs bp number; chain I in blue and chain J in orange. The calculations were performed on three versions of each medoid: the full structure ("total"), a version in which only the charges of the histone tails are non-zero ("only tails"), and a version in which the histone tails are truncated ("no tails"). The charges of the DNA have been put to zero in all structures, in order to consider only the electric field from DNA-histone interactions. SHL regions are highlighted in grey. In the "no tails" structures, the electric field clearly follows the periodicity of the DNA, while in the "only tails" structures the field shows a non-trivial dependence on the positions of the histone tails.

changes between these structures, we attribute this to the conformations of the H3' histone tail near the DNA backbone in medoids 13 and 17. Specifically, residues 27-39 of the H3' tail in medoid13 form a kind of hairpin (Fig. 3.7A). The central residues in this configuration are two glycines, a threonine, and an alanine.

On the other hand, in medoid17 (not pictured) this part of the H3' tail is mostly straight, giving rise to only one instead of two points of close contact with the DNA backbone near the DNA exit site, and overall larger distance between the tail and the DNA in the area between the two gyres. In medoid17 this hairpin structure is closer to the beginning of the tail, and therefore farther from the DNA. We can see the repercussions of these two different conformations on the trend in the electric field in the DNA entry site as well (bps ± 73). Here, medoids 13 and 17 present more moderate electric field values, as less close contact points exist between the tail and the DNA backbone. Finally, in medoids

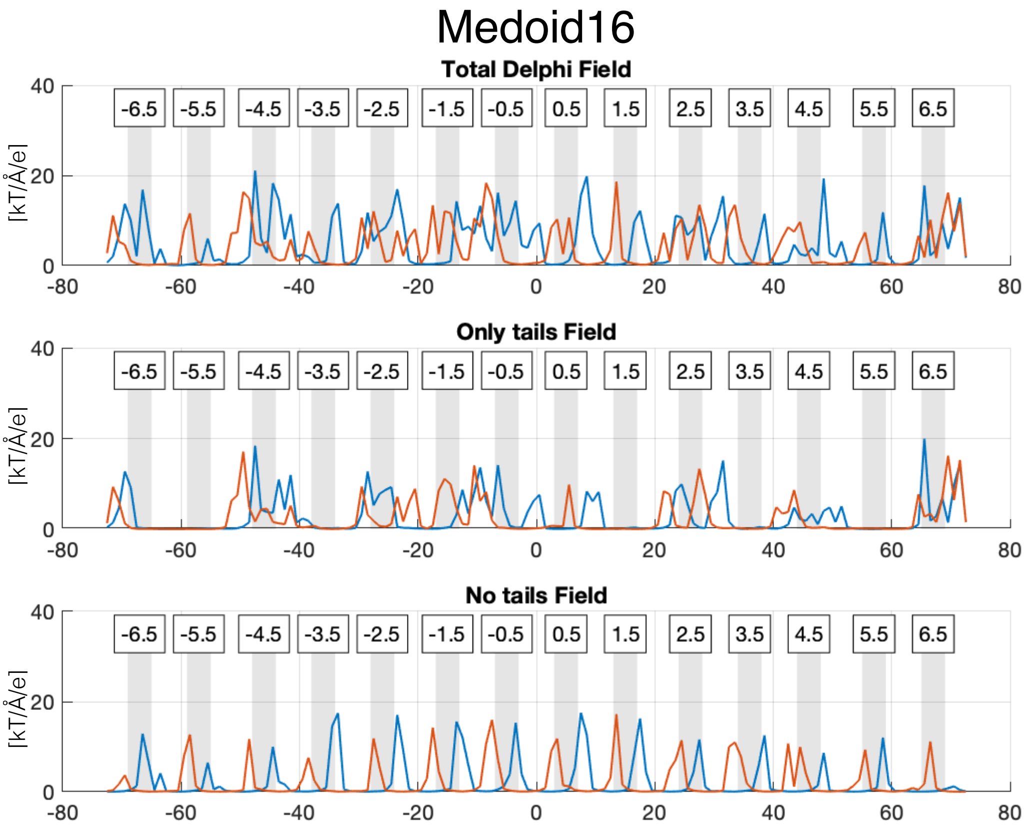


Figure 3.5: Electrostatic field evaluated on the phosphates of the DNA backbone for each DNA chain in medoid16 vs bp number; chain I in blue and chain J in orange. The calculations were performed on three versions of each medoid: the full structure ("total"), a version in which only the charges of the histone tails are non-zero ("only tails"), and a version in which the histone tails are truncated ("no tails"). The charges of the DNA have been put to zero in all structures, in order to consider only the electric field from DNA-histone interactions. SHL regions are highlighted in grey. In the "no tails" structures, the electric field clearly follows the periodicity of the DNA, while in the "only tails" structures the field shows a non-trivial dependence on the positions of the histone tails.

16 and 15 the “hairpin” involves less residues than in medoid13 (just 6-7 residues), as we can determine by visual inspection.

Overall, we observe that the C-terminal tail is fundamental to the changes in the electrostatic interactions in this region of the DNA, since the conformation of the H3’ tail between the two gyres is mostly conserved, while the C-terminal tail’s conformation varies greatly. Since the H3’ tail is known to interact more with linker DNA than with nucleosomal DNA (as confirmed in [20]), this behaviour is consistent with expectations. Furthermore, as we can see in Figs. 3.3 and 3.4, there is a peak in the electric field right on the dyad in medoid13 and medoid15, which is lacking in medoid17 (Fig. 3.6). We remind that the dyad corresponds to bp 0 in our bp numbering convention. The peak is due to a shift of the last two residues of the C-terminal towards the dyad, in a position where the DNA double helix is at its closest to the core. In medoid16 (Fig. 3.5), this peak is shifted

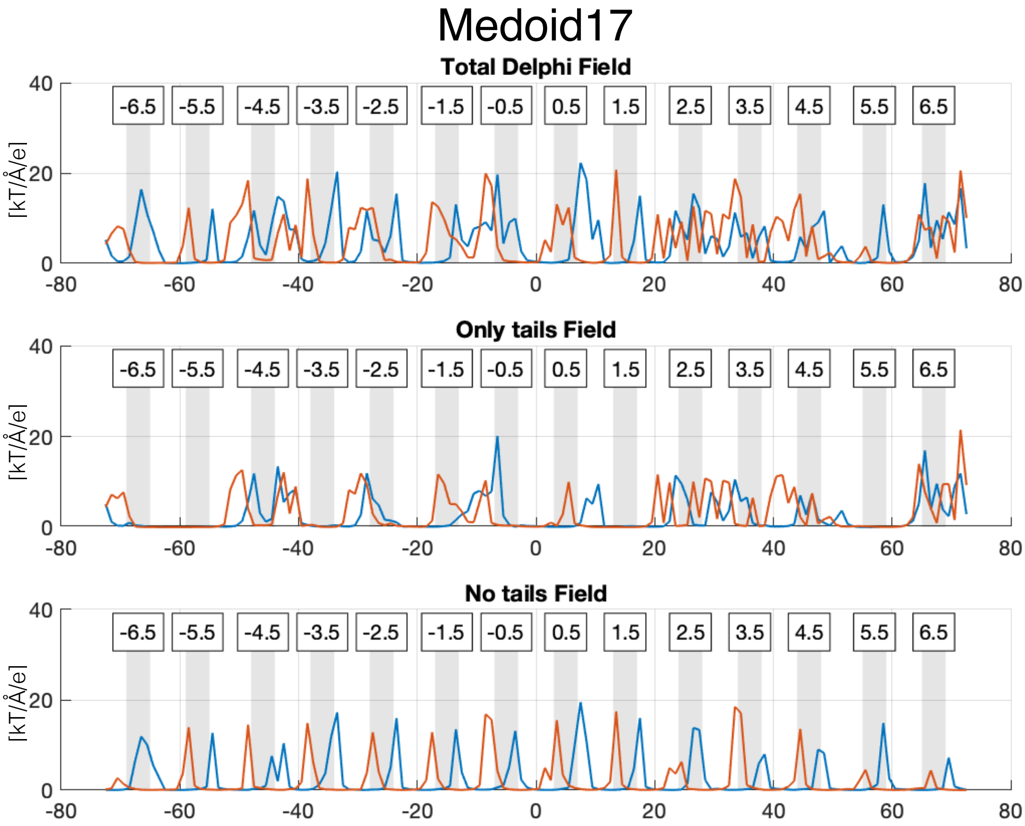


Figure 3.6: Electrostatic field evaluated on the phosphates of the DNA backbone for each DNA chain in medoid17 vs bp number; chain I in blue and chain J in orange. The calculations were performed on three versions of each medoid: the full structure ("total"), a version in which only the charges of the histone tails are non-zero ("only tails"), and a version in which the histone tails are truncated ("no tails"). The charges of the DNA have been put to zero in all structures, in order to consider only the electric field from DNA-histone interactions. SHL regions are highlighted in grey. In the "no tails" structures, the electric field clearly follows the periodicity of the DNA, while in the "only tails" structures the field shows a non-trivial dependence on the positions of the histone tails.

by a couple of bps but it can be ascribed to the same interaction.

Finally, quite a bit of interesting DNA – histone tail interactions occur at the DNA exit site (bp 63-73). These last two superhelical turns are embraced between the short H2A' C-terminal tail from the interior and the long H3 tail, which protrudes between the two DNA gyres, from the exterior. The most diversity in the electric field is seen in medoids 15 and 17, attributable to very different H3 tail conformations. The end of this tail exercises repulsive forces on bps -7 – -10 on chain I, on the other DNA gyre. The shift in the force once again reflects the different conformations of the H3 tail. Observing the 3D structures, we notice that a portion of the H2A' tail forms a short $\alpha$-helix composed by three residues: Ser113, Val114, and Leu115. The final residues of the H2A' C-terminal tail are essentially in the same position in both medoids. Chain I of the DNA mainly interacts with the H2A' C-terminal tail, while chain J mainly does so with the H3 tail. The ability of electrostatics to capture slight structural variations is shown by the presence of a peak in the electric field in bp 66 in chain I, present in medoids 16 and 17 but absent in medoids 13 and 15, correlated to a slightly different conformation of the C-terminal tail (Figs.3.3-3.6).

### 3.4.2 Histone tail effect at negative and positive SHL

We see that all medoids present a peak in the electrostatic field in bp -30 (chain I), because of interactions of the H2B' tail with both DNA gyres. In medoid16, this interaction is displaced with respect to the other medoids: the H2B' tail adopts a different conformation, resembling a double hairpin, and is positioned slightly farther from the DNA than, for example, in medoid17 (Fig. 3.7B). Variations of this "double hairpin" conformation are also found in medoids 13 and 15. The first residues of the H2B' tail are closer to the DNA backbone than the loop of the hairpin. This region is also interesting because, there, the H2B' tail protrudes from between the two DNA gyres, exerting axial and radial forces on the DNA. In conjunction with the presence of the H4 tail in the vicinity, and observing the atomistic structures, we see that the two DNA gyres are slightly closer together.

The same effect can also be seen in other regions of the NCP where the histone tails protrude between the two gyres, especially in the DNA entry and exit sites and the region around bp 50 (chain I). In the region between SHL -1.5 and -2.5 we observe only minor electrostatic interactions in medoids 13 and 17 (Figs.3.3,3.6). This is not the case for medoids 15 and 16 (Figs.3.4,3.5), which present peaks of moderate intensity in those positions on chain J, a behaviour which is once again reflected on the contacts in our proximity analysis by an increase in the number of histone tail atoms within the cutoff (Appendix A Figs. 8-11). There is a larger number of interactions with the H2B' tail at a 10Å threshold in the medoid17, while in the medoid13 they are shifted at distances of 14 and 18Å.

We observe various interactions between the DNA and the N-terminal tails of histones H3 and H4. This part of the DNA is embraced between these two tails, as the H3 tail emerges between the two gyres. Indeed, the variations that we observe in the electrostatic field on the phosphates of the J chain are due to interactions with the H3 and H4 tails in SHL 0.5 (corresponding to -0.5 on chain I). The area between bp 20 to 32 presents a lot of electrostatic features in all medoids, owing to interactions of the DNA backbone with the tail of H4. Only medoid17 presents significant conformational variability with respect to the others: in this case, the tail appears to be shifted in parallel towards the external part of the DNA gyre, a conformational change reflected in a shift in the respective electric field peaks, and on variable axial forces acting on the DNA. In our proximity analysis, we indeed observe more diffuse interactions in the region between bp 20 and 40 in medoid17 (Fig. 11), compared to, for example, much more localised interactions in the same region of medoid16 (Fig. 10). As we will see, this conformational change has repercussions on the forces acting on the DNA backbone.

The region that expresses the greatest heterogeneity across different medoids is the area from bp 20 to bp 53. There, the N-terminal tails of histones H2A', H2B, and H4' interact in succession with the DNA. In the region from residues 47 to 53, the tail of H2B protrudes between the two DNA gyres. However, we can still observe some similarities among medoids. The most striking example is that of bps 32 to 40, in which the electric field on the phosphates is almost null in medoids 13 and 16 (Figs. 3.3 and 3.5). This is caused by the increasing distance of the H2B tail from the DNA, after it emerges from between the two gyres. Once again, we observe an overall lack of contacts between DNA and protein atoms, while in chain I of medoid17 there are at least 100 atoms of the histone tails in the range from 10 to 25Å (Fig. 11). Medoids 16 and 17 are also characterised by an overly low electrostatic signal between bps 40 and 53, owing to the great conformational heterogeneity of the H2B tail. For example, in medoid17 the first residues of this chain fit snugly into the minor groove, between bps 49 and 53, in a completely different conformation compared to medoid16; a sort of hairpin in the latter, a more circular conformation in the former (Fig. 3.7C).

## 3.5 Radial and axial electrostatic forces and their effect on the DNA

It must be noted that the present force calculations are based on representative structures extracted from an MD trajectory where DNA unzipping has not been observed and therefore they could lack some characteristics that are precursory of this process. The plots of the radial and axial electrostatic forces we refer to in the text are provided in Appendix A, for reasons of space. For the normalised radial forces, see Figs. 12-23. For the non-normalised radial forces, see Figs. 48-59. For the normalised axial forces see Figs. 36-47. For the non-normalised axial forces see Figs. 60-71. The normalised forces are used to determine trend changes that are not easily discernible in the plots of the non-normalised values. In the "no tails" structures, we observe a periodicity in the trend of both axial and radial forces roughly corresponding to the double helix pattern, overwhelmingly attractive towards the core. The radial force is particularly attractive at 14 points, as can be seen in Figs. 56-59 in Appendix A. These points correspond to the 14 contact points where the nucleosomal DNA is in proximity of arginines on the histone core, as the minor groove turns to face the histone core.

The histone core and histone tails cause contrasting effects on the DNA backbone, resulting in opposing forces. By examining the plots of the total radial force (Figs. 12-15 and 48-51 in Appendix A) and we see that the repulsive forces are mostly due to interactions with the histone tails, while the radial force is attractive in regions with a low or very low number of DNA/histone tail contacts. Furthermore, we observe sometimes opposed forces acting on the phosphates of pairing bases, this may have mechanical repercussions on the double helix. A characteristic example is the SHL 2.5 region on medoid13. The N-terminal tails of H3 and H2B protrude between the two DNA gyres, stabilising the latter's superhelical structure in the NCP. These interactions are particularly strong near the dyad axis, and in two more regions approximately ±50bp from the dyad, where the H2B tails interact with the DNA [160]. Indeed, by looking at the contributions of the histone tails to the radial and axial components of the electrostatic force we observe that on the dyad, on ±50bp and around that position they are attractive.

In order to have an insight on the dominant effect of the axial component of the

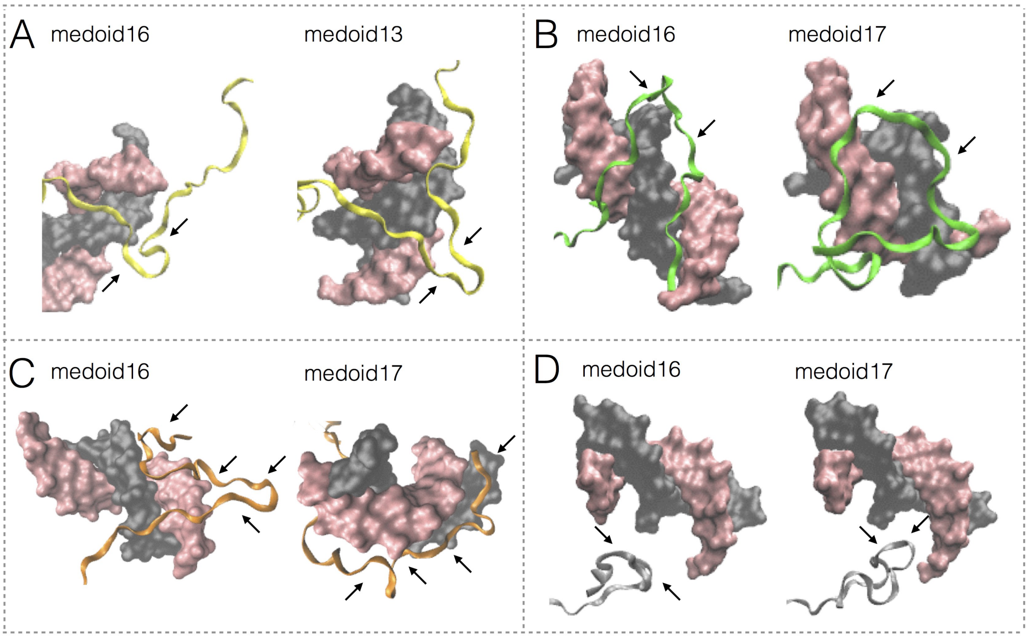


Figure 3.7: **A**: Conformations of the H3' histone tail (yellow) near the DNA backbone in medoids 16 (left) and 13 (right). Surface of DNA chain I coloured in grey and surface of DNA chain J coloured in pink. Residues 27–39 of the H3' tail in medoid13 form a hairpin (indicated by arrows). On the contrary, in medoid17 (not shown) this part of the H3' tail is mostly straight. In medoids 16 and 15 (latter not shown) a similar "hairpin" is present but involves less residues. **B**: All medoids present a peak in the electrostatic field at bp -30 (chain I), because of interactions of the H2B' tail (green) with both DNA gyres. In medoid16 (left), the H2B' tail is positioned farther from the DNA than, for example, in medoid17 (right). **C**: In medoid17 (left) the H2B tail (orange) is inserted into the minor groove, between bps 49 and 53, while in medoid16 (right) it assumes a more circular conformation on the DNA. **D**: The insertion of the H2A' C-terminal tail in the minor groove in medoid 17 (right) versus a larger distance from the DNA in other medoids (medoid16 pictured on the left) causes a more localised and intense peak in the electric field in SHL -0.5.

electrostatic force on the phosphates, we examined their distribution on the different structures (histograms shown in Appendix A, Figs. 24-35). In the absence of histone tails, the axial forces produce a stabilising effect. In the "no tails" version of medoid16 the forces on the lower gyre of the DNA present a wider distribution (Fig. 34). The effect of the histone tails is more markedly attractive/repulsive, with the exception of medoid17, where it is more evenly distributed along the lower gyre, but follows the same trend. Looking at the total axial force on medoid17 we see that there is no clear trend in the lower gyre, and that the forces are more uniformly distributed in the upper gyre as well. The lack of a dominating overall effect could imply a less stable structure in which the two DNA gyres are not kept together as steadily as in other medoids. Similarly, in medoid13 the axial forces are also more evenly distributed, even though slightly attractive and slightly repulsive trends are observed on the upper and lower gyre respectively.

## 3.6 The effect of histone tails on DNA unwrapping

The forces that the histone tails exert on DNA can be studied in conjunction with the positions in which DNA is known to detach from the histone core, or attach more strongly to it. For example, the first barrier encountered during transcription by Polymerase II is at approximately 40 bp from the dyad [159]. We observe strong attractive radial forces immediately after bp ±40 on both DNA chains, even in cases in which the forces were repulsive immediately before that position. This behaviour could be attributed to the nearby arginine residues of the tails of both copies of H2A. The electrostatic interactions on neighbouring phosphates are mostly similar across medoids, with two exceptions: slightly lower intensity of the electric field in medoid13 (Fig.3.3), owing to a larger distance between DNA and protein atoms, and a spike in medoid16 (Fig.3.5), where Lys28 of the H2A tail is found closer to the DNA I chain near bp -47. The interactions with the J chain of the DNA present more heterogeneity, but we observe similarities between medoids 13 and 17 and medoids 15 and 16: in the region from bp -40 to -45, the former present more intense the electric fields, while the latter present lower values. Once again, this is directly linked to the different positioning of the H2A' N-terminal.

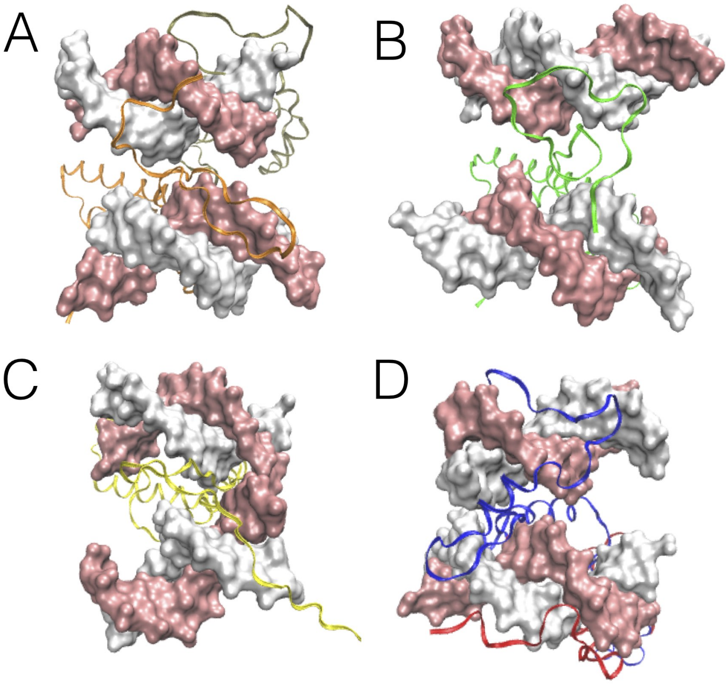


Figure 3.8: Histone tails protruding from between the two DNA gyres of an NCP. Medoid13 was used as reference. Surface of DNA chain I coloured in white and surface of DNA chain J coloured in pink. **A**: H2A (in gray) and H2B (in orange) N-terminal tails. **B**: H2B'. **C**: H3' (NCP DNA exit site). **D**: H3 (in blue) and H4 (in red) embracing the DNA in the NCP entry site. In the sites where histone tails protrude from between the two DNA gyres, the latter are brought closer together, stabilising the structure.

Another important feature that transpires from the total radial components is that the histone tails are responsible for a repulsive radial force of varying intensity across medoids. This force is present when the histone tails protrude from between the two DNA gyres in the DNA entry and exit regions, favouring a possible unwrapping process. It has been observed by Kono et al. that the outer turn of the DNA in NCPs unwraps following a three-step asymmetric procedure [161]. First, 5bps unwrap from one end of the DNA, followed by 5 more bps from the same end, without further increase in free energy. Then, unwrapping starts at the other end, where a total of 10bps is unwrapped in the same way. Finally, the process is completed by the unwrapping of 15 more bps at both ends. NCPs have also been experimentally shown to unwrap in stages [162], with the first half of the bps unwrapping at a low force, while the second half requires a higher force. The required force was significantly lower in cases where the histone tails were truncated, indicating once again their important role in NCP stabilisation. This statement is coherent with our observation that the histone core and the histone tails sometimes produce opposing effects.

The H3 tail remains in contact with the DNA near the entry and exit points of the NCP until the initial 10bp are peeled off both sides. During the following stages, gradual changes are observed in the contacts between the DNA and the H2A/B tails. The positions on the NCP in which unwrapping is seen to be temporarily halted in Kono et al. can be correlated with the positions of the histone tails. In particular, we see that the first five bp on the J chain on the entry sites of our medoids are in contact with histone tails H2A C-terminal and H3, while bps on chain I present contacts in the first 8 bps with these tails (see Figs. 72-79 in Appendix A). Interestingly, in the SHL ±3.5 region, which corresponds to the end of the total 25 bp that unwrap from each side, there is a relative absence of DNA-histone tail interactions, indicated also by a lack of a significant electrostatic field influence due to the histone tails in that region (Figs. 8-11). The total radial force is attractive here, due to the dominant effect of the histone core, but it is zero or even repulsive in the regions immediately before this (Figs. 12-15) . Therefore, we conclude that the histone tails can tune DNA unwrapping, by increasing the tendency of the DNA to unwrap from the histone core.

## 3.7 Repercussions of histone tail truncation

The truncation of the H3 and H4 tails has been shown to produce opposite effects [154]: the removal of the arginine- and lysine-rich H3 N-terminal favours DNA dissociation from the histone core, while H4 truncation stabilises the wrapped conformation. As we previously mentioned in this Chapter, these tails protrude between the two DNA gyres near the linker DNA entry and exit points. The radial forces exerted by the histone tails tell a different story: in this region the force is repulsive in all medoids except for chain J in medoid 17, highlighting the role of Lys36 (Figs. 16-19). Given the contrasting effects of the histone core and the histone tails when it comes to the radial component of the force in the DNA entry and exit regions, we see that the trend is not always markedly repulsive or attractive: for example, it is neutral in SHL -6.5 in medoid13, and in SHL -0.5 in medoid17. The total axial components are mostly repulsive on both chains in the DNA exit site (Figs. 36-39), indicating that it tends to move upwards from the dyad plane, but they are mostly attractive in the entry site.

It has been observed that the end stretches of the DNA spontaneously unwrap 1-10% of the time (the propensity to unwrap is sequence dependent), and that target sites located inside the NCP (at smaller SHL absolute value) are less accessible than sites located near the DNA entry-exit points [163]. In Iwasaki et al., four mutant nucleosome structures were studied, each of which lacked histone tails from a specific histone [160]. They observed that H2B and H3 deletion substantially decreased NCP stability, while they also found that H2A and H4 deletion caused opposite effects. In particular, H3 deletion was seen to enhance unwrapping at the DNA entry and exit sites of the NCP, which are very near the points in which the H3 tail protrudes from the core (Fig. 3.8D), possibly enhancing nucleosome sliding. In our analysis, we see repulsive axial forces with respect to the NCP core near bp 10 of chain I (Figs. 44-47), a region of the DNA in close interaction with the glycine- and lysine-rich - and therefore flexible and positively charged - H4' histone tail.

In the structures where the histone tails were truncated (Figs. 44-47), all medoids present a repulsive trend in the axial component in chain J from bp 0 to -19, with a minimum on bp -7. In the same area chain I presents the same jigsaw pattern we encountered in the electric field and electrostatic potential, while there is an absolute maximum in the force acting on bp 7. In the total axial force (Figs. 36-39), we see a switch from an attractive trend in medoids 13 and 17, where less atoms are found near the DNA backbone around bp 35, to a repulsive trend in medoids 15 and 16, in which there are more atoms in that region. In chain I we observe overall repulsive forces in all medoids from bp 1 to bp 18, with a peak on bp 7, which we attribute to the histone core, and particularly the loop connecting the $\alpha$-helices of histone H4'.

Furthermore, there is a region presenting variability across medoids from bp -40 to -55. As we mentioned in our analysis of the electric field, this region presents interactions with the H2A N-terminal tail, and the contribution is indeed due to the histone tails, as we see from the non-normalised forces. We attribute this force to the action of the H2A N-terminal tail, that embraces the DNA gyre from above in those positions. Comparing the total axial components to the "only tails" structures (Figs. 40-43) we see that there is a repulsive force in SHL -5.5 across medoids due to the core, while a contribution from the tails is absent. As in the case of radial forces, sometimes the core and tails appear to cause contrasting effects. For example, in SHL 5.5 chain J the tails present a consistently repulsive trend (Figs. 16-19), while the total contribution is attractive (Figs. 12-15).

Brower-Toland et al. [162] observed that the removal of the H2A and H2B N-terminal tails induced a decrease in the histone-DNA interaction strength at a position ±36bp from the dyad axis. In our medoids, the H2A and H2B N-terminal tails protrude from between the two DNA gyres in positions near bp ±30 (Fig. 3.8A). Comparing the total electric field on and near those DNA bps with the contributions of the histone core and histone tails, we see that the electric field is indeed mainly due to the histone tails in those points. In order to see if these tails are a stabilising or destabilising component of the NCP, we examine the effect of the histone tails on the axial component of the force. Indeed, in those positions the histone tails have a stabilising effect, as the axial forces present an overall attractive tendency. Juxtaposing the axial force contribution of the histone core (Figs. 68-71) and that of the histone tails (Figs. 64-67) in that position, we see that the latter is of higher intensity. However, the radial component shows that the histone core (Figs. 56-59) exercises a strong attractive force on those positions, while the histone tail contribution (Figs. 52-55) is either close to null for bp-30 or even repulsive for bp 30.

In most approaches, the role of the C-terminal tail appears to be downplayed. It is, however, interesting to see how the truncation of the H2A C-terminal domains change nucleosome dynamics, as they have been observed to increase the nucleosome sliding rate [164]. Cells expressing truncated H2A C-terminals showed increased stress sensitivity and nucleosome mobility. This particular histone tail has two important functions: it stabilises the NCP and mediates interactions with other proteins. Simultaneous deletion of histone tails from more than one histone has been observed to relate to compromised cell survival in yeast [165]. In our analysis, we also observed the important effects that the C-terminal tail has on nucleosome electrostatics, since even a shift of two residues can produce significant electrostatic effects.

Besides histone tail truncation, sometimes entire histones might be missing from the NCP [166] forming Partially Assembled Nucleosome States (PANS). Nucleosomes have the ability to dissociate entirely in histones and DNA, and then reassemble [46], in a process driven by electrostatic interactions. Rychov et al. [26] analysed three types of PANS (hexasomes, tetrasomes, and disomes) through MD simulations. The nucleosome formation procedure was observed to occur as such: the two H3 and H4 dimers bind to the DNA first, forming a tetrasome, followed by the sequential addition of H2A and H2B dimers. MD dynamics studies on PANS reveal that the nucleosomal DNA is drastically deformed when histones H2A or H2B were missing, but that the loss of H3/H4 did not impact on DNA conformations as much, because of the action of the H2A C-terminal tail, which filled the space left empty by the missing histones. The method we propose in this work could be applied on structures representing intermediate states of association/dissociation MD trajectories, to investigate the fleeting contacts formed between the histone tails and the DNA, and the role of the histone tails in NCP assembly.

## 3.8 Arginines and Lysines: key histone tail residues

The particular importance of certain protein residues, particularly arginines and lysines, is evident from several interactions in NCPs. The positions of these residues in the nucleosome is illustrated in Fig. 3.9. For example, looking at bp -20 of chain J in medoids 16 and 17 we see a change in the trend of the force from attractive to repulsive. We can attribute this change to a variation in interactions with two particular arginine residues: Arg17 and Arg19 the H4'. In medoid 16 these arginines are found at a distance of 3Å and 12Å respectively from the phosphate of bp -20. However, in medoid 17 their respective distances from the same phosphate are 8Å and 6Å. In medoid13, on chain J bp 54 there is a strong repulsive force, because of close interactions of the phosphate with Lys17 of chain D (H2B).

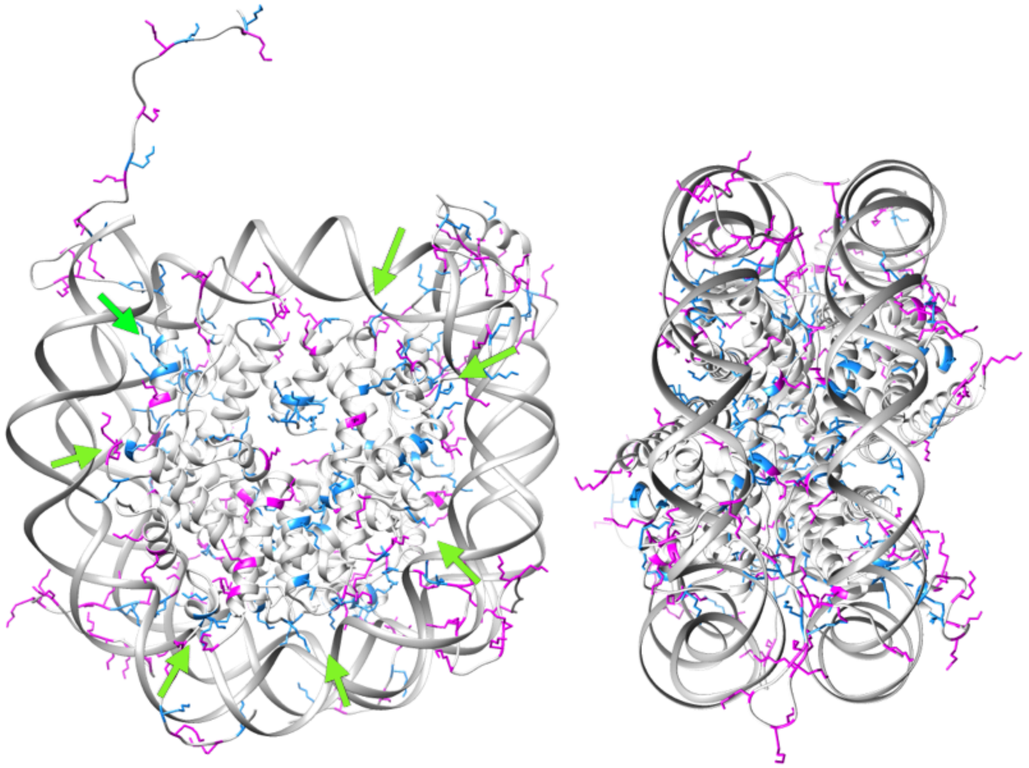


Figure 3.9: Arginine (in blue) and lysine (in magenta) residues in the nucleosome. The presence and location of these residues strongly affects nucleosome electrostatics. The DNA is anchored to the histone core in 14 contact points with arginines, 7 on each side of the NCP, indicated by arrows. Lysines and their PTMs, especially acetylation, impact on inter- and intra-nucleosome interactions.

In medoid16 we observe a repulsive force on bp -65 of chain I, found close to chain E Arg52, and on bp -54 on chain I, which is close to chain D Ser52/53. Interestingly, there is no particular effect on the corresponding bps in chain J in either case. In medoid15, on the other hand, we observe strong repulsive forces on bps belonging to chain J, notably on bp ± 65, because of a contact with chain E Lys56. In this case we notice no corresponding effect on chain I bps. However, a repulsive force is exerted on chain I bp -36, possibly due to a propagated effect from chain J bp 36, on which the total radial component of the electrostatic force is negligible, while in the neighbouring bps the force is attractive, as a result of interactions with the backbone of Arg42 (chain C).

The conformations and dynamics of the histone tails are greatly altered by Post-Translational Modifications (PTMs), sometimes triggering effects that can even supersede native electrostatic interactions. Acetylation, for example, leads to a reduction of the overall positive charge, reducing self-repulsion, and facilitates more hydrophobic interactions. Furthermore, it increases the volume occupied by the side chains of the affected residues, bearing steric implications, and drastically reduces the interactions of the H4 tail with the acidic patch of the same or neighbouring NCPs, with important repercussions on tail-bridging and inter-NCP interactions. Hyperacetylation of the H4 histone tail [167] can ultimately result to enhanced DNA accessibility from DNA-binding proteins [168, 169], acting as a chemical signalling method for transcription, making the H4 N-terminal region vital for chromatin compaction. Acetylation of H4 tails has been shown to increase their propensity in forming $\alpha$-helices, ultimately making them shorter [154], further hindering their interactions with acidic patches.

The H4 tail has been observed to form the largest number of protein-DNA contacts at physiological salt concentration, particularly in the region between residues Lys16 to Arg23 [32]. Naturally, these contacts are disrupted upon the presence of PTMs. Both H3 and H4 acetylation have been observed to enhance DNA unwrapping in ionic concentrations higher than physiological [27]. Moreover, H3 acetylation renders the NCP more sensitive to counterion-induced dissociation and histone dimer exchange between NCPs, while H4 acetylation has opposing consequences [27].

H3 tail acetylation enhances DNA breathing at physiological ionic concentrations, while H4 tail acetylation has this effect at higher ionic concentrations. At low concentrations, acetylation was not observed to correlate with DNA binding propensity. Potoyan and Papoian [167] observed that Lys16 acetylation disrupted H4 tail binding and folding landscapes, enhancing $\alpha$-helix formation. The same PTM was also observed in FRET experiments to result in a tightening of the linker DNA in the entry and exit points of the NCP, opposing the tail-bridging effect and weakening inter-NCP interactions, therefore disrupting NCP stacking in the chromatin fibre. Progressive acetylation of the H4 tail also showed cumulative effects [59], caused by the progressive charge reduction and increased hydrophobicity induced by the clustering of acetyl groups.

Tail truncation and lysine hyperacetylation produce similar results, from the electrostatic point of view, since the positive electric charge is reduced in both cases. For example, H3 and H4 tail truncation has also been observed to increase DNA accessibility, as is the case with hyperacetylation of these tails [170]. We note that the steric implications of these two processes are naturally different, and result in different configurations and interactions between residues in the NCP. PTMs are often hard to study because they are challenging to detect reliably in a sample containing many nucleosomes. However, by understanding the electrostatic implications of the charges present on the histone tails and by examining the electrostatic interactions that take place in their absence, we can improve our knowledge on the consequences of PTMs as well.

## 3.9 Extension to Coarse-Grained Models

Becker and Everaers [155] in their work on the CG base-pair level model of the nucleosomal DNA used X-ray structures to infer the forces acting on the DNA. They observed a periodic pattern of high force peaks, corresponding to the contact points of the nucleosomal DNA and the histone core. They correspond to peaks in the elastic energy of the DNA, and the pattern is similar to the trend we observe for the electric field and the axial component of the force in structures in which the histone tails are absent. Furthermore, Becker and Everaers analysed the deformations of the nucleosomal DNA, and extrapolated the forces and torques that would be necessary to induce such deformations,

| Histone Tail | Properties | Medoid Features |
|---|---|---|
| H3/H3' N | Interacts with linker DNA<br>Can form stable folds<br>Truncation hinders NCP stability | Repulsive radial forces in all medoids (Lys36)<br>(H3') straight in medoid13, hairpin in others |
| H4/H4' N | Truncation favours NCP stability<br>Interacts with acidic patch | Diffuse interactions in medoid17, localised interactions in others<br>Large number of contacts with DNA in SHL -0.5 in all medoids |
| H2A/H2A' N | Halts PolII | Intense electric field in medoids 13 and 17<br>(H2A) Lys28 very close to DNA causes spike in medoid16 |
| H2A/H2A' C | Embraces DNA<br>Interacts with linker DNA<br>Stabilises PANS | Interaction on dyad (H2A') in minor groove in medoid13,<br>off dyad in medoids15/16, No interaction in medoid17<br>Field peak on bp 66 in medoids 16/17 |
| H2B/H2B' N | Protrudes between DNA gyres<br>Truncation hinders NCP stability<br>Great conformational heterogeneity | (H2B) circular conformation in medoid 16, hairpin in medoid17<br>(H2B') double hairpin in medoid16<br>(H2B) inserted in minor groove in medoid17 |

Table 3.3: Summary of properties of histone tails observed in this work: Histone tail Role in the dynamics of the nucleosome, and particular electrostatic and conformational features of different medoids.

interpreting them as interactions between the nucleosomal DNA and the histone tails.

However, in the work of Becker and Everaers a single static structure was used, and therefore the large variety of different histone-tail we observed in our analysis was not taken into account. In our simulations, the electrostatic potential and electric field presented this periodicity in the structures that do not include the histone tails. In addition, we clearly observe peaks in the intensity of the radial component of the electrostatic force on the 14 contact points. Even so, our results clearly indicate that the histone tails are of central importance to the electrostatics of NCPs, and that forces caused by these interactions have important repercussions in the stability of the NCP, in DNA unwrapping propensity, and therefore in transcription. We observe a non-trivial dependence between the presence of histone tails, and their distance from the DNA, and the magnitude of the electric field.

Even though the histone core is overwhelmingly attractive towards the DNA in the radial direction, in some cases the histone tails cause opposing effects. In the entry and exit sites of the NCP, the DNA interacts strongly with H3 and H2A C-terminal tails, by which it is embraced. The electrostatic interactions with these tails were seen to tune spontaneous unwrapping, affirming observations previously made in literature. We consistently observed that, in the places where histone tails protrude from between the two DNA gyres, the DNA gyres appear "pinched" closer together (Fig. 3.8), and perceive a repulsive radial force, pushing them away from the histone core. Notably, the H2A N-terminal tail exercises attractive electrostatic forces towards the histone core in positions where Polymerase II is known to briefly halt its progress along the DNA. We want to stress the importance of H2A C-terminal tail: we clearly saw that small changes in its positioning had a significant impact on the electric field, implying that this often overlooked feature of the NCP holds particular importance in the dynamics of the nucleosome, corroborated by the position of the H2A C-terminal tail, close to both the dyad and the DNA entry/exit sites. The H2A C-terminal tail has been seen to be important for cell survival and function.

Therefore, we believe that the effects of the histone tails should be a central part of CG models regarding nucleosomes and chromatin. Furthermore, studies that infer the forces acted upon DNA by the histone tails and connect them to the structural features of the nucleosome can be particularly useful, since histone tails are notoriously hard to observe experimentally at atom-level, or even residue-level, resolution. Finally, electrostatic forces are a fundamental ingredient in CG models of such highly charged systems and can provide a mechanistic interpretation of chromatin dynamical processes and conformational equilibria.

## 3.10 Conclusions

In this Chapter, we propose a methodology for the study of protein-DNA electrostatic interactions and we apply it to clarify the effect of the histone tails on the NCP. Our methodology correlates electrostatic interactions and spatial conformation at the residue level with mechanical effects and repercussions on their structure and function. Studying the NCP, a complex protein-DNA system, we were able to identify the possible electrostatic origins of many effects, such as spontaneous DNA unwrapping, NCP destabilisation upon histone tail truncation, and the key role of specific arginine and lysine residues. We study four representative structures extracted from a 1μs full-atom MD simulation in explicit solvent, to capture some relevant conformations of the histone tails, and we separate their electrostatic contributions from those of the histone core, to assess their individual importance.

We construct contact maps of the histone tails with the DNA, to study the correlation between the number of histone tail atoms in proximity to the DNA, and the intensity of the electrostatic interaction. We stress the role of H2A C-terminal tail: small changes in the positioning of this short tail had a significant impact on the electric field, implying that this often overlooked feature of the NCP holds particular importance in the dynamics of the nucleosome. We consistently observed that, in the places where histone tails protrude from between the two DNA gyres, the latter appear "pinched" closer together (Fig. 3.8), and perceive a repulsive radial force, pushing them away from the histone core. Notably, the H2A N-terminal tail exerts attractive electrostatic forces towards the histone core in positions where Polymerase II is known to briefly halt its progress along the DNA. Even though the histone core is overwhelmingly attractive towards the DNA in the radial direction, the histone tails can cause opposing effects, such as H3 and H2A C-terminal.

Correlating our observations on the positions of histone tails with the estimates of electrostatic interactions with DNA, we confirm our hypothesis that they tune DNA unwrapping by the repulsive and attractive forces they exert on DNA, and their inherent conformational variability, in a non-trivial way. This "breaks the symmetry" presented in CG models that only consider the deformation of the DNA and interactions with the histone core, and points to more complex interactions taking place in nucleosomal DNA. We therefore propose the application of our methodology in conjunction with CG models that treat the mechanical properties of the DNA.

### CHAPTER 4

# Nucleosome Porosity and Solvation Interactions: Zeta Potential of Individual Nucleosomes

In this Chapter, we conduct a more in-depth analysis of the porosity of NCPs and the importance of solvation phenomena that we touched upon in Chapter 2, Section 2.3. We expand our qualitative analysis of NCP porosity and provide more detailed maps of the cavities present in the NCP. We complement and support our computational findings on inter- and intra-NCP electrostatic interactions reported in Chapters 3 and 5 with experimental data. We discuss Zeta Potential and Dynamic Light Scattering (DLS) measurements on single nucleosomes under varying ionic concentrations, providing information on the surface charge and the size of NCPs.

We then compare these measurements with calculated Zeta Potential ($\zeta$) values using DelPhi, and discuss the position of the shearing layer around the NCP. To our knowledge, this is the first time that $\zeta$ measurements have been employed in the study of nucleosomes. We find that the $\zeta$ generated by nucleosomes becomes monotonically less negative in increasing NaCl concentration, as expected. We also find that the diameter of the structures, as measured with DLS, increases abruptly at 1M NaCl concentration, indicating the formation of agglomerates. It has been reported in the literature that at this ionic concentration nucleosomes unravel [140], consistently with our observations. The agreement between our experimental data and our simulations validates our methodology, and shows the importance of the use of the full non-linear PBE in the study of nucleosomes.

The results and analysis presented in this Chapter are included in a manuscript in preparation at the time of writing.

## 4.1 Introduction

In the previous Chapters, we have seen how the PBE can be exploited to study the electrostatic potential, the electrostatic field and electrostatic force in intra-nucleosomal interactions, as the histone tails exert electrostatic forces on the nucleosomal DNA. We also saw how we can use the PBE to measure the interaction energies of nucleosome pairs. However, the PBE can also provide useful insights about solvent-mediated interactions occurring in a system: in [119], PBE analysis was used, alongside MD, to study the porosity of NCPs. As is the general case in biomolecular dynamics, solvent-mediated interactions are of crucial importance in NCP behaviour, especially because of the electrostatic screening effects exerted by water molecules and ions. It has been observed that the NCP structure is stabilised partly by solvent-related interactions [17, 141].

As we mentioned in Chapter 2, Section 2.3, the notion that the nucleosome is a solid object, impermeable to the solvent has been rebuffed. The high porosity of the NCP facilitates the permeation of water molecules and solvated ions into the histone core, further boosting the neutralising effect of counterions, and enhancing ionic interactions of biological significance [119]. Indeed, the NCP has been found to be solvated by more than 1000 water molecules, leading to an unusually high dielectric environment around the histone core. The role of water molecules is not just electrostatic in nature, as they can cause slight modifications on the surfaces of the histones, facilitating nucleosome dynamical effects, such as nucleosome sliding [24].

One of the few methods that provides experimental information on the electrostatics of solvated particles that can be, for instance, compared to the results of PBE calculations is the measurement of the $\zeta$ of a solute. This technique evaluates the average potential at some distance from the surface of solvated particles, and is primarily used in colloid chemistry and on liposomes and proteins. $\zeta$ measurements are non-invasive, and they do not require aggressive sample preparation procedures. The notion of $\zeta$ refers to that of the shear plane around a solvated particle: the boundary between water molecules the mobility of which is significantly altered by the presence of the solute and those that are less affected and therefore are more similar to the bulk solvent.

The $\zeta$, or electrokinetic potential, is the electrostatic potential averaged over the shear plane of a solvated particle. The shear plane is defined as the boundary between the solvation layer around the particle and the bulk solvent. The width of the solvation layer depends on the composition of the solvent and some thermodynamic parameters (temperature and pressure). For a solvent containing ions, an electric double-layer is formed around the charged particle, in which counterions seek to neutralise the charge of the solute (see Chapter 2). The double-layer consists of two regions, the internal Stern layer or adsorbed layer, in which ions are forming specific interactions, not only of electrostatic nature, and the external diffuse layer, in which ion mobility is possible. Experimentally, $\zeta$ is determined from measurements of electrophoretic mobility in suspension. A more detailed discussion of $\zeta$ and DLS is not the aim of this work. For more information readers can refer to works such as [172, 173].

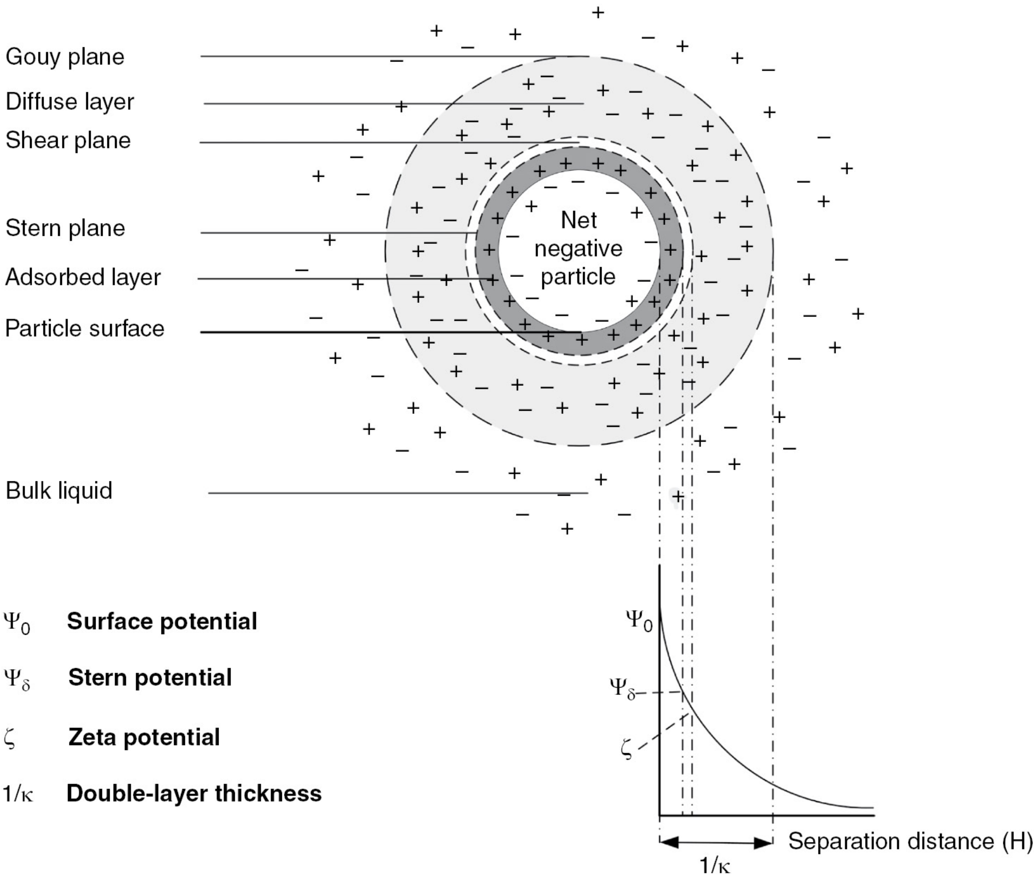


Figure 4.1: Schematic representation of the electric double-layer surrounding charged particle immersed in a dielectric solution, and of the trend followed by the electrostatic potential as a function of radial distance from the particle surface. Distinction between surface potential, Stern potential, and $\zeta$. Illustration from [171].

## 4.2 Computational and Experimental Methods

### 4.2.1 NanoShaper Nucleosome Porosity Analysis Parameters

In order to create Fig.2.3 illustrating the electrostatic potential and the central cavity and channel traversing the nucleosome, surfaces were generated with VMD using NanoShaper under "Drawing Method" in the "Draw Style" tab of the "Graphical Representations" menu. We used the SES Surface Type, a Solid Surface Representation Method for the electrostatic potential case, and a Wireframe Representation Method in the case of the cavities of the nucleosome. The values of the electrostatic potential on the surface of the DNA and histones were calculated with DelPhi and visualised by loading a .cube file (DelPhi output, having provided an .frc file as input) on the PDB file in VMD. In order to view the internal cavities, we used NanoShaper in `pockets` Operative Mode, with a big probe radius of 3A and a small probe radius of 1.5. We selected 12 water molecules per pocket, and a grid scale of 2, with a grid perfil of 90%.

### 4.2.2 Zeta Potential Calculations with DelPhi

We calculated the electrostatic potential at points placed at given distances from the molecular surface of 1KX5, using DelPhi interfaced with NanoShaper. We generated the points by increasing the radii of the atoms in the original PQR file by a desired amount using a Python script. The distances at which the points were generated were determined by considering the position of the shear plane as a function of ionic concentration (further analysis on this can be found in the subsection 4.4). The following example illustrates the procedure we followed for the generation of the input files: let us think of an increase of 5.6Å, approximately corresponding to the diameter of two water molecules. This PQR file is passed to DelPhi as input, generating an scrg file (Delphi parameter file setting `out(scrg,file='1kx5_5.60.scrg',form='pdb')`). For the point generation step, the `scale` parameter was put to 1 (grid spacing 1Å) and `perfil` to 95. Ionic concentration and probe radius were put to zero, as this run was performed solely to generate the points on which to calculate the potential. For more information on the 1KX5 nucleosome structure and other nucleosome crystal structures used in this work see Table 3.3.

We then calculated the potential on every point included in the scrg files at varying ionic concentrations. The NaCl concentrations we have selected were 5, 10, 20, 50, 137 and 250 mM, in order to compare our measurements with the data from $\zeta$ measurement experiments we performed. In the second part of the simulation, we used the non-linear PBE, because the system in question is highly charged, and therefore the linearised PBE is less accurate. We impose `ni=1200` (non linear iteration parameter in DelPhi). A cubic box of grid spacing 2 and `perfil` value of 80%. This run generates an .frc file, which contains the values of the potential and the coordinates of the points the potential was calculated on. The potential in Delphi is given in kT/e, where k is the Boltzmann constant, T the temperature, and e the elementary electric charge. We then calculated the mean value of the potential, and converted it to mV. Electrostatic potentials were computed using the DelPhi non-linear PBE solver, interfaced with NanoShaper [135]. The solvent was assigned a dielectric value of 2, and the solute a value of 80. NanoShaper probe radius was 1.4. Other parameters assumed default values.

### 4.2.3 Zeta Potential Measurements on Single Nucleosomes

Mononucleosomes assembled from recombinant human histones expressed in *E. coli* (two each of histones H2A, H2B, H3 and H4) wrapped by 147 base pairs of 601 positioning sequence DNA, were a product of EpiCypher (Durham, NC, USA). Mononucleosomes (100 µg total mass (protein and DNA), 54.6 µg protein) were delivered in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 25 mM NaCl, 2 mM DTT, and 20% glycerol. Aliquots (10 µg) were kept at the temperature of $-20^{o}$C and defreezed before sample preparation. NaCl was a product of Sigma Aldrich (Saint Louis, MO, USA). Buffer solution containing NaCl in concentration 5, 10, 20, 50, 137, 250 and 1000 mM were prepared. For the DLS and $\zeta$ measurements, 2.5 µg of nucleosomes were dissolved in 600 µL of ultrapure water (resistivity 18 MOhm). DLS and potential measurements were collected with a NanoSizer Instrument (Malvern Panalytical, Malvern, UK). Three independent measurements were collected for each sample.

## 4.3 Nucleosome Porosity: A Quantitative Analysis

As we mention in Chapter 2, Section 2.3, in order to illustrate the porosity of the nucleosome, particularly described in [119], we conducted a study on the nucleosome crystal structure (PDB code 1kx5 [90]) using NanoShaper interfaced with VMD [135, 149]. We visualised the channel traversing the nucleosome core, and constructed an electrostatic map of the nucleosome, using data from the DelPhi Poisson Boltzmann solver [131] on the potential and constructing the SASA of the nucleosome with NanoShaper, as seen in Fig.2.3.B, where it is possible to clearly see, among other features, the position of the acidic patch (residues E56, E61, E64, D90, E91, E92 of H2A and E102, E110 on histone H2B [15]), and the highly charged histone tails, both key elements in chromatin compaction and chromatin interaction with DNA-binding proteins. This analysis showed a minor acidic region, on the surface of histone H4.

The nucleosome is usually treated in models as a solid particle (a sphere, or oblate spheroid, or ellipsoid) [60, 66, 69], a treatment which directly implicates the omission of any solvation effects in the interior of the structure. However, it has been shown that nucleosomes are highly porous, enabling water molecules and ions to penetrate to the core of the particle [119]. To analyse the porosity of the nucleosome, we have conducted a quantitative study on the nucleosome crystal structure (PDB code 1KX5 [90]) using NanoShaper [135] interfaced with VMD [149], providing the values of the Surface to Volume Ratio (SVR), the number of internal and superficial cavities. We observe an SVR of $0.387\ Å^{-1}$, which reflects a quite high porosity [150], 11 closed cavities with volumes ranging from $20.62Å^3$ to $55.75Å^3$ and 31 pockets. The NCP features a channel traversing its core, which significantly impacts on NCP accessibility to water and ions (Fig.2.3.B). In Fig.4.3, we show the superficial cavities of 1KX5.

Our results are consistent with previous qualitative analyses by Materese et al. [119], and indicate that the nucleosome is highly solvated and porous. We have also constructed an electrostatic map of the nucleosome, using data from DelPhi on the potential and constructing the SASA of the nucleosome with NanoShaper , as seen in Fig.2.3.C, where it is possible to clearly see, among other features, the position of the acidic patch (residues E56, E61, E64, D90, E91, E92 of H2A and E102, E110 on histone H2B [15]), and the highly charged histone tails, both key elements in chromatin compaction and chromatin interaction with DNA-binding proteins. We also observe a minor acidic region, on the surface of histone H4.

## 4.4 Zeta Potential Measurements on Single Nucleosomes in Varying Ionic Conditions

Non-invasive experimental observations of quantities related to electrostatics and solvation in chromatin can be done via DLS and $\zeta$ measurements. Such measurements can be done on single nucleosome dispersions under varying ionic conditions, providing a high

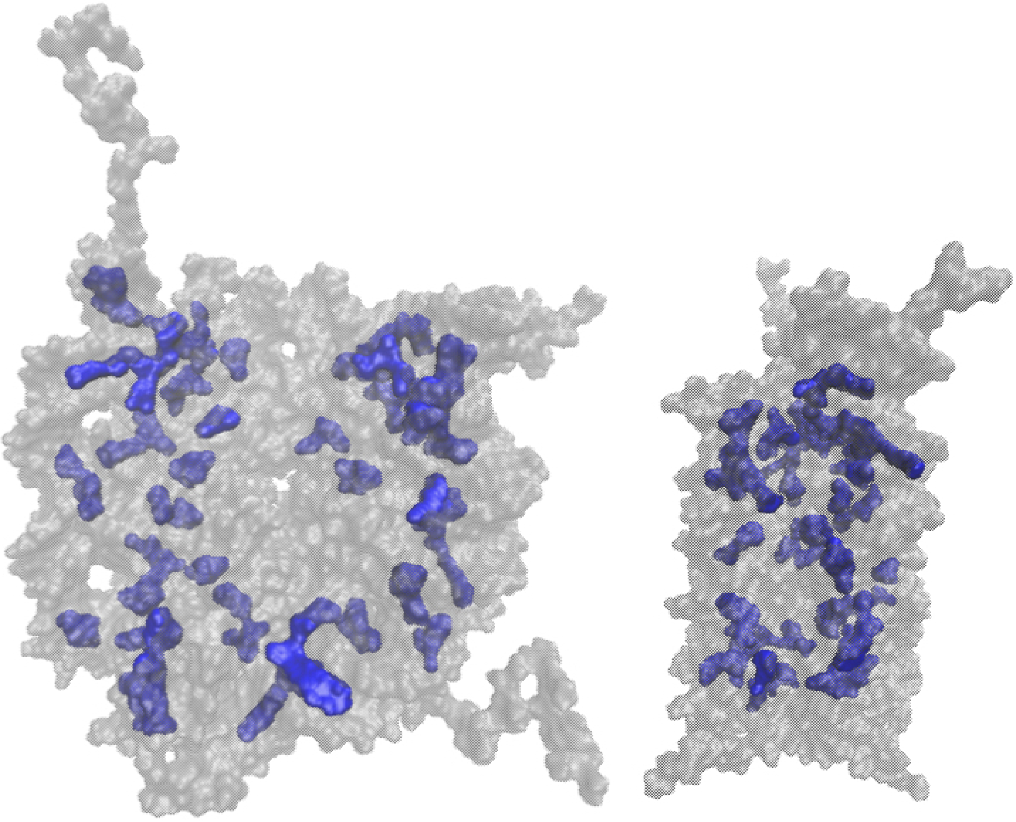


Figure 4.2: Superficial cavities on surface of nucleosome crystal structure front and side (PDB code 1KX5), in blue, constructed with NanoShaper. Solvent Excluded Surface (SES) in grey, generated with NanoShaper interfaced with VMD. A minimum threshold of a capacity of 12 water molecules per cavity was used.

level of specificity. To our knowledge, this is the first application of this technique on nucleosomes. Fig.4.3 summarises the results, displaying both size d and surface potential $\zeta$ measured at different salt concentrations. In the interval between 5 and 250 mM NaCl concentration, the absolute value of $\zeta$ was found to decrease monotonically from 45 ± 7 mV to 17 ± 7 mV when the salt concentration increases. This result is in good agreement with $\zeta$ values extrapolated from electrophoretic mobility and reported in the literature [174], and in agreement with our expectations: elevated ionic concentrations improve the electrostatic screening of the negative charges on the DNA backbone, and therefore the overall potential should, and does, become less negative as counterion concentration increases. In the same NaCl concentration range (5-250 mM), the size measured by DLS varied between 1.9 ± 1.3 and 4.6 ± 1.7 nm. At the highest concentration (i.e., 1 M NaCl) large aggregates were present in the sample ($d$ = 109 ± 23 nm), and the corresponding $\zeta$ was −4.6 ± 1.8 mV. It has been mentioned before in the literature that, at such high NaCl concentrations, nucleosomes have been observed to be unstable, and disassembly is possible [140].

$\zeta$ measurements can be compared to the average value of the electrostatic potential at a suitable distance from the Solvent Excluded Surface (SES) of the solute, obtained, for

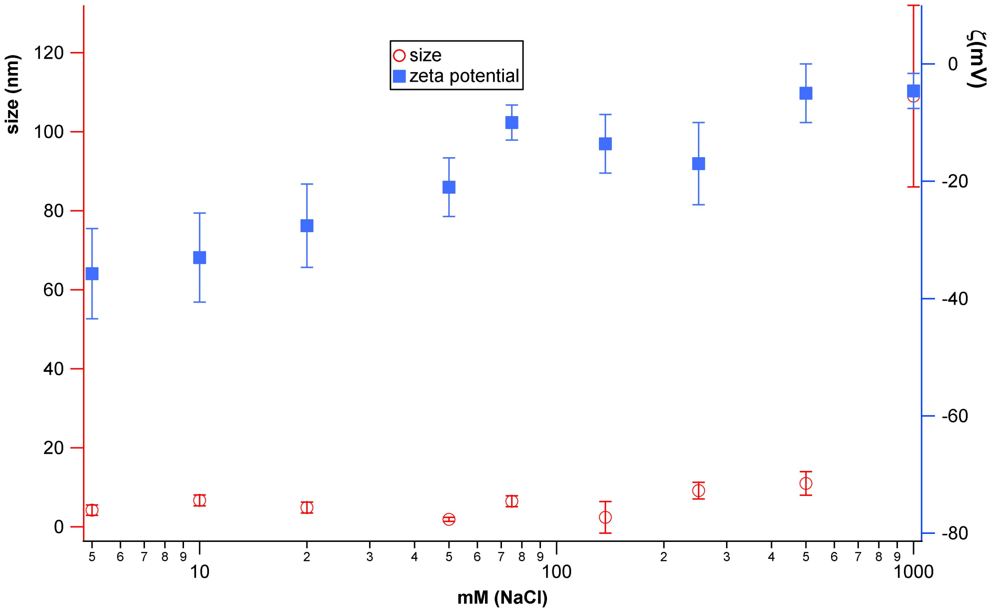


Figure 4.3: $\zeta$ and particle size measured under varying ionic conditions. A clear trend is seen in the decrease of $\zeta$ as a function of NaCl concentration. At high concentration (1M), aggregates are observed in the sample, possibly owing to the disassembly of NCPs.

instance, by solving the PBE. This can be done by exploiting already existing features of PBE solvers, such as DelPhi [131]. We performed measurements of the electrostatic potentials on points generated at the calculated positions of the shear plane, as described in the Methods section. We estimated the position of the shear plane, on which $\zeta$ is defined as a function of the ionic concentration, as follows: we take the value of $\zeta$ at physiological ionic concentration (in our data 145mM), and we find the distance from the NCP at which the potential presents a similar value. In our case, we found this distance to be 5.6Å. Taking into account the dependence of the position of the shear plane, we can write $x_i = \lambda_0\sqrt{I_0/I_i}$, where $x_i$ is the position of the shearing plane, $\lambda_0$ the reference Debye length, $I_0$ = 145mM, and $I_0/I_i$ are the respective ionic strengths. Taking $\lambda_0$ = 8.30Å, the distances for the shear plane we estimate are presented in Table 4.1.

| Ionic Concentration (mM) | Shear layer width (Å) | $\zeta$ Calculated Value (mV) | $\zeta$ Experimental Value (mV) |
|---|---|---|---|
| 5 | 20.8 | -34.51 | -44.9 ±7.7 |
| 10 | 15.3 | -26.31 | -33 ±7.6 |
| 20 | 11.4 | -21.41 | -27.6 ±7.1 |
| 50 | 7.96 | -17.13 | -21 ±5 |
| 137 | 5.60 | -13.80 | -13.6 ±5 |
| 250 | 4.66 | -12.05 | -17 ±7 |

Table 4.1: Experimental values of $\zeta$ with experimental error and calculated values of $\zeta$ in points found on the calculated shear plane in different ionic concentrations.

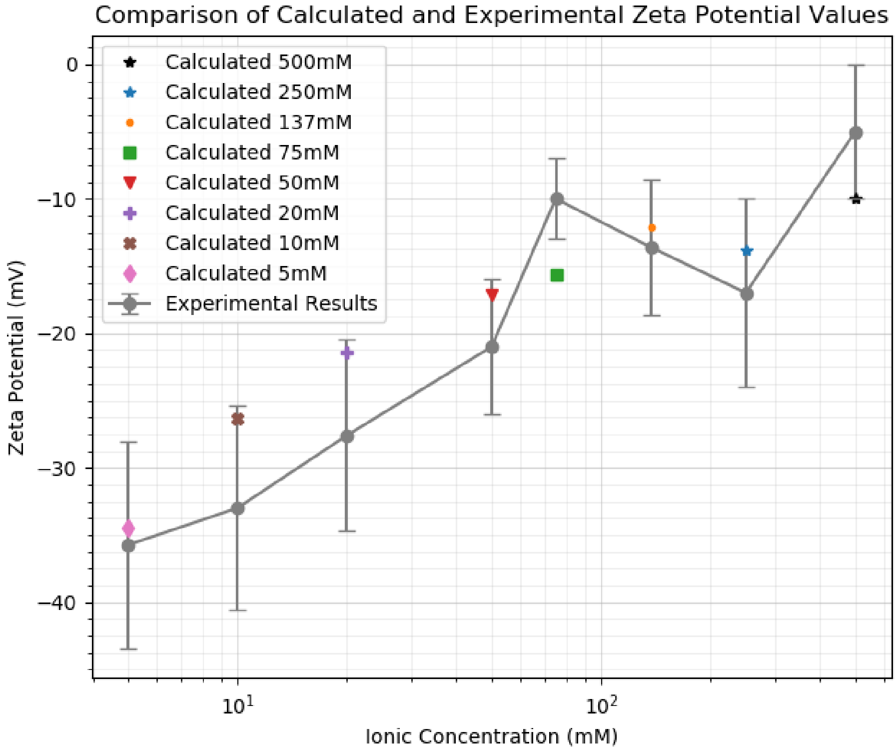


Figure 4.4: Calculated and experimental values of $\zeta$ at varying ionic concentrations. The values of the potential calculated by DelPhi on the estimated positions of the shearing plane for each ionic concentration reproduce the trend of the experimental measurements, and are within the margins of error, with the exception of the value for 75mM NaCl concentration. The computational values slightly overestimate the potential, with respect to the experimental values, an effect that we believe is tied to the approximation used for the estimation of the position of the shearing plane.

## 4.5 Conclusions

In this Chapter, we have followed a quantitative approach to describe the solvation interactions in the nucleosome. We discuss the importance of taking into account nucleosome porosity in chromatin models, instead of treating nucleosomes as solid objects, and we motivate our point by providing quantitative information on the number of cavities and pockets present in the nucleosome crystal structure, and their volume. The high porosity of the nucleosome facilitates the permeation of biologically and electrostatically significant counterions, such as $Na^{+1}$ and $Mg^{+2}$, and enhances electrostatic screening of the negative charge of the DNA backbone.

We discuss the Zeta Potential and Dynamic Light Scattering measurements that we conducted on the nucleosome. We observed a quasi-monotonical relationship between the Zeta Potential and NaCl concentration, and a leap in structure size for very high NaCl concentration. Both of these results are consistent with our expectations and with the literature. We then attempted to reproduce these results with DelPhi calculations. Since the position of the shearing plane cannot be explicitly determined, we estimated it based on the dependence on the square root of the ionic concentration, taking physiological ionic conditions as a benchmark. This allowed us to reproduce the experimental Zeta Potential measurements with good agreement.

What was particularly interesting in the process of conducting these calculations with DelPhi was how much the potential was underestimated if we used the linearised PBE instead of the full non-linear PBE. As we have discussed in previous Chapters, the nucleosome is a highly charged structure and the linearised PBE can only be used in situations where the electrostatic potential is relatively low. Our results clearly show the necessity of using the full non-linear PBE when treating nucleosomes, particularly when interested in measuring electrostatic quantities, such as the potential or the energy, in the vicinity of the nucleosome. However, a single nucleosome is already a quite large structure; 1KX5, the "industry standard" nucleosome crystal structure in chromatin and nucleosome simulations contains 25086 atoms. Therefore, a reasonable question arises: how are we to study the electrostatic interactions between two or more nucleosomes in the chromatin fibre, cases in which the use of the full non-linear PBE becomes prohibitively computationally demanding? This is a matter that we will attempt to face in the next Chapter, in which we present our approach in describing electrostatic interactions between nucleosomes.

CHAPTER 5

# Electrostatic Interactions of Free Nucleosome Pairs

The high electric charge present on nucleosomes, their size and porosity make electrostatics and solvation crucial factors in inter-nucleosome interactions, affecting them at all distances. In this Chapter, we present a comprehensive analysis concerning the electrostatic interactions between nucleosome pairs. We use docked nucleosome pairs generated with the HADDOCK molecular docking software, and nucleosome pairs sampling different distances and rotations in the intermediate distance range.

Our analysis aims to provide a concise description of nucleosome electrostatic interactions in the chromatin fibre, combining information on the physical feasibility of different relative positions of nucleosomes, especially in very tight packing situations. We believe that this data is useful for the development of coarse-grained models of chromatin, and also as a stepping stone for future studies of inter-nucleosome interactions. Our data confirm that after a certain distance, the shape of nucleosomes does not influence the electrostatic energy of their interactions, and the Coulombic energy values follow the monopole approximation. However, certain structures with particularly favourable orientations of the dipole moment do not follow this trend.

In addition to numerical estimates of electrostatic interaction energy of nucleosomes at different relative distances and orientations, obtained within the Poisson-Boltzmann framework, we focus on their approximation by analytical asymptotic expressions, where nucleosomes are approximated as monopoles and dipoles centred in dielectric spheres immersed in an electrolytic solution. We parametrise the analytical expressions using computational calculations of the electrostatic potential between the spheres, and of the interaction energy. We are able to identify a non-linearity region around the nucleosomes; in points outside that region the electrostatic potential can be described by the linearised PBE.

The results and analysis presented in this Chapter are included in a manuscript in preparation at the time of writing.

## 5.1 Introduction

As we discussed in Chapter 2, electrostatics acts as a stabilising force for nucleosome core particles (NCPs). Molecular Dynamics (MD) simulations of the order of µs of single NCPs [20] have shown that the NCPs are stable in dynamics, and do not disassemble even in a solution of 1M NaCl, where they have been experimentally observed to do so [140], indicating that unravelling occurs in longer timescales. Another important feature of NCPs, which points to electrostatics is the "acidic patch", a small groove formed by eight residues on the surface of the histone octamer. The acidic patch constitutes a region of negative charge density on the nucleosome surface, which interacts with DNA-binding proteins and histone tails of adjacent nucleosomes, affecting NCP relative positioning [15, 17, 139].

The electrostatic interactions occurring between charged particles in solution can be described to some extent by the Poisson-Boltzmann Equation (PBE). The PBE combines the electrostatics of continuum media and a mean-field approach for the effect of the ionic environment to describe these interactions [131]. Notably, a PBE-based approach has been used by Tamar Schlick et al. in DiSCO, a coarse-grained model which uses different bead representations for different parts of the nucleosome, such as the histone core and the DNA, incorporating PBE data to parametrise electrostatic interactions between beads [175]. This model has been used, among other applications, to study ionic concentrations of shielding charges around NCPs, and for correlating counterion layer differences to different fibre conformations [106]. The applicability of DiSCO is by no means limited to NCPs: in [36], this model was used to study the topology of oligonucleosome fibres as a function of linker DNA length.

The PBE is also used as a source of information for models aimed at treating larger structures; for example, electrostatic potential distribution derived by PBE has been used to improve the accuracy of a multiscale Generalised Born model, applied to a 40-nucleosome structure [37], even though the direct use of PBE solution via traditional solvers becomes impractical for very large systems. The PBE, although computationally intensive, can however provide fundamental information on the electrostatics of larger structures. It can also provide useful insights about solvent-mediated interactions occurring in a system: in [119], PBE analysis was used, alongside MD, to study the porosity of NCPs.

As we discuss in Chapter 2, analytical solutions to the PBE are difficult to obtain. Some approaches have been made to solve the full non-linear PBE for the case of one dielectric sphere immersed in an electrolyte solution, using for example perturbation expansions [123], or for two particles with known and constant surface potentials, using an iterative scheme [124]. However, it would be remiss not to mention that the PBE is not the only approach for the study of electrostatic interactions between dielectric particles. Approaches have been attempted, for example, through a series expansion of the solution to Gauss's equation [176]. The numerous contributions of Onufriev and collaborators have been of notable importance in this field, particularly their Generalized Born model, optimised to treat biomolecules [177, 178] and able to treat variable dielectric environments [179], and recently proposed a grid-based molecular surface implementation as an alternative to numerical PBE approaches, applied to the calculation of solvation and binding free energies in small protein-ligand complexes [180]. Furthermore, Onufriev and collaborators have used the Kirkwood solution to the Poisson equation to derive a closed-form analytical approximation [180], validated on realistic molecular surfaces [181].

Derbenev and Filippov have published several works in which they explore the electrostatic interactions between charged polarisable particles immersed in equilibrium plasma or in an electrolyte solution, mainly focusing on the electrostatic force [128, 182–185]. We expand upon their analysis to include cases in which azimuthal symmetry is broken. Our formalism enables us to find exact closed-form expressions for the total electrostatic energy of particles under the linear PBE validity assumption, and to explicitly quantify the doubly screened corrections to the leading Coulombic terms (single screening) of the mutual interaction energy between particle centres. We also investigate dipole-dipole and charge-dipole interactions, while for the charge-charge interaction of small weakly screened particles our results recover the expressions previously derived by Fisher et al. [186]. The analytical derivation, not included in this thesis, was conducted by Dr. Sergii Siryk (manuscript in preparation).

In this Chapter, we present a study of the electrostatic interactions between nucleosomes at short and intermediate distances. We perform PBE calculations on nucleosome pairs at different relative orientations and distances, using the DelPhi solver. At short distances, we provide electrostatic energy values on pairs of docked NCP pair configurations, generated using the HADDOCK molecular docking software [187], and we present the most energetically favourable configurations. At intermediate distances we used nucleosomes with attached linker DNA segments, and we show that the electrostatic energies follow a general trend, unless there are particularly favourable orientations of the linker DNA, and therefore the dipole moment. We compare our numerical results with the leading terms of the analytical asymptotic expression of the interaction energies of two dielectric spheres immersed in a dielectric solution, containing a monopole and a dipole. These expressions for the energy were derived using the linearised PBE. We show that charged dielectric particles are surrounded by a non-linearity layer, beyond which the electrostatic potential drops enough for the linearised PBE to be adequate. Our analysis constitutes a first building block towards a description of electrostatic interactions between nucleosomes in the chromatin fibre.

## 5.2 Analytical Methods: A Linearised Poisson-Boltzmann Model

Let us consider a general system consisting of two dielectric spherical particles $i, j$, with dielectric constants $\varepsilon_i$, $\varepsilon_j$ and radii $a_i$, $a_j$, separated by a distance $R$ between their centres $\mathbf{x}_i$ and $\mathbf{x}_j$, $R > a_i + a_j$. Without loss of generality we will assume henceforth that $i \in \{1, 2\}$ while $j = 3 - i$ stands for the second particle. The particles are suspended in an electrolyte solution with dielectric constant $\varepsilon_{\mathrm{m}}$ and Debye length $\kappa^{-1}$. The corresponding graphical representation of the system is shown in Figure 5.1.

The electrostatic potential $\Phi_{\text{in},i}$ inside the $i$-th particle ($i = 1, 2$) satisfies the Poisson equation [188]

$$\Delta\Phi_{\text{in},i}(\mathbf{x}) = -\frac{\rho_i(\mathbf{x})}{\varepsilon_i\varepsilon_0} \quad \text{as} \quad r_i < a_i, \tag{5.1}$$

where $r_i$ is the radial coordinate of $\mathbf{x} \in \mathbb{R}^3$ measured from the centre $\mathbf{x}_i$ of the $i$-th sphere (so that, $r_i = \|\mathbf{r}_i\|$, $\mathbf{r}_i = \mathbf{x} - \mathbf{x}_i$, see Fig. 5.1), and $\rho_i(\mathbf{x})$ denotes the charge density inside the $i$-th sphere. Simultaneously, within the Debye-Hückel (DH) approximation, the potential $\Phi_{\text{out},i}$ in the surrounding medium caused by the presence of the $i$-th particle satisfies the linearised Poisson-Boltzmann (PB) equation:

$$\Delta\Phi_{\text{out},i}(\mathbf{x}) - \kappa^2\Phi_{\text{out},i}(\mathbf{x}) = 0 \quad \text{as} \quad r_i > a_i. \tag{5.2}$$

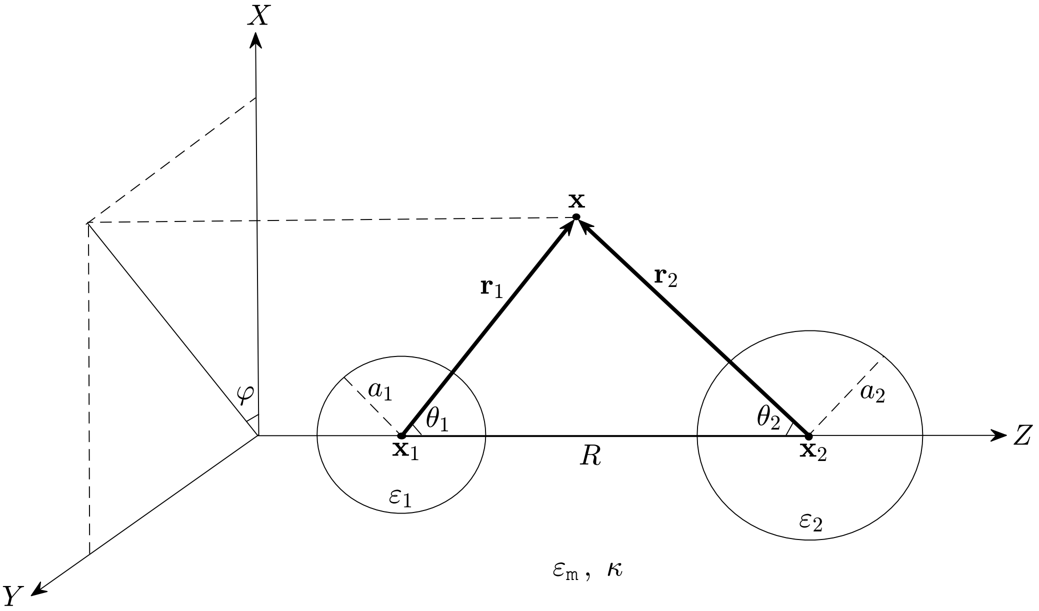


Figure 5.1: Depiction of the general geometry of the system under consideration: two dielectric spheres of dielectric constants $\varepsilon_1$, $\varepsilon_2$ and radii $a_1$, $a_1$, separated by a distance $R$ between their centres $\mathbf{x}_1$ and $\mathbf{x}_2$, $R > a_1 + a_2$. The spheres are immersed in a medium of dielectric constant $\epsilon_m$. The electrostatic potential is calculated at point $\mathbf{x}$.

Due to the superposition principle the self-consistent total electrostatic potential $\Phi(\mathbf{x})$ of the whole system is then [128, 182–185]:

$$\Phi(\mathbf{x}) = \left[\begin{array}{ll} \Phi_{\text{in},i}(\mathbf{x}), & r_i < a_i, \\ \Phi_{\text{out}}(\mathbf{x}) := \Phi_{\text{out},i}(\mathbf{x}) + \Phi_{\text{out},j}(\mathbf{x}), & \text{outside particles,} \end{array}\right. \tag{5.3}$$

subject to the following general boundary conditions [188]:

$$\Phi_{\text{in},i}\big|_{r_i=a_i} = \Phi_{\text{out}}\big|_{r_i=a_i}, \qquad \varepsilon_i\,(\mathbf{n}_i\cdot\nabla\Phi_{\text{in},i})\big|_{r_i=a_i-0} - \varepsilon_{\text{m}}\,(\mathbf{n}_i\cdot\nabla\Phi_{\text{out}})\big|_{r_i=a_i+0} = \sigma_i/\varepsilon_0, \qquad i = 1, 2, \tag{5.4}$$

where $\mathbf{n}_i$ is the unit normal vector and $\sigma_i$ is a charge density distribution on the surface of the $i$-th sphere. The notation $r_i = a_i \pm 0$ indicates approaching the surface of the sphere from the interior or the exterior. In the cases that are of interest to us, we assume point charges/dipoles centred in the spherical particles (5.5), and therefore $\sigma_i = 0$ [186].

In the analysis that follows, we will treat cases where the charge density $\rho_i(\mathbf{x})$ represents point monopoles or dipoles located in the particle's centre. We express our solutions in the form (5.1)-(5.2)

$$\Phi_{\mathrm{in},i} = \hat{\mathit{\Phi}}_{\mathrm{in},i} + \tilde{\Phi}_{\mathrm{in},i}, \qquad \Phi_{\mathrm{out},i} = \hat{\mathit{\Phi}}_{\mathrm{out},i} + \tilde{\Phi}_{\mathrm{out},i}, \tag{5.5}$$

where $\hat{\mathit{\Phi}}_{\mathrm{in},i}$ and $\hat{\mathit{\Phi}}_{\mathrm{out},i}$ are particular solutions to (5.1) and (5.2), respectively, which represent the standard Coulombic and screened Coulombic (DH) potentials in infinite space for a point charge/dipole. Explicitly expressing these contributions $\hat{\mathit{\Phi}}_{\mathrm{in},i}$, $\hat{\mathit{\Phi}}_{\mathrm{out},i}$ in the general solutions of equations (5.1)-(5.2), coupled with boundary conditions (5.4), provides a convenient way to subtract the self-energy contributions from the expression for the total electrostatic energy of a system, as well as to treat $\tilde{\Phi}_{\mathrm{in},i}$, $\tilde{\Phi}_{\mathrm{out},i}$ as perturbations added to the corresponding infinite space point-charge/dipole potentials $\hat{\mathit{\Phi}}_{\mathrm{in},i}$, $\hat{\mathit{\Phi}}_{\mathrm{out},i}$ to account for the finiteness of particle sizes and their mutual influence. Substituting (5.5) in (5.1)-(5.2), we arrive at the following homogeneous equations:

$$\begin{cases} \Delta\tilde{\Phi}_{\mathrm{in},i} = 0, \quad \text{as} \quad r_i < a_i, \quad i = 1, 2, \\ \Delta\tilde{\Phi}_{\mathrm{out},i} - \kappa^2\tilde{\Phi}_{\mathrm{out},i} = 0, \quad \text{as} \quad r_i > a_i, \quad i = 1, 2. \end{cases} \tag{5.6}$$

Since the linearized PBE (5.2) is a Helmholtz-type equation with $\kappa \neq 0$, it cannot be solved using a bi-spherical coordinate system through a separation of variables [185]. We therefore use two spherical coordinate systems with their origins associated with centres of the particles $i$ and $j$, as shown on Figure 5.1. Without loss of generality, we can assume that the spheres' centres lie on the Cartesian axis $Z$, while the axes $X$ and $Y$ are fixed. Then, for $i = 1, 2$ the $i$-th spherical coordinate system operates with coordinates $(r_i, \theta_i, \varphi)$, where $\theta_i$, $\varphi$ are the polar and azimuthal angles, respectively, and $r_i$ is the radial distance. Note that a point $\mathbf{x}$ between the two spheres in Fig. 5.1 has different $r_i$, $\theta_i$ and $r_j$, $\theta_j$, but $\varphi$ is the same in both spherical coordinate systems. That is to say, azimuthal symmetry is conserved.

Let us introduce a new dimensionless radial coordinate $\tilde{r}_i := \kappa r_i$ and also denote $\tilde{a}_i := \kappa a_i$, $i = 1, 2$. With this scaling, equations (5.6) can be rewritten in the following form:

$$\begin{cases} \Delta_{\tilde{r}_i}\tilde{\Phi}_{\mathrm{in},i}(\tilde{r}_i, \theta_i, \varphi) = 0, \quad \text{as} \quad \tilde{r}_i < \tilde{a}_i, \quad i = 1, 2, \\ \Delta_{\tilde{r}_i}\tilde{\Phi}_{\mathrm{out},i}(\tilde{r}_i, \theta_i, \varphi) - \tilde{\Phi}_{\mathrm{out},i}(\tilde{r}_i, \theta_i, \varphi) = 0, \quad \text{as} \quad \tilde{r}_i > \tilde{a}_i, \quad i = 1, 2, \end{cases} \tag{5.7}$$

where $\Delta_{\tilde{r}_i}$ denotes the Laplace operator with $\tilde{r}_i$ as the radial spherical coordinate. This translates to the following boundary conditions:

$$\begin{cases} \tilde{\Phi}_{\mathrm{in},i}\Big|_{\tilde{r}_i=\tilde{a}_i} - \tilde{\Phi}_{\mathrm{out},i}\Big|_{\tilde{r}_i=\tilde{a}_i} - \tilde{\Phi}_{\mathrm{out},j}\Big|_{\tilde{r}_i=\tilde{a}_i} = \hat{\varPhi}_{\mathrm{out},i}\Big|_{\tilde{r}_i=\tilde{a}_i} + \hat{\varPhi}_{\mathrm{out},j}\Big|_{\tilde{r}_i=\tilde{a}_i} - \hat{\varPhi}_{\mathrm{in},i}\Big|_{\tilde{r}_i=\tilde{a}_i}, \\ \varepsilon_i \dfrac{\partial \tilde{\Phi}_{\mathrm{in},i}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i-0} - \varepsilon_{\mathrm{m}} \dfrac{\partial \tilde{\Phi}_{\mathrm{out},i}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i+0} - \varepsilon_{\mathrm{m}} \dfrac{\partial \tilde{\Phi}_{\mathrm{out},j}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i+0} = \\ -\varepsilon_i \dfrac{\partial \hat{\varPhi}_{\mathrm{in},i}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i-0} + \varepsilon_{\mathrm{m}} \dfrac{\partial \hat{\varPhi}_{\mathrm{out},i}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i+0} + \varepsilon_{\mathrm{m}} \dfrac{\partial \hat{\varPhi}_{\mathrm{out},j}}{\partial \tilde{r}_i}\Bigg|_{\tilde{r}_i=\tilde{a}_i+0}, \end{cases} \tag{5.8}$$

where $i = 1, 2$ and $j = 3 - i$. General solutions to equations (5.7), such that $|\tilde{\Phi}_{\mathrm{in},i}| < \infty$ as $\tilde{r}_i \longrightarrow 0$, and $\tilde{\Phi}_{\mathrm{out},i} \longrightarrow 0$ as $\tilde{r}_i \longrightarrow +\infty$ can be expanded in modified Bessel functions of the 2nd kind $K_{n+1/2}(\cdot)$, so called Macdonald functions, [189] and complex-valued spherical harmonics $Y_{nm}$ [190] as follows:

$$\tilde{\Phi}_{\mathrm{in},i}(\tilde{r}_i, \theta_i, \varphi) = \sum_{n=0}^{+\infty} \tilde{r}_i^n \sum_{m=-n}^{n} \mathfrak{A}_{nm,i} Y_{nm}(\theta_i, \varphi), \tag{5.9}$$

$$\tilde{\Phi}_{\mathrm{out},i}(\tilde{r}_i, \theta_i, \varphi) = \sum_{n=0}^{+\infty} \frac{K_{n+1/2}(\tilde{r}_i)}{\sqrt{\tilde{r}_i}} \sum_{m=-n}^{n} \mathfrak{B}_{nm,i} Y_{nm}(\theta_i, \varphi) \tag{5.10}$$

where $\mathfrak{A}_{nm,i}$ and $\mathfrak{B}_{nm,i}$ are expansion coefficients. An alternative is to represent spherical harmonics through associated Legendre polynomials $P_n^m(x) = (-1)^m (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_n(x)$, where $P_n(x)$ is the $n$-th standard Legendre polynomial [190]. For real values this can be written as follows:

$$\tilde{\Phi}_{\mathrm{in},i}(\tilde{r}_i, \theta_i, \varphi) = \sum_{n=0}^{+\infty} \tilde{r}_i^n \sum_{m=0}^{n} \big(L_{nm,i} \cos(m\varphi) + M_{nm,i} \sin(m\varphi)\big) P_n^m(\mu_i), \tag{5.11}$$

$$\tilde{\Phi}_{\mathrm{out},i}(\tilde{r}_i, \theta_i, \varphi) = \sum_{n=0}^{+\infty} \frac{K_{n+1/2}(\tilde{r}_i)}{\sqrt{\tilde{r}_i}} \sum_{m=0}^{n} \big(G_{nm,i} \cos(m\varphi) + H_{nm,i} \sin(m\varphi)\big) P_n^m(\mu_i), \tag{5.12}$$

where $\mu_i := \cos\theta_i$, $i = 1, 2$, with real expansion coefficients $L_{nm,i}$, $M_{nm,i}$, $G_{nm,i}$, $H_{nm,i}$ to be determined from the boundary conditions (5.8). Note that $P_n^0(x) = P_n(x)$ for $\forall n \geq 0$, and we define $P_n^m(x)$ with the Condon-Shortley factor $(-1)^m$ included. The expressions for all the coefficients can be found in Appendix B.

## 5.3 Numerical Methods

### 5.3.1 DelPhi Simulations: Calculations of Electrostatic Interaction Energy between NCP pairs

#### NCP Structures

The original PDB files we used are from publicly available data [137], used in the work of Shaytan et al [20]. In summary, we have used the "FN model" (file name `only_nucl_init.pdb`, following the naming scheme in [137]). The structure is composed of the PDB code 1KX5 crystal structure [90] and two 20bp-long B-DNA segments of linker DNA. A 1μs MD trajectory for these structures is also publicly available. We subsequently performed k-medoids clustering analysis on this structure, based on RMSD distance, as done in [191], with the software BiKi Life Sciences [149]. This particular variant does not rely on the assumption of an underlying vector space and does not require to perform the superposition of all the frames with respect to a common reference structure. We selected the medoid representing the largest cluster, cluster number 2, in order to have the most representative structure. The resulting structure, which we will be referring to as `medoid2` is represented in Fig.5.2. We then generated PQR files, i.e. PDB format files where atomic charge and radius replace beta-factor and occupancy, from the medoid PDB files via the PDB2PQR web server[156], in order to use them as input files for the DelPhi PBE solver, which requires atomic radii and charges. We used the CHARMM force field for the PDB2PQR conversion, as it was the one used by Shaytan et al. in the MD simulations. After file conversion with PDB2PQR, we had to perform some manual changes to the DNA charge parameters due to nomenclature issues causing only partial recognition of the DNA structure.

In order to evaluate inter-nucleosome electrostatic interactions between nucleosomes with different structural characteristics, we also performed simulations on pairs of 3AFA NCPs with a single 25-bp linker DNA piece attached on each NCP. PQR files were generated using tleap, a program included in the Ambertools package. For more information on the nucleosome crystal structures used in this work see Table 3.3.

### NCP Pair Generation

Nucleosome pairs at different relative distances and orientations were constructed using custom Python scripts in the following manner:

- Two identical PQR files are parsed.
- One of them is kept fixed and the other is rotated, first with respect to the polar and then with respect to the azimuthal angle. The angles sampled are in the following ranges: $\theta$ (polar angle) from 0 to $\pi$ and $\phi$ (azimuthal angle) from 0 to $2\pi$.
- The rotated nucleosomes are saved, and then each rotated structure is translated by increments of x/N, where x is a maximum imposed by the user on the distance between the nucleosomes. For the calculations presented here, we have selected the maximum distance between structures to be approximately 5 Debye lengths in water in physiological NaCl concentration (147 mM), and N to be 15.
- Finally, the nucleosome pairs are saved in a separate PQR files. Examples of the structures on which the DelPhi calculations were performed on are presented in Fig.5.6.

In order to take into consideration only physically feasible nucleosome pairs, we then removed structures containing steric clashes between the two nucleosomes from our data set using NanoShaper [135]. We calculated the total surface area of each nucleosome pair and compared it to twice the surface area of a single nucleosome. Precise distances between the geometric centres of the nucleosomes were calculated by DelPhi, for added accuracy. For the data presented here, we generated a total of medoid2 pair 2115 structures and 621 3AFA pair structures.

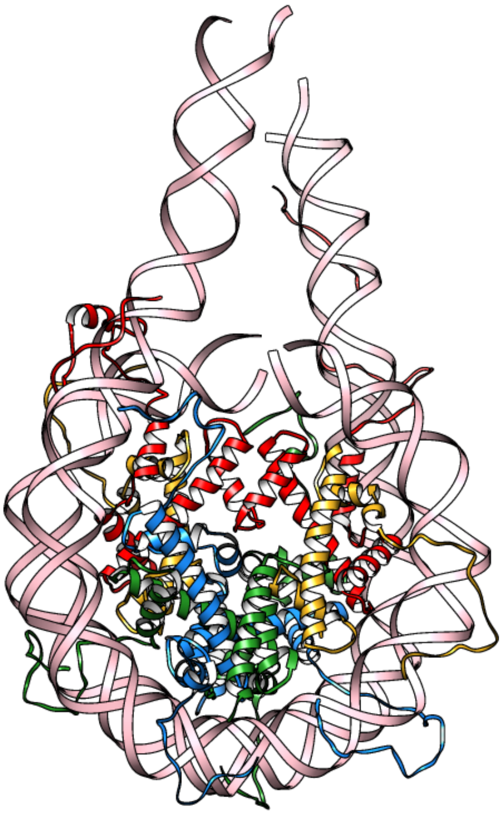


Figure 5.2: To construct the fundamental structure (medoid2) used in our NCP pair construction we performed k-medoids clustering analysis based on RMSD distance on an original PDF file from [137]. The structure is composed of the PDB code 1KX5 crystal structure [90] and two 20bp-long B-DNA segments of linker DNA. Note: the nucleosomal DNA is continuous along the dyad; the current visualisation is a result of a change in chain naming in the PDB file.

The Python scripts that we have developed use the Bio.PDB package included in Biopython [192] to parse and manipulate molecular structures. However, since the parsing and input/output of PQR files was not supported by Biopython, we extended the already present PDB file parser and input/output methods to support PQR files. We addressed these changes in a pull request (`https://github.com/biopython/biopython/pull/2338`) which has been accepted and merged with the Biopython source code, and the implemented support for parsing, reading and writing PQR files is included in the current release of Biopython. Our contribution is part of Biopython as of release 1.77.

### 5.3.2 HADDOCK Docked Nucleosome Structures

Two NCPs were docked with an 70Å restraint between the DNA ends and tight (HADDOCK default) centre of mass restraints. We used the final minimised structures, which have been equilibrated to take into account solvation in water. The total number of structures that we performed DelPhi calculations on was 400. The docked structures are two copies of PDB code 2PYO. Both 2PYO and 1KX5 contain DNA sequences from *Homo Sapiens*, but 2PYO contains histones from *Drosophila Melanogaster*, while the histones in 1KX5 are from *Xenopus Laevis*. In order to check whether the two structures presented a degree of structural similarity high enough to allow for a comparison of subsequent measurements, we compared them using the BLAST alignment web server [193]. Core histones H3 and H4 differ only in one amino acid each, but the two H2A sequences score 90.57% similarity and the H2B sequences score 95.60%. The total number of different amino acids between the two structures is 16, 15 of which are in H2B and H2A histones. Four of these different amino acids are found in the histone tails, and the rest are histone core residues of very minor importance in NCP interactions [118]. The total electric charge and dipole moment of medoid2 and 2PYO were calculated by DelPhi, and found to be: -226e and -216e for the monopole electric charges, while the dipole moments are greatly different, due to the presence of the linker DNA segments in medoid2: approximately -92.23eÅ for 2PYo and -1725.12eÅ for medoid2. To calculate the dipole moments we selected the system geometric centre as the pole. For more information on the nucleosome crystal structures used in this work see Table 3.3.

In order to provide consistent data between the two types of input structures, rotation angles and distances between the centroids of the docked nucleosomes were extracted using Python scripts, in the following manner: Taking medoid2 as a reference structure, we superimpose one of the two docked NCPs on it (model 1 in the HADDOCK PBD files), in order to establish a common reference frame between the structures used in the measurements of the electrostatic field at close and intermediate distances. The Superimposer module of the Bio.PDB package is used to perform this step. We then calculate the translation vector between the centroids of the two HADDOCK NCPs, and the relevant rotation angles (azimuth and polar), implementing the Kabsch algorithm with the rmsd module of PyPI [194].

#### DelPhi Simulation Conditions and Parameters

Electrostatic energies were computed using the DelPhi non-linear PBE solver, interfaced with NanoShaper [135]. The solvent was assigned a dielectric value of 2, and the solute a value of 80. A cubic box of grid spacing 2 and `perfil` value of 80% (percentage of box filled by solute). Salt concentration was set at 145mM. Atomic electric charge values and atomic radii were taken from the PQR files. Connolly surface probe radius was 1.4. Other parameters assumed default values. For all simulations we have used the 5.1 patched version of DelPhi as distributed by the CONCEPT Lab in IIT (Dr. Walter Rocchia).

### 5.3.3 DelPhi Simulations on Dielectric Spherical Particles

DelPhi can generate geometric objects and charge distributions on which to perform simulations by including the keyword `insobj` in the parameter file (prm file). When running DelPhi using this prm file, the user is asked to input some information regarding the charge distributions that they want to construct, interactively via the command line. The user can choose the type of charge distribution, its shape, the value of the charge and the dielectric constant, the coordinates of the geometric centre of the distribution and add more distributions. In order to generate the desired charge distributions we used a prm file with scale=1, perfil=95 and no salt concentration. We assimilated the characteristics of the nucleosome by constructing two identical spheres of charge -144 and radius 60Å, corresponding to the electric charge and mean radius of 1KX5. The spheres have dielectric constants equal to 2 and are immersed in solvent of dielectric constant 80. The geometric centre of the first sphere is placed at the origin of the axes, while the centre of second sphere is on the Z axis and its distance from the first sphere varies in a range from 160 to 250Å. For the electrostatic potential calculations we used scale=3 and perfil=65. The rest of the parameters assumed default values.

## 5.4 Numerical Study of the Electrostatic Interaction Energy of Nucleosome Pairs

### 5.4.1 Nucleosome Electrostatics at Close Distances: Electrostatic Interaction Energy of Docked Nucleosome Pairs

To enrich our statistics on electrostatic interactions of NCP pairs at close distances, we performed DelPhi PBE calculations on docked nucleosome structures, generated with HADDOCK. Out of the 400 structures of which we calculated the electrostatic interaction energy, we represent the 12 energetically most favourable configurations in Figs.5.3 – 5.5. The total electrostatic interaction energy we consider for the classification of these structures is the sum of the Coulombic and the Reaction Field energies as calculated by DelPhi. For visualisations purposes, we have omitted conformations that presented significant structural overlap. The structures are listed in Table 5.1, along with their geometric characteristics (distance between NCP centres, relative rotation angles as calculated by the Kabsch algorithm, electrostatic interaction energy values). The naming scheme used for the structures is the one used by HADDOCK. We note here that the docking procedure performed by HADDOCK is not only based on continuum electrostatics; rather, this process includes equilibration in explicit solvent using a molecular mechanics force field.

### 5.4.2 Nucleosome Electrostatics at Intermediate Distances

In order to study the electrostatic energies of nucleosome pairs and their dependence on distances between the centres of the nucleosomes and rotations, we generated nucleosome pairs as described in the Numerical Methods section, and calculated the Coulombic Interaction Energy and the Reaction Field Energy using DelPhi. In our calculations, the Self-Reaction Field Energy, as defined in Chapter 2, of a single nucleosome was subtracted from the values, in order to observe only the contribution of the interactions in the nucleosome pair, and the trends of those interactions. We observe that we have data on very few structures at close distances; this is due to the steric clash criterion that we have

| Structure | Distance (Å) | $\theta$ (rad) | $\phi$ (rad) | Electrostatic Energy (kT) |
|---|---|---|---|---|
| 213 | 7.164 | -0.708 | 0.061 | -907209.111 |
| 90 | 7.353 | -1.404 | 0.095 | -907044.998 |
| 105 | 7.309 | 0.112 | 0.041 | -906910.640 |
| 120 | 7.221 | -0.445 | 0.018 | -906464.964 |
| 319 | 7.385 | -0.892 | 0.052 | -906452.998 |
| 262 | 7.503 | -1.449 | -0.034 | -906394.677 |
| 190 | 7.543 | 0.757 | 0.447 | -906382.876 |
| 188 | 7.334 | -0.977 | 0.424 | -906300.913 |
| 112 | 7.538 | 0.913 | 0.128 | -906265.329 |
| 304 | 7.298 | 0.984 | 0.036 | -906264.065 |
| 307 | 7.323 | -1.38 | 0.042 | -906119.669 |
| 390 | 7.256 | -0.844 | -0.041 | -906067.399 |

Table 5.1: Energetically favourable docked nucleosome structures. Relative rotation angles are calculated using the Kabsch algorithm and electrostatic interaction energy is calculated with DelPhi.

applied. Given that we have generated the same number of structures at all distances, and structures at closer distances are more likely to present steric clashes, more structures are excluded when the distance between nucleosome centres is smaller. However, this is a source of information on chromatin topology as well: since we generate the NCP pairs in a controlled manner, we have information on which combinations of NCP distances and relative rotations are feasible in the chromatin fibre.

Our results, shown in Fig.5.7 for the medoid2 pairs and in Fig.5.8 for the 3AFA with linker DNA pairs, show that the Coulombic interaction energy coincides with the monopole approximation as we move towards larger inter-NCP distances, while, as we expected, in closer distances relative rotations and the inherent three-dimensionality and characteristic geometry of the structures become very significant in inter-NCP electrostatic interactions, as we can see in Fig.5.7. Specifically, the Coulombic Interaction Energy follows closely the asymptotic trend of the monopole approximation for inter-NCP distances larger than approximately 175Å. The Reaction field energy becomes less relevant as the distance between the nucleosomes becomes larger, as the effect of the induced surface charges becomes less and less significant. As this energy is negative, it progressively becomes less so as its absolute value decreases.

As it can be seen by inspecting Fig. 5.7, the interactions energies of certain medoid2 pairs do not follow the general trend. The interactions energies of these structures are significantly lower than the mean energy in that range, which points to them being particularly favourable configurations, as lower energy indicates a more stable configuration. By visually inspecting these structures we see that the linker DNA segments are found in particular configurations, either (anti)parallel or perpendicular. Bearing in mind that the linker DNA segments bear a high negative electric charge, because of the charges of the phosphate groups present on the DNA backbone, we deduce that the favourable energies that we are calculating for these configurations are a result of particular relative

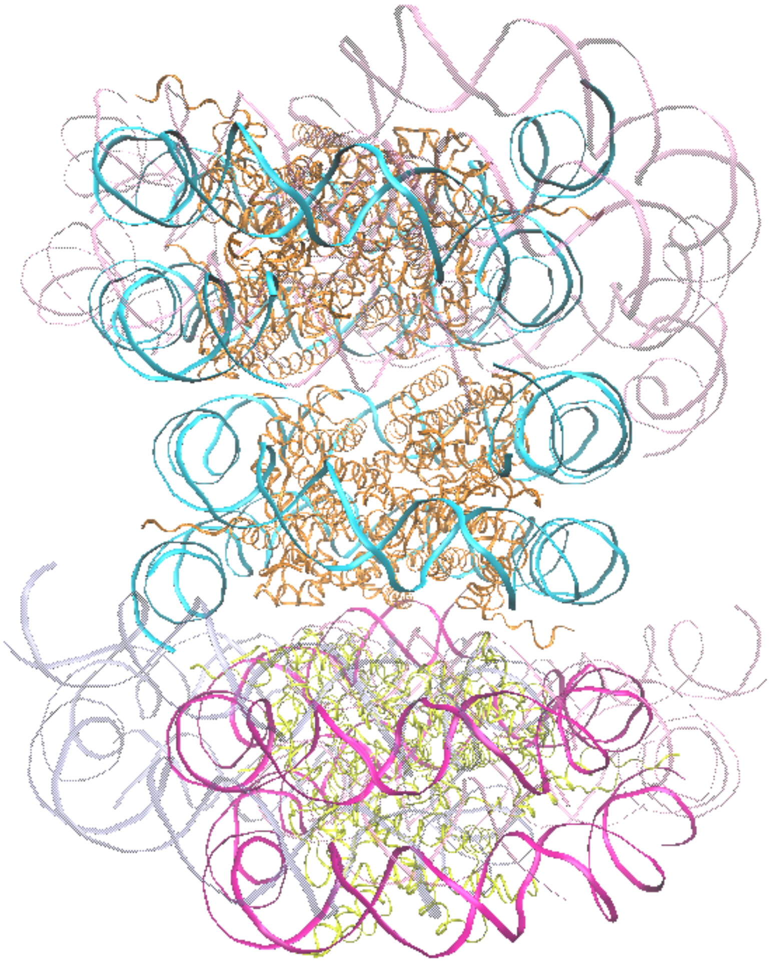


Figure 5.3: Energetically favourable docked nucleosome structures: following the HADDOCK naming scheme, structures 213 and 190 are opaque, while the backbone of the DNA in relatively overlapping structures 90, 390, 304, and 112 is pictured using a semi-transparent representation. Figure created using VMD.

orientations of the dipole moments of each structure. On the other hand, for intermediate angles between the two dipole moment vectors, the monopole properties of the structures prevail, and indeed the electrostatic interaction energies follow the monopole approximation. Notice the absence of such "outliers" in the values of the electrostatic energy for the measurements performed on the 3AFA structure (Fig.5.8). Here there is only one linker DNA segment and the histone tails are absent, overall decreasing the dipole moment of the structure (-1725.13eÅ for medoid2 versus -298.65eÅ for the 3AFA/linker DNA complex).

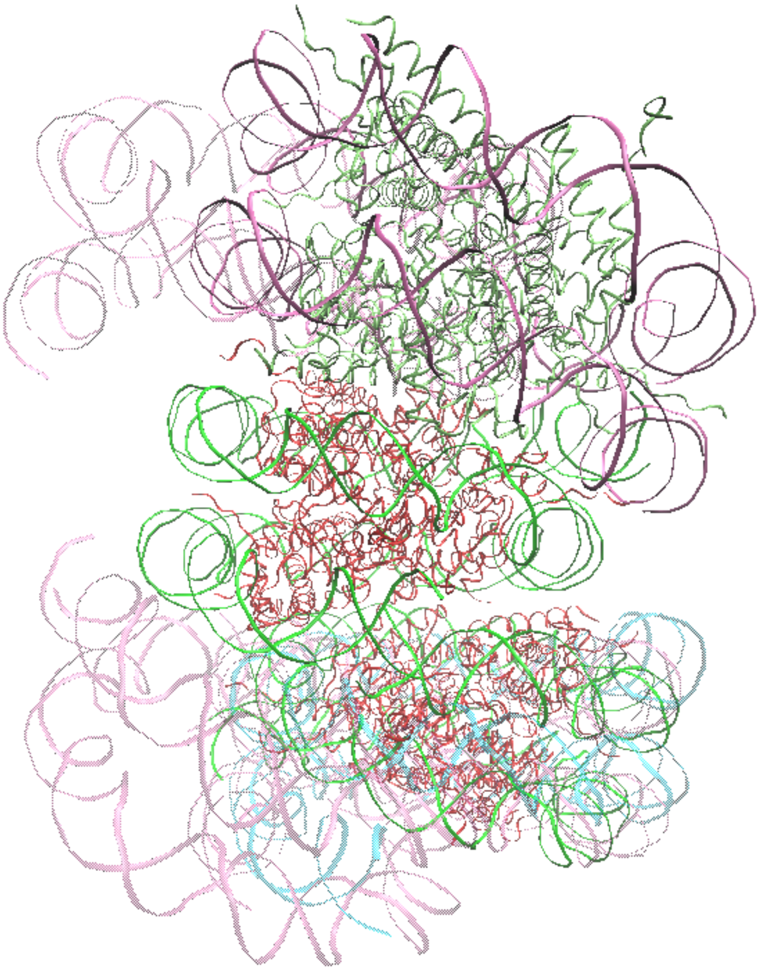


Figure 5.4: Energetically favourable docked nucleosome structures: following the HADDOCK naming scheme, structures 304 and 262 are opaque, while the backbone of the DNA in relatively overlapping structures 120, 112, 307, and 390 is pictured using a semi-transparent representation. Figure created using VMD.

## 5.5 Electrostatic Interactions of Charged Dielectric Spheres

In our study of the electrostatic interactions between two nucleosomes at large distances, we employ a simplistic representation of nucleosomes as spherical dielectric particles, with monopole charges (or, for added detail, dipoles) placed at their centre. We choose this approximation as, due to the large size of the overall system, numerical solution of PBE becomes impractical at large distances. As we discussed in Chapters 2 and 4, for highly charged systems, such as nucleosomes, the non-linear PBE is more accurate. However, after a certain distance the linearised approximation and the full non-linear PBE exhibit the same behaviour. The region in which the non-linear regime holds, which we term "non-linear corona", depends on the charge of the particle, and the Debye length. The non-linear corona is delimited by the potential surface around the spheres where the

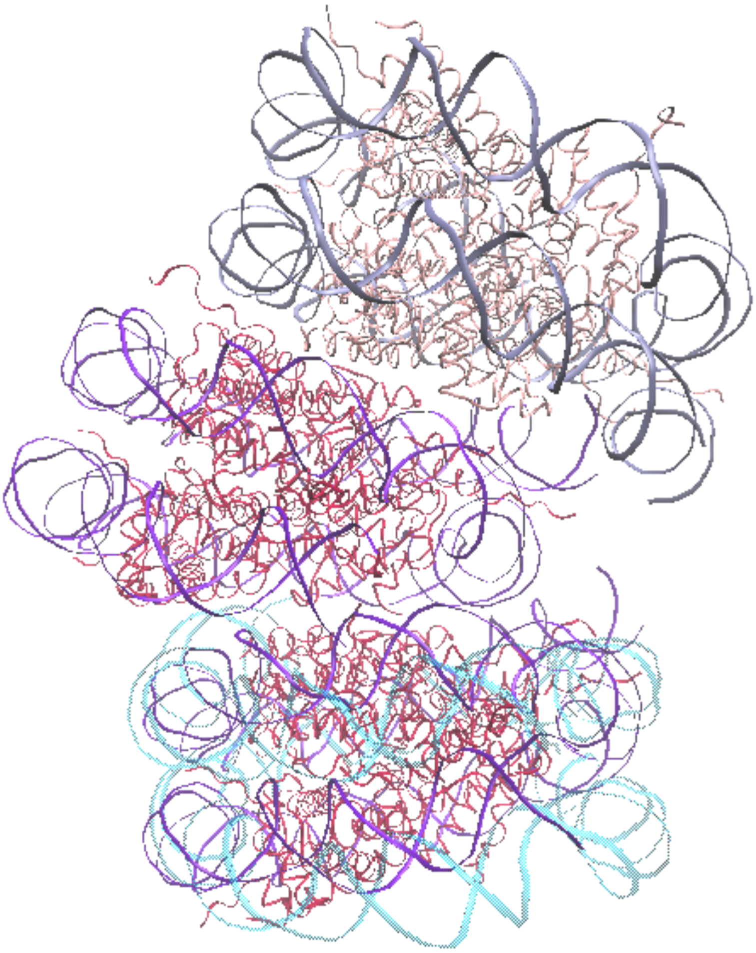
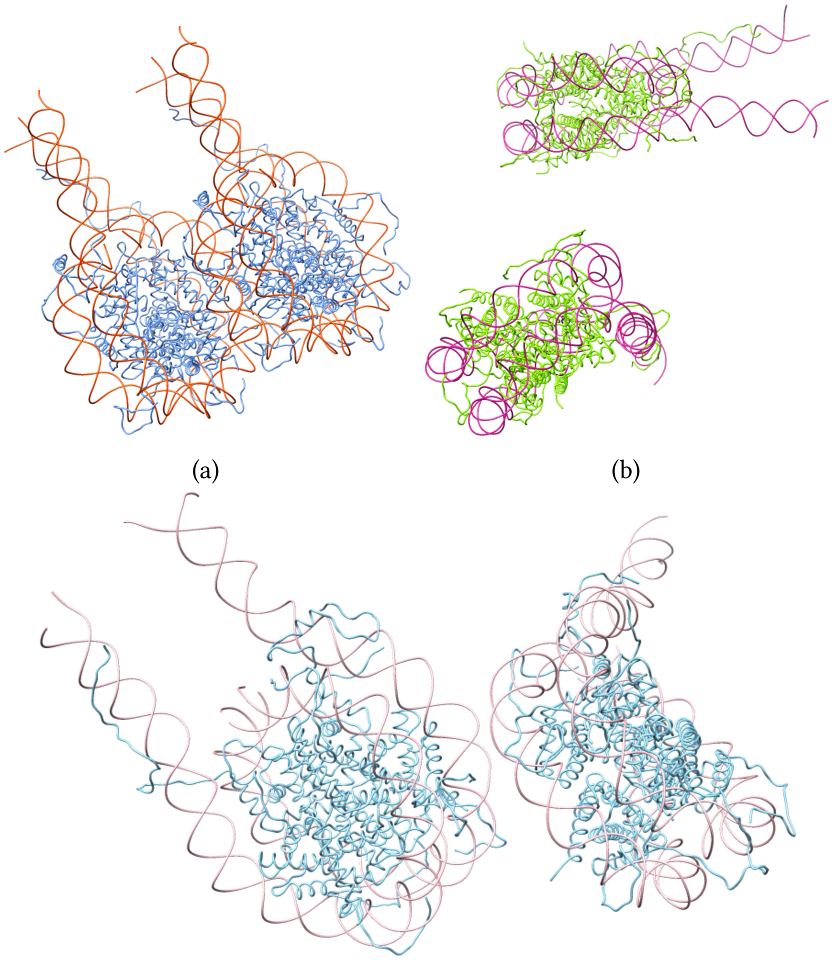


Figure 5.5: Energetically favourable docked nucleosome structures: following the HADDOCK naming scheme, structures 105 and 188 are opaque, while the backbone of the DNA in relatively overlapping structure 90 is pictured using a semi-transparent representation. Figure created using VMD.

absolute value of the normalised electrostatic potential is 0.1. In points outside this region, it is a valid approximation to state that the linearised and non-linear PBE provide the same behaviour for the potential (Fig.5.9). An increase in the distance between spheres causes the thickness of the non-linearity layer to slightly decrease, but the mean value remains approximately 20.5Å. In our model, we are also able to calculate a "renormalised charge", which is the total charge included in the non-linearity layer (see Appendix B). We used DelPhi to calculate the electrostatic potential and the electrostatic interaction energies on specific points between and around the two spheres. We then used these measurements to parametrise the values of the parameters of our model: the electric charge and the dipole moment.

In our numerical results, we have not included the trend of the ionic contribution to the interaction energy, focusing on the Coulombic and Reaction Field contributions. The ionic contribution, which we have introduced in Chapter 2, is two orders of magnitude



(a)

(b)

(c)

Figure 5.6: Examples of medoid2 combinations on which DelPhi calculations were performed. We explored a wide range of differences and relative orientations, such as stacked NCPs (Fig. 5.6a) that would show up as successive nucleosomes in the chromatin fibre, and structures with larger distances between NCP centres (Figs. 5.6b, 5.6c) that represent interactions between nucleosomes in different parts of the fibre that approach each other through secondary folding.

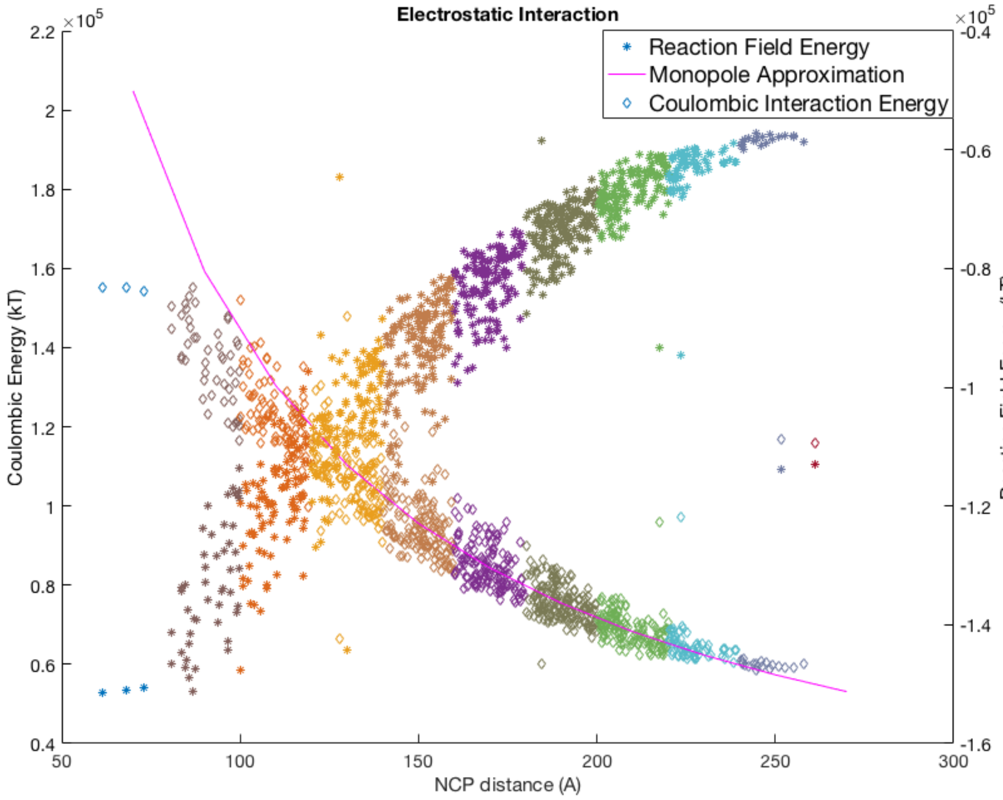


Figure 5.7: Scatter plot of electrostatic interaction energies (Coulombic interaction energy and reaction field energy as defined in Section 5.4.2) calculated by DelPhi on pairs of medoid2 structures in different relative distances and rotations. Asterisks indicate the values of reaction field energy and diamonds indicate the Coulombic energy. The magenta line indicates the monopole approximation of the Coulombic Energy for two particles bearing the same charge as medoid2.

smaller than the Reaction Field contribution, and therefore we have omitted it from our qualitative description of nucleosome electrostatic interactions in intermediate distances. We have verified this by calculating the ionic contribution in a system of two interacting spheres of radii $a_i = a_j = 60$Å, each bearing a monopole charge of $q_i = q_j = -144e$, placed at a distance of 200Å.

### 5.5.1 General Analytical Asymptotic Expressions of Electrostatic Energy

We state the leading terms of the electrostatic energy in different systems, differentiated by different charge combinations in the two spheres (monopole - monopole, dipole - dipole, monopole - dipole). The total electrostatic energy of the system is

$$\mathcal{E} = \mathcal{E}_{\mathrm{Born}} + \mathcal{E}_{\mathrm{Int}}, \tag{5.13}$$

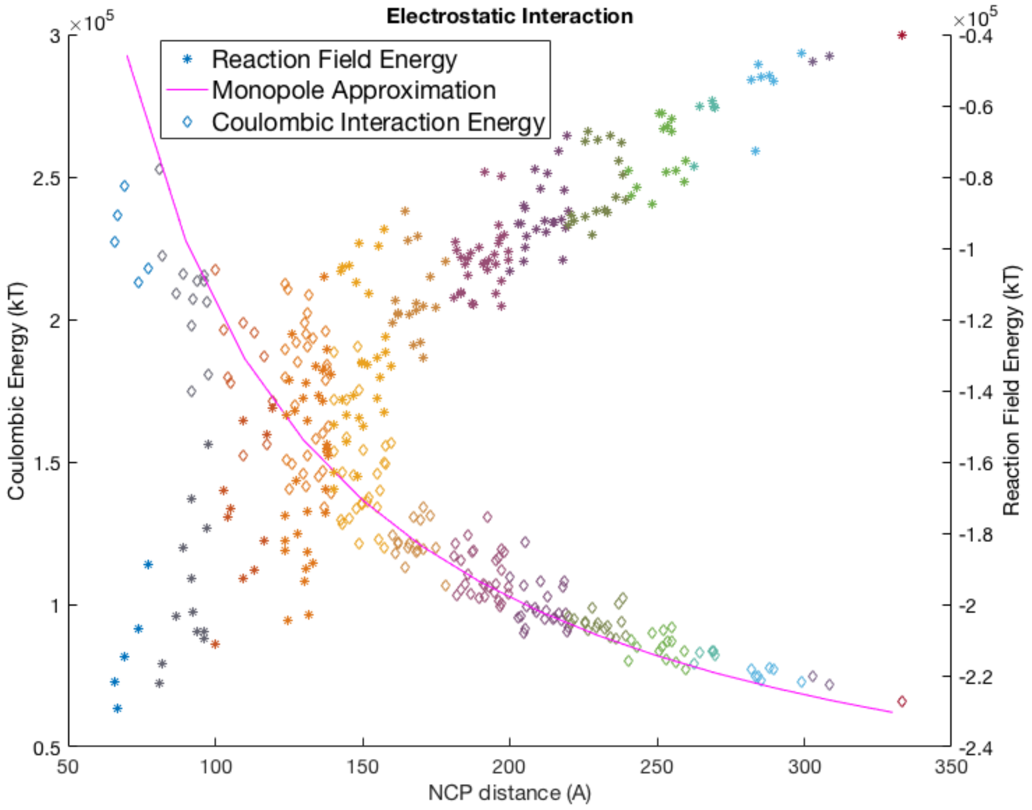


Figure 5.8: Scatter plot of electrostatic interaction energies (Coulombic interaction energy and reaction field energy as defined in Section 5.4.2) calculated by DelPhi on pairs of 3AFA NCPs with a 20-bp piece of linker DNA attached to each NCP in different relative distances and rotations. Asterisks indicate the values of reaction field energy and diamonds indicate the Coulombic energy. The magenta line indicates the monopole approximation of the Coulombic Energy for two particles bearing the same charge as the 3AFA/linker DNA complex.

where $\mathcal{E}_{\text{Born}}$ is the Born contribution, while $\mathcal{E}_{\text{Int}}$ denotes the Coulombic interaction contribution, taking into account electrostatic screening. This Interaction term is be represented by the following expression:

$$\mathcal{E}_{\text{Int}} = \hat{\mathcal{E}}_{\text{Int}} + O(\hat{\omega}) \quad \text{as} \quad \hat{\omega} \longrightarrow 0; \tag{5.14}$$

$\hat{\mathcal{E}}_{\text{Int}}(R)$ is the leading term of the Interaction part, and $\hat{\omega} := \tilde{a}^2 e^{-2\tilde{R}}/\tilde{R}^2 = a^2 e^{-2\kappa R}/R^2$ ($a := \max\{a_i, a_j\}$) is the doubly-screened dimensionless factor. Let there be two particles, $i$ and $j$, with radii $a_i$, $a_j$, and dielectric constants $\varepsilon_i$, $\varepsilon_j$, immersed in a dielectric medium with relative dielectric constant $\varepsilon_{\text{m}}$. Tildes denote dimensionless quantities scaled by the Debye length $\kappa^{-1}$: $\tilde{a}_i = \kappa a_i$, $\tilde{a}_j = \kappa a_j$, $\tilde{R} = \kappa R$, etc. The following expressions are given for the case of identical particles, i.e. $q_i = q_j = q$, $a_i = a_j = a$ and $\varepsilon_i = \varepsilon_j = \varepsilon$.

In the next paragraphs we provide the expressions for the energy in different cases: Monopole - Monopole, Dipole - Dipole, Monopole - Dipole, and the general Monopole - Dipole – Monopole - Dipole case. The plots representing the energy as a function of the

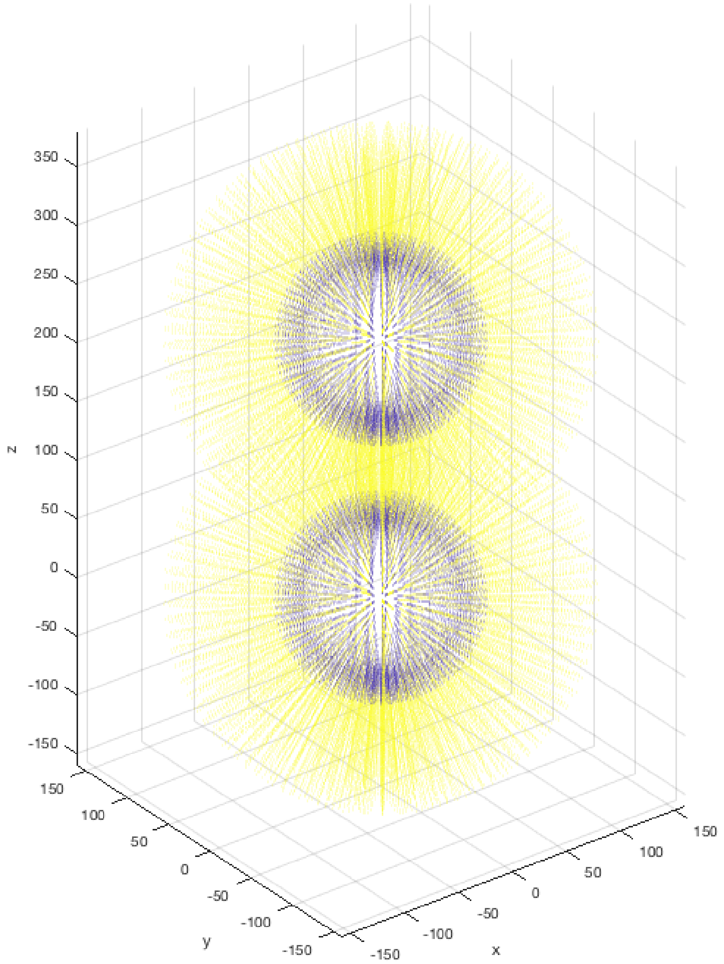


Figure 5.9: Electrostatic potential distribution between two dielectric spheres of radii $a_i = a_j = 60$Å, each bearing a monopole charge of $q_i = q_j = -144e$ in its interior, where $e$ is the charge of an electron. The distance between the centres of the two spheres is 220Å. The points in yellow indicate the region where the linearised approximation of the PBE holds, while the points belonging to the non-linearity corona are denoted in blue. The approximate width of the non-linearity region is 20.5Å.

distance between the centres of the two spheres are provided in Appendix B for the first three cases (Figs. 3-6) and in Figs. 5.10-5.11 below for the general case.

### Monopole - Monopole system

We assume that the two spheres bear only a single charge, placed in their geometric centres. The expressions for the Interaction and Born energies are given by:

$$\mathcal{E}_{\text{Born}} = \frac{q^2\kappa}{4\pi\varepsilon_0\tilde{a}}\left(\frac{1}{(1+\tilde{a})\varepsilon_{\text{m}}} - \frac{1}{\varepsilon}\right), \tag{5.15}$$

$$\hat{\mathcal{E}}_{\text{Int}} = \frac{q^2 \kappa e^{2\tilde{a}-\tilde{R}}}{4\pi\varepsilon_0\varepsilon_{\text{m}}\tilde{R}(1+\tilde{a})^2}. \tag{5.16}$$

These two expressions were derived in [186]. We see that the Interaction energy in this case is given by the well-known DLVO expression (see Chapter 2).

### Dipole - Dipole system

To improve the level of detail of this reductionistic representation, we now consider that each sphere contains a dipole. The expressions for the energy now become:

$$\mathcal{E}_{\text{Born}} = \frac{\kappa^3 p_i^2}{4\pi\varepsilon_0\tilde{a}^3}\frac{2(1+\tilde{a})(\varepsilon-\varepsilon_{\text{m}}) - \tilde{a}^2\varepsilon_{\text{m}}}{\varepsilon\left(\tilde{a}^2\varepsilon_{\text{m}} + (1+\tilde{a})(\varepsilon+2\varepsilon_{\text{m}})\right)}, \tag{5.17}$$

$$\begin{aligned}\hat{\mathcal{E}}_{\text{Int}} =& \frac{9\kappa^3\varepsilon_{\text{m}}e^{-(\tilde{R}-2\tilde{a})}}{4\pi\varepsilon_0\tilde{R}^2}\left(1+\frac{1}{\tilde{R}}\right)\frac{\mathbf{p}_i\cdot\mathbf{p}_j}{\left(\tilde{a}^2\varepsilon_{\text{m}} + (1+\tilde{a})(\varepsilon+2\varepsilon_{\text{m}})\right)^2} \\ &- \frac{9\kappa^3\varepsilon_{\text{m}}e^{-(\tilde{R}-2\tilde{a})}}{4\pi\varepsilon_0\tilde{R}}\left(1+\frac{3}{\tilde{R}}+\frac{3}{\tilde{R}^2}\right)\frac{(\mathbf{p}_i\cdot\hat{\mathbf{n}}_0)(\mathbf{p}_j\cdot\hat{\mathbf{n}}_0)}{\left(\tilde{a}^2\varepsilon_{\text{m}} + (1+\tilde{a})(\varepsilon+2\varepsilon_{\text{m}})\right)^2},\end{aligned} \tag{5.18}$$

where $\hat{\mathbf{n}}_0$ is a unit vector in the direction connecting the spheres' centres and $\mathbf{p}_{i,j}$ is the dipole moment.

Note that in the limit $\kappa \longrightarrow 0$, the above expression for the Born energy $\mathcal{E}_{\text{Born}}$ turns into the expression [195]:

$$\mathcal{E}_{\text{Born}}\big|_{\kappa=0} = \frac{p_i^2}{4\pi\varepsilon_0 a^3}\frac{\varepsilon-\varepsilon_{\text{m}}}{\varepsilon(\varepsilon+2\varepsilon_{\text{m}})} + \frac{p_j^2}{4\pi\varepsilon_0 a^3}\frac{\varepsilon-\varepsilon_{\text{m}}}{\varepsilon(\varepsilon+2\varepsilon_{\text{m}})}. \tag{5.19}$$

Similarly, in the limit $\kappa \longrightarrow 0$ and when imposing the condition $\varepsilon = \varepsilon_{\text{m}}$, the above expression for the Interaction energy turns into the well-known formula for the interaction energy of two dipoles [188, 196]:

$$\hat{\mathcal{E}}_{\text{Int}} = \frac{\mathbf{p}_i\cdot\mathbf{p}_j - 3(\mathbf{p}_i\cdot\hat{\mathbf{n}}_0)(\mathbf{p}_j\cdot\hat{\mathbf{n}}_0)}{4\pi\varepsilon_0\varepsilon_{\text{m}}R^3}. \tag{5.20}$$

### Monopole - Dipole system

Finally, we assume that sphere $i$ carries a point charge $q_i$, while the opposite sphere $j$ carries a point dipole $\mathbf{p}_j$. Then

$$\mathcal{E}_{\text{Born}} = \frac{q_i^2\kappa}{8\pi\varepsilon_0\tilde{a}}\left(\frac{1}{(1+\tilde{a})\varepsilon_{\text{m}}} - \frac{1}{\varepsilon}\right) + \frac{\kappa^3 p_j^2}{8\pi\varepsilon_0\tilde{a}^3}\frac{2(1+\tilde{a})(\varepsilon-\varepsilon_{\text{m}}) - \tilde{a}^2\varepsilon_{\text{m}}}{\varepsilon\left(\tilde{a}^2\varepsilon_{\text{m}} + (1+\tilde{a})(\varepsilon+2\varepsilon_{\text{m}})\right)}, \tag{5.21}$$

$$\hat{\mathcal{E}}_{\text{Int}} = \frac{3\kappa^2 e^{-(\tilde{R}-2\tilde{a})}q_i(\mathbf{p}_j\cdot\hat{\mathbf{n}}_0)}{4\pi\varepsilon_0(1+\tilde{a})\left(\tilde{a}^2\varepsilon_{\text{m}} + (1+\tilde{a})(\varepsilon+2\varepsilon_{\text{m}})\right)\tilde{R}}\left(1+\frac{1}{\tilde{R}}\right), \tag{5.22}$$

where $\hat{\mathbf{n}}_0$ is the unit vector directed from the $j$-th sphere's centre to the $i$-th sphere's centre. In the limit $\kappa \longrightarrow 0$ and when imposing the condition $\varepsilon = \varepsilon_\mathrm{m}$, the above expression immediately yields the classical result [188, 196].

$$\hat{\mathcal{E}}_\mathrm{Int} = \frac{(\mathbf{p}_j \cdot \hat{\mathbf{n}}_0) q_i}{4\pi\varepsilon_0\varepsilon_\mathrm{m} R^2} \tag{5.23}$$

### Monopole-Dipole – Monopole-Dipole system (general case)

Finally, let us assume that both spheres $i$ and $j$ carry point charges $q_i$, $q_j$, and point dipoles $\mathbf{p}_i$, $\mathbf{p}_j$, located at the spheres' centres, the previous expressions can be combined to give the overall energy estimate:

$$\mathcal{E}_\mathrm{Born} = \mathcal{E}_\mathrm{Born}^{(\mathrm{mon}\ i)} + \mathcal{E}_\mathrm{Born}^{(\mathrm{mon}\ j)} + \mathcal{E}_\mathrm{Born}^{(\mathrm{dip}\ i)} + \mathcal{E}_\mathrm{Born}^{(\mathrm{dip}\ j)}, \tag{5.24}$$

$$\hat{\mathcal{E}}_\mathrm{Int} = \hat{\mathcal{E}}_\mathrm{Int}^{(\mathrm{mon}\ i,\ \mathrm{mon}\ j)} + \hat{\mathcal{E}}_\mathrm{Int}^{(\mathrm{mon}\ i,\ \mathrm{dip}\ j)} + \hat{\mathcal{E}}_\mathrm{Int}^{(\mathrm{mon}\ j,\ \mathrm{dip}\ i)} + \hat{\mathcal{E}}_\mathrm{Int}^{(\mathrm{dip}\ i,\ \mathrm{dip}\ j)} \tag{5.25}$$

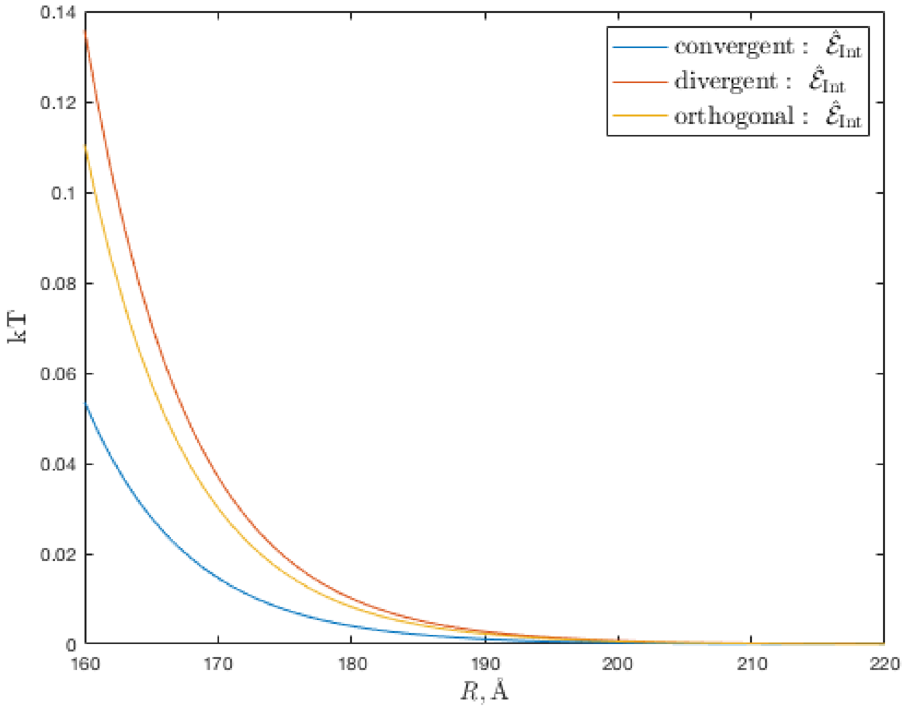


Figure 5.10: Electrostatic energies ($\mathcal{E}_\mathrm{Born}$ and $\hat{\mathcal{E}}_\mathrm{Int}$) in the general Monopole-Dipole – Monopole-Dipole system case, as defined in eqs. (5.24) - (5.25). The values of the energy are given as a function of the distance between the centres of the spheres for different relative dipole orientations: convergent, divergent, and orthogonal. The Born energy is constant.

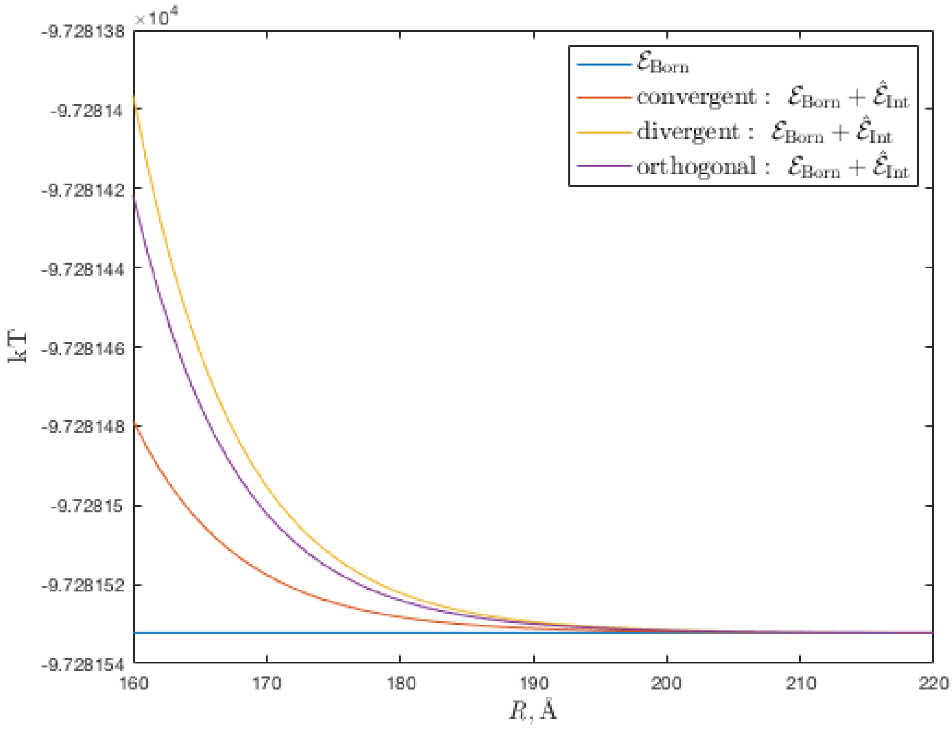


Figure 5.11: Electrostatic interaction energy $\hat{\mathcal{E}}_{\text{Int}}$ in the general Monopole-Dipole – Monopole-Dipole system case, as defined in eq. (5.25). The values of the energy are given as a function of the distance between the centres of the spheres for different relative dipole orientations: convergent, divergent, and orthogonal.

## 5.6 Conclusions

In this Chapter, we have discussed electrostatic interactions between nucleosomes at close and intermediate distances, and we have provided values for the electrostatic interaction energies. As we saw in Chapter 4, electrostatic calculations on nucleosomes require the use of the full non-linear PBE, as the negative charge of the phosphate groups in the DNA backbone is such that the electrostatic potential is elevated near the nucleosome, rendering the linearised approximation of the PBE unsuitable. However, this poses great computational limitations in the use of the PBE framework for the study of electrostatics in the chromatin fibre; the use of the full non-linear PBE in such calculations would be prohibitively time-consuming and demanding in terms of computational resources.

Our attempt to solve this conundrum makes use of the following assumption: there is a region around the nucleosome where the electrostatic potential is elevated and indeed requires the full non-linear PBE treatment, but beyond this region the potential drops sufficiently for the linearised approximation to be viable without significant loss of accuracy in the calculations. Performing simulations on simplified two-sphere systems, placing the charge of the 1KX5 nucleosome in the centre of the spheres, we show that our assumption holds and indeed the difference between the estimates of the electrostatic potential using the full non-linear and the linearised treatment after a certain distance from the surface of the sphere is negligible. We term the non-linearity region a "non-linear" corona around the sphere, and discuss its thickness for different sphere separation distances. In our expressions for the electrostatic energy terms we split the total energy into two contributions: the Born energy term, which is independent of the distance between the particles, and the Interaction energy term, a screened Coulombic interaction term. The Interaction energy term includes the ionic interactions, providing a significant simplification to the "traditional" calculation procedure for this energy contribution with the PBE; as we illustrated in Chapter 2, the PBE formalism requires either an integration on the entire volume of the solvent where the potential is non-zero, or two calculations, with and without ions in the solvent, and a subsequent calculation of the ionic component of the interaction energy through the difference in the grid energy contribution in the two cases.

Therefore, in the case of close inter-nucleosome distances we have used the full non-linear PBE to study their electrostatic interactions. We calculated the electrostatic interaction energy between docked nucleosomes generated by HADDOCK using the full non-linear PBE, and we have presented the graphic representations of the most favourable docking poses from an electrostatic point of view, i.e. the ones that present the lowest energies. The author wishes to thank Prof. A. Bonvin for having provided the HADDOCK docked nucleosome poses. Information on the electrostatics of docked nucleosome pairs becomes very important when considering tight packing situations in chromatin. Furthermore, we have studied the electrostatic interactions between nucleosome pairs at intermediate distances, this time using the linearised PBE, sampling different relative translations and rotations. Once again, taken in the wider context of chromatin compaction, these provide both electrostatic and topological information, as the presence of steric overlap of the generated conformations is also evaluated.

The research conducted in this Chapter is a building block towards a comprehensive model of electrostatic interactions between nucleosomes at all scales relevant to chromatin compaction, from stacked nucleosomes to nucleosomes found in positions appropriate for non -neighbouring nucleosomes in the chromatin fibre. Eventually, we believe that our results will be a first step towards an ultra-CG force field for chromatin, integrating our electrostatic description and results into a description of the non-bonded interactions between nucleosomes, and combining it with a mechanical counterpart for the treatment of the bonded interactions between nucleosomes.

# CHAPTER 6

# Conclusions & Outlook

Chromatin is a complex of proteins and DNA found in the nuclei of eukaryotic cells. Its primary functions are to compact and structurally reinforce the DNA whilst keeping it accessible to DNA-binding proteins, such as transcription factors. The building blocks of chromatin are the nucleosomes, structures in which approximately 147 base pairs of DNA wrap around a histone protein octamer. Due to its molecular composition, its size, and its ability to influence gene expression and transcription through its dynamics, chromatin is an extremely complex system. Its intricate dynamics are tuned by both mechanical and electrostatic factors and by biomolecular interactions occurring in the cell nucleus. Chromatin is, therefore, a system the study of which requires multiscale analysis and the synergy of both computational and experimental techniques.

In my work, I focus on the role of electrostatics and the often underrated role of solvation as the driving mechanisms of chromatin topology and as the determinants of the relative conformational equilibria. The high negative electric charge of the DNA backbone is partially neutralised through direct interactions with histones, but electrostatic stabilisation of the chromatin fibre is achieved through a combination of this effect with long-range electrostatics and solvent screening. I studied electrostatics at different levels in chromatin, starting from the study of intra-nucleosome electrostatic interactions and their repercussions on nucleosome structure and dynamics. I also investigated inter-nucleosome electrostatic interactions at close and intermediate distances, and I considered the role of linker DNA in such interactions. Finally, I developed asymptotic expressions for the interaction energy between nucleosomes at larger distances. I conducted my analysis and simulations in the Poisson Boltzmann Equation (PBE) framework.

In my study of intra-nucleosome electrostatic interactions, I focused on the histone tails, the intrinsically disordered terminal domains of the histone proteins, which bear a positive electric charge and are important for the electrostatic stabilisation of nucleosomes. I proposed a methodology for the study of protein-DNA electrostatic interactions and I apply it to clarify the effect of the histone tails. My methodology correlates electrostatic interactions and spatial conformation at the residue level with mechanical effects and repercussions on their structure and function. During my analysis, I was able to identify the possible electrostatic influence on many effects, such as spontaneous DNA unwrapping, nucleosome destabilisation upon histone tail truncation, and the key role of specific arginine and lysine residues. I studied the correlation between the number of histone tail atoms in proximity to the DNA, and the intensity of the electrostatic interaction. Finally, I stressed the role of H2A C-terminal tail: small changes in the positioning of this short tail had a significant impact on the electric field, implying that this often overlooked feature of the nucleosome actually holds particular importance. I proposed the application of my methodology in conjunction with CG models that treat the mechanical properties of the DNA. Such models often infer the mechanical forces exerted on the DNA by examining the known mechanical properties of the DNA, such as characteristic torsion angles, and completely omit the histone tails.

We subsequently followed a quantitative approach to describe the solvation interactions in the nucleosome, discussing the importance of taking into account nucleosome porosity in chromatin models, instead of treating nucleosomes as solid objects. I motivated my point by providing quantitative information on the number of cavities and pockets present in the nucleosome crystal structure, and their volume. Nucleosomes can be likened much more to sponges than to solid objects; their overall high porosity and particularly the presence of a central channel facilitates the permeation of biologically and electrostatically significant counterions, such as $Na^{+1}$ and $Mg^{+2}$, and enhances electrostatic screening of the negative charge of the DNA backbone.

In order to evaluate the importance of solvation and of the ionic environment, but also to evaluate the intensity of the electrostatic potential around nucleosomes, I performed Zeta Potential measurements, accompanied by Dynamic Light Scattering measurements. I observed a quasi-monotonical relationship between the Zeta Potential and NaCl concentration, and a leap in structure size for very high NaCl concentration. I was able to reproduce these results in my computational framework, supporting the validity of my methodology. This process led reminded us of the fact that the potential was underestimated if one used the linearised PBE instead of the full non-linear PBE, clearly showing the necessity of using the full non-linear PBE when treating nucleosomes. However, a single nucleosome is already a quite large structure, containing tens of thousands of atoms. Therefore, a reasonable question arises: how are we to study the electrostatic interactions between two or more nucleosomes in the chromatin fibre, where the large number of atoms and the size of the system under scrutiny would make the use of the full non-linear PBE prohibitively computationally demanding?

Our attempt to solve this dilemma of accuracy versus computational realisability makes use of the following observation: there is a region around the nucleosome where the electrostatic potential is elevated and indeed requires the full non-linear PBE treatment, but beyond this region the potential drops sufficiently for the linearised approximation to be viable without significant loss of accuracy in the calculations. I tested this assumption by performing simulations on simplified two-sphere systems bearing the approximate radius and the charge of a nucleosome. I was able to show that my "non-linearity region" assumption holds and indeed the difference between the estimates of the electrostatic potential using the full non-linear and the linearised treatment after a certain distance from the surface of the sphere is negligible. I term the non-linearity region a "non-linear" corona around the sphere, the thickness of which varies slightly as the distance between the two spheres increases, but can be considered approximately constant.

Finally, making use of this knowledge, I discussed the electrostatic interaction energies between nucleosomes at close distances obtained using HADDOCK, and their trends for nucleosome pairs at intermediate distances. At close distances, I used docked nucleosome structures and determined the most electrostatically favourable conformations at very tight chromatin packing situations, using the full non-linear PBE, as such situations correspond to the interior of the non-linearity region. Furthermore, I have studied the electrostatic interactions between nucleosome pairs at intermediate distances, this time using the linearised PBE, sampling different relative translations and rotations. Once again, taken in the wider context of chromatin compaction, these provide both electrostatic and topological information, as the feasibility of the generated conformations is also evaluated.

The research I conducted on inter-nucleosome electrostatic interactions, and particularly my focus on nucleosome electrostatic energies, is a building block towards a comprehensive model of electrostatic interactions between nucleosomes at all scales relevant to chromatin compaction. Eventually, I believe that my results will be a first step towards an ultra-coarse-grained force field for chromatin, integrating my electrostatic description and results into a description of the non-bonded interactions between nucleosomes, and combining it with a mechanical counterpart for the treatment of the bonded interactions between nucleosomes. In this case, the asymptotic expressions I have derived for the electrostatic energy between dielectric spheres immersed in a dielectric solution can be used when nucleosomes are sufficiently far away from each other for their structural characteristics to become irrelevant for the purposes of an electrostatic characterisation. These asymptotic expressions form the basis of the terms of the force field, the parameters of which will be parametrised using the explicit numerical calculations of nucleosome interaction energies.

# List of Publications

[1] **Artemi Bendandi**, Sergii Siryk, Alberto Diaspro, and Walter Rocchia. The Role of Histone Tails in Nucleosome Stability: an Electrostatic Perspective. *In preparation.*

[2] Sergii Siryk, **Artemi Bendandi**, Alberto Diaspro, and Walter Rocchia. Monopolar and dipolar dielectric spheres interacting in electrolytic solution: a linearized Poisson-Boltzmann equation model. *In preparation.*

[3] Sergii Siryk, **Artemi Bendandi**, Alberto Diaspro, and Walter Rocchia. Monopolar and dipolar dielectric spheres interacting in electrolytic solution: a full non linear Poisson-Boltzmann equation model. *In preparation.*

[4] Aymeric Le Gratiet, **Artemi Bendandi**, Colin J.R. Sheppard, and Alberto Diaspro. Polarimetric optical scanning microscopy of zebrafish embryonic development using the coherency matrix. *Journal of Biophotonics*, 2021. DOI: `10.1002/jbio.202000494`.

[5] Aymeric Le Gratiet, Luca Lanzano, **Artemi Bendandi**, Riccardo Marongiu, Paolo Bianchini, and Alberto Diaspro. Phasor approach of Mueller matrix optical scanning microscopy for label-free biological tissue orientation imaging. *Under Review*, 2020.

[6] **Artemi Bendandi**, Alessandro S. Patelli, Alberto Diaspro, and Walter Rocchia. The Role of Histone Tails in Nucleosome Stability: an Electrostatic Perspective. *Computational and Structural Biotechnology*, 18:2799–2809, 2020. DOI: `10.1016/j.csbj.2020.09.034`.

[7] Colin J. R. Sheppard, **Artemi Bendandi**, Aymeric Le Gratiet, and Alberto Diaspro. Eigenvectors of polarization coherency matrices. *Journal of the Optical Society of America A*, 37(7):1143, June 2020. DOI: `10.1364/josaa.391902`.

[8] **Artemi Bendandi**, Silvia Dante, Syeda Rehana Zia, Alberto Diaspro, and Walter Rocchia. Chromatin Compaction Multiscale Modeling: A Complex Synergy Between Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences*, 7, February 2020. DOI: `10.3389/fmolb.2020.00015`.

[9] Colin J. R. Sheppard, **Artemi Bendandi**, Aymeric Le Gratiet, and Alberto Diaspro. Polarization in reflectance imaging. *Journal of the Optical Society of America A*, 37(3):491, February 2020. DOI: `10.1364/josaa.379327`.

[10] Aymeric Le Gratiet, Marta d'Amora, Marti Duocastella, Riccardo Marongiu, **Artemi Bendandi**, Silvia Giordani, Paolo Bianchini, and Alberto Diaspro. Zebrafish structural development in mueller-matrix scanning microscopy. *Scientific Reports*, 9(1), December 2019. DOI: `10.1038/s41598-019-56610-9`.

[11] Colin J. R. Sheppard, **Artemi Bendandi**, Aymeric Le Gratiet, and Alberto Diaspro. Eigenvalues of the coherency matrix for exact backscattering. *Journal of the Optical Society of America A*, 36(9):1540, August 2019. DOI: `10.1364/josaa.36.001540`.

[12] Colin J. R. Sheppard, **Artemi Bendandi**, Aymeric Le Gratiet, and Alberto Diaspro. Coherency and differential mueller matrices for polarizing media. *Journal of the Optical Society of America A*, 35(12):2058, November 2018. DOI: `10.1364/josaa.35.002058`.

# List of Posters & Oral Presentations

- **CECAM Mixed-Gen** – *Session 2: Active Matter*
  04/03/2021
  ***Poster***

- **Protein Electrostatics Conference Webinar Series**
  24/02/2021
  ***Oral Presentation***

- **BioExcel Student Webinar** – *Winter School 2020 Edition*
  18/01/2021
  ***Invited Oral Presentation***

- **Remote BioExcel Winter School On Biomolecular Simulations**
  30/11 – 04/12/2020
  ***Poster, Winner of Poster Competition***

- **BImBS 2019**
  Lugano, Switzerland 05 – 12/10/2019
  *BioInformatics meets BioSimulations in protein and DNA studies: from theory to practice*
  ***Oral Presentation***

- **CECAM50**
  Lausanne, Switzerland 09 – 12/09/2019
  *Molecular and materials simulations at the turn of the decade: Celebrating 50 years of CECAM*
  ***Invited Participation, Contribution to Italian CECAM Node Poster***

- **Italian Celebration Day for the 50 Years of CECAM**
  Bologna, Italy 09/07/2019
  ***Poster, Winner of Poster Competition***

- **Computational mathematics for model reduction and predictive modelling in molecular and complex systems**
  Lausanne, Switzerland 21 – 29/05/2019

*CIB - CECAM Workshop*
***Invited Poster***

- **Multiscale Modelling of Chromatin**
  Les Houches, France 31/03 – 05/04/2019
  *Biophysical Society Thematic Meeting*
  ***Oral Presentation***

# Bibliography

[1] Artemi Bendandi et al. «Chromatin Compaction Multiscale Modeling: A Complex Synergy Between Theory, Simulation, and Experiment». In: *Frontiers in Molecular Biosciences* 7 (Feb. 2020). DOI: `10.3389/fmolb.2020.00015` (cit. on pp. 1, 18).

[2] Clara L. van Emmerik and Hugo van Ingen. «Unspinning chromatin: Revealing the dynamic nucleosome landscape by NMR». In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 110 (Feb. 2019), pp. 1–19. DOI: `10.1016/j.pnmrs.2019.01.002` (cit. on pp. 3, 13).

[3] T. Dechat et al. «Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin». In: *Genes & Development* 22.7 (Apr. 2008), pp. 832–853. DOI: `10.1101/gad.1652708` (cit. on p. 2).

[4] Andrew D. Stephens et al. «Chromatin and lamin A determine two different mechanical response regimes of the cell nucleus». In: *Molecular Biology of the Cell* 28.14 (July 2017). Ed. by Dunn Alex R., pp. 1984–1996. DOI: `10.1091/mbc.e16-09-0653` (cit. on p. 2).

[5] Sarah M. Schreiner et al. «The tethering of chromatin to the nuclear envelope supports nuclear mechanics». In: *Nature Communications* 6.1 (June 2015). DOI: `10.1038/ncomms8159` (cit. on p. 2).

[6] Andrew D. Stephens et al. «Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins». In: *Molecular Biology of the Cell* 29.2 (Jan. 2018). Ed. by Tom Misteli, pp. 220–233. DOI: `10.1091/mbc.e17-06-0410` (cit. on p. 2).

[7] Anne R. Shim et al. «Dynamic Crowding Regulates Transcription». In: *Biophysical Journal* 118.9 (May 2020), pp. 2117–2129. DOI: `10.1016/j.bpj.2019.11.007` (cit. on p. 2).

[8] Davide Marenduzzo, Cristian Micheletti, and Peter R. Cook. «Entropy-Driven Genome Organization». In: *Biophysical Journal* 90.10 (May 2006), pp. 3712–3721. DOI: `10.1529/biophysj.105.077685` (cit. on p. 2).

[9] C. A. Brackley et al. «Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization». In: *Proceedings of the National Academy of Sciences* 110.38 (Sept. 2013), E3605–E3611. DOI: `10.1073/pnas.1302950110` (cit. on p. 2).

[10] Chris A. Brackley et al. «Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains». In: *Nucleic Acids Research* 44.8 (Apr. 2016), pp. 3503–3512. DOI: `10.1093/nar/gkw135` (cit. on p. 2).

[11] Yasumasa Joti et al. «Chromosomes without a 30-nm chromatin fiber». In: *Nucleus* 3.5 (Sept. 2012), pp. 404–410. DOI: `10.4161/nucl.21222` (cit. on pp. 2, 16).

[12] Tadasu Nozaki et al. «Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging». In: *Molecular Cell* 67.2 (July 2017), 282–293.e7. DOI: `10.1016/j.molcel.2017.06.018` (cit. on p. 2).

[13] Horng D. Ou et al. «ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells». In: *Science* 357.6349 (July 2017), eaag0025. DOI: `10.1126/science.aag0025` (cit. on p. 2).

[14] Rabia Laghmach, Michele Di Pierro, and Davit A. Potoyan. «Mesoscale Liquid Model of Chromatin Recapitulates Nuclear Order of Eukaryotes». In: *Biophysical Journal* 118.9 (May 2020), pp. 2130–2140. DOI: `10.1016/j.bpj.2019.09.013` (cit. on p. 2).

[15] Anna A. Kalashnikova et al. «The role of the nucleosome acidic patch in modulating higher order chromatin structure». In: *Journal of The Royal Society Interface* 10.82 (May 2013), p. 20121022. DOI: `10.1098/rsif.2012.1022` (cit. on pp. 3, 29, 32, 62, 68).

[16] Mithun Biswas et al. «Role of Histone Tails in Structural Stability of the Nucleosome». In: *PLoS Computational Biology* 7.12 (Dec. 2011). Ed. by Marc A. Marti-Renom, e1002279. DOI: `10.1371/journal.pcbi.1002279` (cit. on pp. 4, 5).

[17] Robert K. McGinty and Song Tan. «Nucleosome Structure and Function». In: *Chemical Reviews* 115.6 (Dec. 2014), pp. 2255–2273. DOI: `10.1021/cr500373h` (cit. on pp. 4, 29, 30, 59, 68).

[18] Jochen Erler et al. «The Role of Histone Tails in the Nucleosome: A Computational Study». In: *Biophysical Journal* 107.12 (Dec. 2014), pp. 2911–2922. DOI: `10.1016/j.bpj.2014.10.065` (cit. on p. 4).

[19] E. G. Iashina et al. «Additive scaling law for structural organization of chromatin in chicken erythrocyte nuclei». In: *Physical Review E* 96.1 (July 2017). DOI: `10.1103/physreve.96.012411` (cit. on pp. 4, 12, 15, 17).

[20] Alexey K. Shaytan et al. «Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions». In: *Journal of Molecular Biology* 428.1 (Jan. 2016), pp. 221–237. DOI: `10.1016/j.jmb.2015.12.004` (cit. on pp. 4, 5, 8, 17, 29, 36, 44, 68, 72).

[21] Marco Pasi and Richard Lavery. «Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics». In: *Nucleic Acids Research* 44.11 (Apr. 2016), pp. 5450–5456. DOI: `10.1093/nar/gkw293` (cit. on pp. 4, 15, 29).

[22] Nils B. Becker and Ralf Everaers. «DNA nanomechanics: How proteins deform the double helix». In: *The Journal of Chemical Physics* 130.13 (Apr. 2009), p. 135102. DOI: `10.1063/1.3082157` (cit. on pp. 4, 35).

[23] Heather J. Szerlong and Jeffrey C. Hansen. «Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structureThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process.» In: *Biochemistry and Cell Biology* 89.1 (Feb. 2011), pp. 24–34. DOI: `10.1139/o10-139` (cit. on pp. 4, 6, 30).

[24] Giovanni B Brandani et al. «DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations». In: *Nucleic Acids Research* 46.6 (Feb. 2018), pp. 2788–2801. DOI: `10.1093/nar/gky158` (cit. on pp. 5, 29, 59).

[25] Gordon S. Freeman et al. «DNA Shape Dominates Sequence Affinity in Nucleosome Formation». In: *Physical Review Letters* 113.16 (Oct. 2014). DOI: `10.1103/physrevlett.113.168101` (cit. on p. 5).

[26] Georgy N. Rychkov et al. «Partially Assembled Nucleosome Structures at Atomic Detail». In: *Biophysical Journal* 112.3 (Feb. 2017), pp. 460–472. DOI: `10.1016/j.bpj.2016.10.041` (cit. on pp. 5, 7, 15, 17, 52).

[27] Alexander Gansen et al. «Opposing roles of H3- and H4-acetylation in the regulation of nucleosome structure—a FRET study». In: *Nucleic Acids Research* 43.3 (Jan. 2015), pp. 1433–1443. DOI: `10.1093/nar/gku1354` (cit. on pp. 5, 54).

[28] Maria L Kireeva et al. «Nucleosome Remodeling Induced by RNA Polymerase II». In: *Molecular Cell* 9.3 (Mar. 2002), pp. 541–552. DOI: `10.1016/s1097-2765(02)00472-0` (cit. on p. 5).

[29] Gu Li et al. «Rapid spontaneous accessibility of nucleosomal DNA». In: *Nature Structural & Molecular Biology* 12.1 (Dec. 2004), pp. 46–53. DOI: `10.1038/nsmb869` (cit. on p. 5).

[30] Gungor Ozer, Antoni Luque, and Tamar Schlick. «The chromatin fiber: multiscale problems and approaches». In: *Current Opinion in Structural Biology* 31 (Apr. 2015), pp. 124–139. DOI: `10.1016/j.sbi.2015.04.002` (cit. on p. 5).

[31] Hidenori Kato et al. «Characterization of the N-Terminal Tail Domain of Histone H3 in Condensed Nucleosome Arrays by Hydrogen Exchange and NMR». In: *Journal of the American Chemical Society* 131.42 (Oct. 2009), pp. 15104–15105. DOI: `10.1021/ja9070078` (cit. on pp. 5, 13, 29, 35).

[32] Suman Saurabh et al. «Atomistic Simulation of Stacked Nucleosome Core Particles: Tail Bridging, the H4 Tail, and Effect of Hydrophobic Forces». In: *The Journal of Physical Chemistry B* 120.12 (Mar. 2016), pp. 3048–3060. DOI: `10.1021/acs.jpcb.5b11863` (cit. on pp. 5, 54).

[33] Jenna M. Buckwalter et al. «Regulation of chromatin folding by conformational variations of nucleosome linker DNA». In: *Nucleic Acids Research* 45.16 (June 2017), pp. 9372–9387. DOI: `10.1093/nar/gkx562` (cit. on pp. 5, 7, 8, 17).

[34] Pablo D Dans et al. «Multiscale simulation of DNA». In: *Current Opinion in Structural Biology* 37 (Apr. 2016), pp. 29–45. DOI: `10.1016/j.sbi.2015.11.011` (cit. on p. 5).

[35] Yanping Fan et al. «An Advanced Coarse-Grained Nucleosome Core Particle Model for Computer Simulations of Nucleosome-Nucleosome Interactions under Varying Ionic Conditions». In: *PLoS ONE* 8.2 (Feb. 2013). Ed. by Jie Zheng, e54228. DOI: `10.1371/journal.pone.0054228` (cit. on pp. 6, 14, 17, 27, 29, 35).

[36] R. Collepardo-Guevara and T. Schlick. «Chromatin fiber polymorphism triggered by variations of DNA linker lengths». In: *Proceedings of the National Academy of Sciences* 111.22 (May 2014), pp. 8061–8066. DOI: `10.1073/pnas.1315872111` (cit. on pp. 6–9, 17, 34, 68).

[37] Saeed Izadi, Ramu Anandakrishnan, and Alexey V. Onufriev. «Implicit Solvent Model for Million-Atom Atomistic Simulations: Insights into the Organization of 30-nm Chromatin Fiber». In: *Journal of Chemical Theory and Computation* 12.12 (Nov. 2016), pp. 5946–5959. DOI: `10.1021/acs.jctc.6b00712` (cit. on pp. 6, 17, 28, 31, 34, 68).

[38] Surya K. Ghosh and Daniel Jost. «How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes». In: *PLOS Computational Biology* 14.5 (May 2018). Ed. by Sheng Zhong, e1006159. DOI: `10.1371/journal.pcbi.1006159` (cit. on pp. 6, 12, 17).

[39] Alexey Savelyev, Christopher K. Materese, and Garegin A. Papoian. «Is DNA's Rigidity Dominated by Electrostatic or Nonelectrostatic Interactions?» In: *Journal of the American Chemical Society* 133.48 (Dec. 2011), pp. 19290–19293. DOI: `10.1021/ja207984z` (cit. on p. 6).

[40] Luca Giorgetti et al. «Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription». In: *Cell* 157.4 (May 2014), pp. 950–963. DOI: `10.1016/j.cell.2014.03.025` (cit. on pp. 6, 9, 17).

[41] Simona Bianco et al. «Predicting chromatin architecture from models of polymer physics». In: *Chromosome Research* 25.1 (Jan. 2017), pp. 25–34. DOI: `10.1007/s10577-016-9545-5` (cit. on pp. 6, 17).

[42] Maxim V. Imakaev, Geoffrey Fudenberg, and Leonid A. Mirny. «Modeling chromosomes: Beyond pretty pictures». In: *FEBS Letters* 589.20PartA (Sept. 2015), pp. 3031–3036. DOI: `10.1016/j.febslet.2015.09.004` (cit. on p. 6).

[43] Geoffrey Fudenberg et al. «Formation of Chromosomal Domains by Loop Extrusion». In: *Cell Reports* 15.9 (May 2016), pp. 2038–2049. DOI: `10.1016/j.celrep.2016.04.085` (cit. on pp. 6, 11, 17).

[44] M. Tompitak. «The mechanical genome : inquiries into the mechanical function of genetic information». PhD thesis. Leiden University, 2017 (cit. on p. 6).

[45] Joshua Lequieu et al. «1CPN: A coarse-grained multi-scale model of chromatin». In: *The Journal of Chemical Physics* 150.21 (June 2019), p. 215102. DOI: `10.1063/1.5092976` (cit. on p. 6).

[46] Olga I. Kulaeva et al. «Internucleosomal Interactions Mediated by Histone Tails Allow Distant Communication in Chromatin». In: *Journal of Biological Chemistry* 287.24 (Apr. 2012), pp. 20248–20257. DOI: `10.1074/jbc.m111.333104` (cit. on pp. 6, 9, 52).

[47] Aurélien Bancaud et al. «Nucleosome Chiral Transition under Positive Torsional Stress in Single Chromatin Fibers». In: *Molecular Cell* 27.1 (July 2007), pp. 135–147. DOI: `10.1016/j.molcel.2007.05.037` (cit. on pp. 7, 28).

[48] 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». In: *Nucleic Acids Research* 37.9 (Mar. 2009), pp. 2818–2829. DOI: `10.1093/nar/gkp113` (cit. on p. 7).

[49] Leonid A. Mirny. «The fractal globule as a model of chromatin architecture in the cell». In: *Chromosome Research* 19.1 (Jan. 2011), pp. 37–51. DOI: `10.1007/s10577-010-9177-0` (cit. on pp. 8, 12).

[50] Georgi V. Pachov, Razif R. Gabdoulline, and Rebecca C. Wade. «On the structure and dynamics of the complex of the nucleosome and the linker histone». In: *Nucleic Acids Research* 39.12 (Feb. 2011), pp. 5255–5263. DOI: `10.1093/nar/gkr101` (cit. on p. 7).

[51] Rosana Collepardo-Guevara and Tamar Schlick. «Crucial role of dynamic linker histone binding and divalent ions for DNA accessibility and gene regulation revealed by mesoscale modeling of oligonucleosomes». In: *Nucleic Acids Research* 40.18 (July 2012), pp. 8803–8817. DOI: `10.1093/nar/gks600` (cit. on pp. 7, 28).

[52] Pablo D. Dans et al. «How accurate are accurate force-fields for B-DNA?» In: *Nucleic Acids Research* (Jan. 2017), gkw1355. DOI: `10.1093/nar/gkw1355` (cit. on p. 7).

[53] Rodrigo Galindo-Murillo et al. «Assessing the Current State of Amber Force Field Modifications for DNA». In: *Journal of Chemical Theory and Computation* 12.8 (July 2016), pp. 4114–4127. DOI: `10.1021/acs.jctc.6b00186` (cit. on p. 7).

[54] S. A. Grigoryev et al. «Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions». In: *Proceedings of the National Academy of Sciences* 106.32 (July 2009), pp. 13317–13322. DOI: `10.1073/pnas.0903280106` (cit. on pp. 8, 9, 17, 27).

[55] Tamar Schlick and Ognjen Perišić. «Mesoscale simulations of two nucleosome-repeat length oligonucleosomes». In: *Physical Chemistry Chemical Physics* 11.45 (2009), p. 10729. DOI: `10.1039/b918629h` (cit. on p. 8).

[56] P. J. J. Robinson et al. «EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure». In: *Proceedings of the National Academy of Sciences* 103.17 (Apr. 2006), pp. 6506–6511. DOI: `10.1073/pnas.0601212103` (cit. on p. 8).

[57] R. Garcés, R. Podgornik, and V. Lorman. «Antipolar and Anticlinic Mesophase Order in Chromatin Induced by Nucleosome Polarity and Chirality Correlations». In: *Physical Review Letters* 114.23 (June 2015). DOI: `10.1103/physrevlett.114.238102` (cit. on p. 8).

[58] Nick Gilbert et al. «Chromatin Architecture of the Human Genome». In: *Cell* 118.5 (Sept. 2004), pp. 555–566. DOI: `10.1016/j.cell.2004.08.011` (cit. on p. 9).

[59] David Winogradoff et al. «The Acetylation Landscape of the H4 Histone Tail: Disentangling the Interplay between the Specific and Cumulative Effects». In: *Journal of the American Chemical Society* 137.19 (May 2015), pp. 6245–6253. DOI: `10.1021/jacs.5b00235` (cit. on pp. 9, 35, 54).

[60] Philipp M. Diesinger and Dieter W. Heermann. «Depletion Effects Massively Change Chromatin Properties and Influence Genome Folding». In: *Biophysical Journal* 97.8 (Oct. 2009), pp. 2146–2153. DOI: `10.1016/j.bpj.2009.06.057` (cit. on pp. 9, 17, 62).

[61] Philipp M. Diesinger et al. «Histone Depletion Facilitates Chromatin Loops on the Kilobasepair Scale». In: *Biophysical Journal* 99.9 (Nov. 2010), pp. 2995–3001. DOI: `10.1016/j.bpj.2010.08.039` (cit. on pp. 9, 17).

[62] Nick Kepper et al. «Nucleosome Geometry and Internucleosomal Interactions Control the Chromatin Fiber Conformation». In: *Biophysical Journal* 95.8 (Oct. 2008), pp. 3692–3705. DOI: `10.1529/biophysj.107.121079` (cit. on pp. 9, 10, 17).

[63] Elena F. Koslover et al. «Local Geometry and Elasticity in Compact Chromatin Structure». In: *Biophysical Journal* 99.12 (Dec. 2010), pp. 3941–3950. DOI: `10.1016/j.bpj.2010.10.024` (cit. on pp. 9, 17).

[64] Elena F. Koslover and Andrew J. Spakowitz. «Twist- and Tension-Mediated Elastic Coupling between DNA-Binding Proteins». In: *Physical Review Letters* 102.17 (Apr. 2009). DOI: `10.1103/physrevlett.102.178102` (cit. on p. 9).

[65] J. Langowski. «Polymer chain models of DNA and chromatin». In: *The European Physical Journal E* 19.3 (Mar. 2006), pp. 241–249. DOI: `10.1140/epje/i2005-10067-9` (cit. on p. 9).

[66] Gero Wedemann and Jörg Langowski. «Computer Simulation of the 30-Nanometer Chromatin Fiber». In: *Biophysical Journal* 82.6 (June 2002), pp. 2847–2859. DOI: `10.1016/s0006-3495(02)75627-0` (cit. on pp. 9, 62).

[67] MacPherson et al. «A Polymer Physics Model for Epigenetic Control of Chromatin Compaction». In: *Proc. of the 62nd Biophysical Society Annual Meeting*. 2018, 563a (cit. on p. 9).

[68] Helmut Schiessel, William M. Gelbart, and Robijn Bruinsma. «DNA Folding: Structural and Mechanical Properties of the Two-Angle Model for Chromatin». In: *Biophysical Journal* 80.4 (Apr. 2001), pp. 1940–1956. DOI: `10.1016/s0006-3495(01)76164-4` (cit. on p. 9).

[69] René Stehr et al. «The Effect of Internucleosomal Interaction on Folding of the Chromatin Fiber». In: *Biophysical Journal* 95.8 (Oct. 2008), pp. 3677–3691. DOI: `10.1529/biophysj.107.120543` (cit. on pp. 9, 17, 62).

[70] Y. Cui and C. Bustamante. «Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure». In: *Proceedings of the National Academy of Sciences* 97.1 (Jan. 2000), pp. 127–132. DOI: `10.1073/pnas.97.1.127` (cit. on p. 10).

[71] F. A. Manna et al. «Polarity and Chirality in NCP Mesophases and Chromatin Fibers». In: *Molecular Crystals and Liquid Crystals* 478.1 (Dec. 2007), 83/[839]–97/[853]. DOI: `10.1080/15421400701686918` (cit. on p. 10).

[72] Annick Lesne et al. «3D genome reconstruction from chromosomal contacts». In: *Nature Methods* 11.11 (Sept. 2014), pp. 1141–1143. DOI: `10.1038/nmeth.3104` (cit. on p. 10).

[73] N. Varoquaux et al. «A statistical approach for inferring the 3D structure of the genome». In: *Bioinformatics* 30.12 (June 2014), pp. i26–i33. DOI: `10.1093/bioinformatics/btu268` (cit. on p. 10).

[74] Alexandra Gesine Cauer et al. «Inferring diploid 3D chromatin structures from Hi-C data». In: *bioRxiv* (May 2019). DOI: `10.1101/644294` (cit. on pp. 10, 17).

[75] Masae Ohno et al. «Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs». In: *Cell* 176.3 (Jan. 2019), 520–534.e25. DOI: `10.1016/j.cell.2018.12.014` (cit. on pp. 10, 17).

[76] Oliver Wiese, Davide Marenduzzo, and Chris A. Brackley. «Nucleosome positions alone can be used to predict domains in yeast chromosomes». In: *Proceedings of the National Academy of Sciences* 116.35 (Aug. 2019), pp. 17307–17315. DOI: `10.1073/pnas.1817829116` (cit. on p. 10).

[77] Anton Goloborodko, John F. Marko, and Leonid A. Mirny. «Chromosome Compaction by Active Loop Extrusion». In: *Biophysical Journal* 110.10 (May 2016), pp. 2162–2168. DOI: `10.1016/j.bpj.2016.02.041` (cit. on p. 11).

[78] Gavin D. Bascom, Karissa Y. Sanbonmatsu, and Tamar Schlick. «Mesoscale Modeling Reveals Hierarchical Looping of Chromatin Fibers Near Gene Regulatory Elements». In: *The Journal of Physical Chemistry B* 120.33 (June 2016), pp. 8642–8653. DOI: `10.1021/acs.jpcb.6b03197` (cit. on pp. 11, 17).

[79] Edward J. Banigan et al. «Chromosome organization by one-sided and two-sided loop extrusion». In: *bioRxiv* (Oct. 2019). DOI: `10.1101/815340` (cit. on p. 11).

[80] Nick Gilbert and Davide Marenduzzo. «Genome organization: experiments and modeling». In: *Chromosome Research* 25.1 (Feb. 2017), pp. 1–4. DOI: `10.1007/s10577-017-9551-2` (cit. on p. 12).

[81] E. Lieberman-Aiden et al. «Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome». In: *Science* 326.5950 (Oct. 2009), pp. 289–293. DOI: `10.1126/science.1181369` (cit. on pp. 12, 15, 17).

[82] A. Yu. Grosberg, S.K. Nechaev, and E.I. Shakhnovich. «The role of topological constraints in the kinetics of collapse of macromolecules». In: *Journal de Physique* 49.12 (1988), pp. 2095–2100. DOI: `10.1051/jphys:0198800490120209500` (cit. on p. 12).

[83] M. V. Tamm et al. «Anomalous Diffusion in Fractal Globules». In: *Physical Review Letters* 114.17 (Apr. 2015). DOI: `10.1103/physrevlett.114.178102` (cit. on p. 12).

[84] Suhas S.P. Rao et al. «A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping». In: *Cell* 159.7 (Dec. 2014), pp. 1665–1680. DOI: `10.1016/j.cell.2014.11.021` (cit. on p. 12).

[85] Andrey V Ilatovskiy et al. «SANS spectra of the fractal supernucleosomal chromatin structure models». In: *Journal of Physics: Conference Series* 351 (Mar. 2012), p. 012007. DOI: `10.1088/1742-6596/351/1/012007` (cit. on pp. 12, 15).

[86] L. M. Almassalha et al. «The Global Relationship between Chromatin Physical Topology, Fractal Structure, and Gene Expression». In: *Scientific Reports* 7.1 (Jan. 2017). DOI: `10.1038/srep41061` (cit. on p. 12).

[87] A. Bancaud et al. «A fractal model for nuclear organization: current evidence and biological implications». In: *Nucleic Acids Research* 40.18 (July 2012), pp. 8783–8792. DOI: `10.1093/nar/gks586` (cit. on p. 12).

[88] Karolin Luger et al. «Crystal structure of the nucleosome core particle at 2.8 AA resolution». In: *Nature* 389.6648 (Sept. 1997), pp. 251–260. DOI: `10.1038/38444` (cit. on pp. 13, 17).

[89] J Widom. «Chromatin: The nucleosome unwrapped». In: *Current Biology* 7.10 (Oct. 1997), R653–R655. DOI: `10.1016/s0960-9822(06)00327-7` (cit. on pp. 13, 35).

[90] Curt A. Davey et al. «Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution». In: *Journal of Molecular Biology* 319.5 (June 2002), pp. 1097–1113. DOI: `10.1016/s0022-2836(02)00386-8` (cit. on pp. 13, 17, 29, 31, 32, 36, 62, 72, 74).

[91] H. Tachiwana et al. «Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T». In: *Proceedings of the National Academy of Sciences* 107.23 (May 2010), pp. 10454–10459. DOI: `10.1073/pnas.1003064107` (cit. on pp. 13, 16, 39).

[92] Yoshimasa Takizawa et al. «Cryo-EM structure of the nucleosome containing the ALB1 enhancer DNA sequence». In: *Open Biology* 8.3 (Mar. 2018), p. 170255. DOI: `10.1098/rsob.170255` (cit. on p. 13).

[93] Wataru Kobayashi et al. «Structural and biochemical analyses of the nuclear pore complex component ELYS identify residues responsible for nucleosome binding». In: *Communications Biology* 2.1 (May 2019). DOI: `10.1038/s42003-019-0385-7` (cit. on pp. 13, 14).

[94] Silvija Bilokapic, Mike Strauss, and Mario Halic. «Cryo-EM of nucleosome core particle interactions in trans». In: *Scientific Reports* 8.1 (May 2018). DOI: `10.1038/s41598-018-25429-1` (cit. on p. 13).

[95] Nina A. Filenko, Dmytro B. Palets, and Yuri L. Lyubchenko. «Structure and Dynamics of Dinucleosomes Assessed by Atomic Force Microscopy». In: *Journal of Amino Acids* 2012 (2012), pp. 1–6. DOI: `10.1155/2012/650840` (cit. on p. 14).

[96] Katarzyna M. Krzemien et al. «Atomic force microscopy of chromatin arrays reveal non-monotonic salt dependence of array compaction in solution». In: *PLOS ONE* 12.3 (Mar. 2017). Ed. by Yamini Dalal, e0173459. DOI: `10.1371/journal.pone.0173459` (cit. on pp. 14, 15).

[97] Aymeric Le Gratiet et al. «Circular intensity differential scattering (CIDS) scanning microscopy to image chromatin-DNA nuclear organization». In: *OSA Continuum* 1.3 (Nov. 2018), p. 1068. DOI: `10.1364/osac.1.001068` (cit. on pp. 14, 15, 17).

[98] Babatunde Ekundayo, Timothy J. Richmond, and Thomas Schalch. «Capturing Structural Heterogeneity in Chromatin Fibers». In: *Journal of Molecular Biology* 429.20 (Oct. 2017), pp. 3031–3042. DOI: `10.1016/j.jmb.2017.09.002` (cit. on p. 13).

[99] Thomas Schalch et al. «X-ray structure of a tetranucleosome and its implications for the chromatin fibre». In: *Nature* 436.7047 (July 2005), pp. 138–141. DOI: `10.1038/nature03686` (cit. on p. 13).

[100] Bing-Rui Zhou et al. «Histone H4 K16Q Mutation, an Acetylation Mimic, Causes Structural Disorder of Its N-Terminal Basic Patch in the Nucleosome». In: *Journal of Molecular Biology* 421.1 (Aug. 2012), pp. 30–37. DOI: `10.1016/j.jmb.2012.04.032` (cit. on pp. 13, 35).

[101] Min Gao et al. «Histone H3 and H4 N-Terminal Tails in Nucleosome Arrays at Cellular Concentrations Probed by Magic Angle Spinning NMR Spectroscopy». In: *Journal of the American Chemical Society* 135.41 (Oct. 2013), pp. 15278–15281. DOI: `10.1021/ja407526s` (cit. on pp. 13, 35).

[102] Aurélie Bertin et al. «Role of Histone Tails in the Conformation and Interactions of Nucleosome Core Particles†». In: *Biochemistry* 43.16 (Apr. 2004), pp. 4773–4780. DOI: `10.1021/bi036210g` (cit. on pp. 14, 17, 35).

[103] Alexander W. Mauney et al. «Local DNA Sequence Controls Asymmetry of DNA Unwrapping from Nucleosome Core Particles». In: *Biophysical Journal* 115.5 (Sept. 2018), pp. 773–781. DOI: `10.1016/j.bpj.2018.07.009` (cit. on pp. 14, 35).

[104] Davood Norouzi and Victor B. Zhurkin. «Dynamics of Chromatin Fibers: Comparison of Monte Carlo Simulations with Force Spectroscopy». In: *Biophysical Journal* 115.9 (Nov. 2018), pp. 1644–1655. DOI: `10.1016/j.bpj.2018.06.032` (cit. on pp. 14, 17, 35).

[105] Steven C. Howell. «Dynamic Conformations of Nucleosome Arrays in Solution from Small-Angle X-ray Scattering». PhD thesis. George Washington University, 2016 (cit. on p. 14).

[106] Hin Hark Gan and Tamar Schlick. «Chromatin Ionic Atmosphere Analyzed by a Mesoscale Electrostatic Approach». In: *Biophysical Journal* 99.8 (Oct. 2010), pp. 2587–2596. DOI: `10.1016/j.bpj.2010.08.023` (cit. on pp. 15, 17, 27, 29, 34, 68).

[107] Noa Plavner Hazan et al. «Nucleosome Core Particle Disassembly and Assembly Kinetics Studied Using Single-Molecule Fluorescence». In: *Biophysical Journal* 109.8 (Oct. 2015), pp. 1676–1685. DOI: `10.1016/j.bpj.2015.07.004` (cit. on p. 15).

[108] Marco Pasi, John H. Maddocks, and Richard Lavery. «Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics». In: *Nucleic Acids Research* 43.4 (Feb. 2015), pp. 2412–2423. DOI: `10.1093/nar/gkv080` (cit. on p. 15).

[109] Marco Pasi et al. «Analyzing DNA curvature and its impact on the ionic environment: application to molecular dynamics simulations of minicircles». In: *Nucleic Acids Research* 45.7 (Feb. 2017), pp. 4269–4277. DOI: `10.1093/nar/gkx092` (cit. on p. 15).

[110] Remo Rohs et al. «The role of DNA shape in protein–DNA recognition». In: *Nature* 461.7268 (Oct. 2009), pp. 1248–1253. DOI: `10.1038/nature08473` (cit. on pp. 15, 17, 29).

[111] Maria Aurelia Ricci et al. «Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo». In: *Cell* 160.6 (Mar. 2015), pp. 1145–1158. DOI: `10.1016/j.cell.2015.01.054` (cit. on pp. 15, 17).

[112] Shujun Cai et al. «Cryo-ET reveals the macromolecular reorganization of S. pombe mitotic chromosomes in vivo». In: *Proceedings of the National Academy of Sciences* 115.43 (Oct. 2018), pp. 10977–10982. DOI: `10.1073/pnas.1720476115` (cit. on p. 16).

[113] Yoshinori Nishino et al. «Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure». In: *The EMBO Journal* 31.7 (Feb. 2012), pp. 1644–1653. DOI: `10.1038/emboj.2012.35` (cit. on p. 16).

[114] Kazuhiro Maeshima et al. «Liquid-like behavior of chromatin». In: *Current Opinion in Genetics & Development* 37 (Apr. 2016), pp. 36–45. DOI: `10.1016/j.gde.2015.11.006` (cit. on p. 16).

[115] Kazuhiro Maeshima et al. «Chromatin as dynamic 10-nm fibers». In: *Chromosoma* 123.3 (Apr. 2014), pp. 225–237. DOI: `10.1007/s00412-014-0460-2` (cit. on p. 16).

[116] C.A. Davey et al. *X-Ray Structure of the Nucleosome Core Particle, NCP147, at 1.9 A Resolution.* Dec. 2002. DOI: `10.2210/pdb1kx5/pdb` (cit. on pp. 16, 28).

[117] Jan Bednar et al. «Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1». In: *Molecular Cell* 66.3 (May 2017), 384–397.e8. DOI: `10.1016/j.molcel.2017.04.012` (cit. on p. 16).

[118] Cedric R. Clapier et al. «Structure of theDrosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer». In: *Proteins: Structure, Function, and Bioinformatics* 71.1 (2008), pp. 1–7. DOI: `10.1002/prot.21720` (cit. on pp. 16, 75).

[119] Christopher K. Materese, Alexey Savelyev, and Garegin A. Papoian. «Counterion Atmosphere and Hydration Patterns near a Nucleosome Core Particle». In: *Journal of the American Chemical Society* 131.41 (Oct. 2009), pp. 15005–15013. DOI: `10.1021/ja905376q` (cit. on pp. 17, 31, 32, 34, 59, 62, 68).

[120] Peter Debye and Erich Hückel. «Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen». In: *Physikalische Zeitschrift* 24.185 (1923), p. 305 (cit. on p. 20).

[121] Michael J. Holst. *The Poisson-Boltzmann Equation: Analysis and Multilevel Numerical Solution.* 1994 (cit. on pp. 21, 22).

[122] W. Rocchia. «Poisson-boltzmann equation boundary conditions for biological applications». In: *Mathematical and Computer Modelling* 41.10 (May 2005), pp. 1109–1118. DOI: `10.1016/j.mcm.2005.05.006` (cit. on p. 24).

[123] Barbara Abraham-Shrauner. «Nonlinear poisson-Boltzmann potential for a uniformly charged dielectric sphere in an electrolyte». In: *Journal of Colloid and Interface Science* 44.1 (July 1973), pp. 79–84. DOI: `10.1016/0021-9797(73)90194-x` (cit. on pp. 24, 68).

[124] Eytan Barouch and Sanjeev Kulkarni. «Exact solution of the Poisson-Boltzmann equation for two spheres with fixed surface potentials». In: *Journal of Colloid and Interface Science* 112.2 (Aug. 1986), pp. 396–402. DOI: `10.1016/0021-9797(86)90107-4` (cit. on pp. 24, 68).

[125] John Lekner. «Electrostatics of two charged conducting spheres». In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 468.2145 (May 2012), pp. 2829–2848. DOI: `10.1098/rspa.2012.0133` (cit. on p. 24).

[126] A.B Glendinning and W.B Russel. «The electrostatic repulsion between charged spheres from exact solutions to the linearized poisson-boltzmann equation». In: *Journal of Colloid and Interface Science* 93.1 (May 1983), pp. 95–104. DOI: `10.1016/0021-9797(83)90388-0` (cit. on p. 24).

[127] Joseph E. Ledbetter, Thomas L. Croxton, and Donald A. McQuarrie. «The interaction of two charged spheres in the Poisson–Boltzmann equation». In: *Canadian Journal of Chemistry* 59.13 (July 1981), pp. 1860–1864. DOI: `10.1139/v81-277` (cit. on p. 24).

[128] Ivan N. Derbenev et al. «Electrostatic interactions between charged dielectric particles in an electrolyte solution». In: *The Journal of Chemical Physics* 145.8 (Aug. 2016), p. 084103. DOI: `10.1063/1.4961091` (cit. on pp. 24, 69, 70).

[129] M. Gouy. «Sur la constitution de la charge électrique à la surface d'un électrolyte». In: *Journal de Physique Théorique et Appliquée* 9.1 (1910), pp. 457–468. DOI: `10.1051/jphystap:019100090045700` (cit. on p. 24).

[130] David Leonard Chapman. «LI. A contribution to the theory of electrocapillarity». In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 25.148 (Apr. 1913), pp. 475–481. DOI: `10.1080/14786440408634187` (cit. on p. 24).

[131] W. Rocchia, E. Alexov, and B. Honig. «Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions†». In: *The Journal of Physical Chemistry B* 105.28 (July 2001), pp. 6507–6514. DOI: `10.1021/jp010454y` (cit. on pp. 24, 26, 32, 34, 37, 62, 64, 68).

[132] N. A. Baker et al. «Electrostatics of nanosystems: Application to microtubules and the ribosome». In: *Proceedings of the National Academy of Sciences* 98.18 (Aug. 2001), pp. 10037–10041. DOI: `10.1073/pnas.181342398` (cit. on pp. 24, 34).

[133] Claudia N. Schutz and Arieh Warshel. «What are the dielectric ?constants? of proteins and how to validate electrostatic models?» In: *Proteins: Structure, Function, and Genetics* 44.4 (2001), pp. 400–417. DOI: `10.1002/prot.1106` (cit. on p. 24).

[134] M. L. Connolly. «Analytical molecular surface calculation». In: *Journal of Applied Crystallography* 16.5 (Oct. 1983), pp. 548–558. DOI: `10.1107/s0021889883010985` (cit. on p. 25).

[135] Sergio Decherchi and Walter Rocchia. «A general and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale». In: *PLoS ONE* 8.4 (Apr. 2013). Ed. by Chris Lorenz, e59744. DOI: `10.1371/journal.pone.0059744` (cit. on pp. 25, 30, 32, 37, 61, 62, 73, 75).

[136] Anatoly Zinchenko et al. «Single-molecule compaction of megabase-long chromatin molecules by multivalent cations». In: *Nucleic Acids Research* 46.2 (Nov. 2017), pp. 635–649. DOI: `10.1093/nar/gkx1135` (cit. on p. 27).

[137] Alexey K. Shaytan et al. «Trajectories of microsecond molecular dynamics simulations of nucleosomes and nucleosome core particles». In: *Data in Brief* 7 (June 2016), pp. 1678–1681. DOI: `10.1016/j.dib.2016.04.073` (cit. on pp. 28, 35, 36, 72, 74, 115).

[138] Nikolay Korolev et al. «A Coarse-Grained DNA Model Parameterized from Atomistic Simulations by Inverse Monte Carlo». In: *Polymers* 6.6 (May 2014), pp. 1655–1675. DOI: `10.3390/polym6061655` (cit. on p. 29).

[139] Keda Zhou, Guillaume Gaullier, and Karolin Luger. «Nucleosome structure and dynamics are coming of age». In: *Nature Structural & Molecular Biology* 26.1 (Dec. 2018), pp. 3–13. DOI: `10.1038/s41594-018-0166-x` (cit. on pp. 29, 68).

[140] F. X. Wilhelm et al. «Reconstitution of chromatin: assembly of the nucleosome». In: *Nucleic Acids Research* 5.2 (1978), pp. 505–521. DOI: `10.1093/nar/5.2.505` (cit. on pp. 29, 58, 63, 68).

[141] Magdalena Gebala et al. «Ion counting demonstrates a high electrostatic field generated by the nucleosome». In: *eLife* 8 (June 2019). DOI: `10.7554/elife.44993` (cit. on pp. 30, 59).

[142] Rudi Podgornik, Donald C. Rau, and V. Adrian Parsegian. «The action of interhelical forces on the organization of DNA double helixes: fluctuation-enhanced decay of electrostatic double-layer and hydration forces». In: *Macromolecules* 22.4 (July 1989), pp. 1780–1786. DOI: `10.1021/ma00194a048` (cit. on p. 31).

[143] R. Podgornik, D.C. Rau, and V.A. Parsegian. «Parametrization of direct and soft steric-undulatory forces between DNA double helical polyelectrolytes in solutions of several different anions and cations». In: *Biophysical Journal* 66.4 (Apr. 1994), pp. 962–971. DOI: `10.1016/s0006-3495(94)80877-x` (cit. on p. 31).

[144] Julija Zavadlav, Rudolf Podgornik, and Matej Praprotnik. «Order and interactions in DNA arrays: Multiscale molecular dynamics simulation». In: *Scientific Reports* 7.1 (July 2017). DOI: `10.1038/s41598-017-05109-2` (cit. on p. 31).

[145] A. Cuervo et al. «Direct measurement of the dielectric polarization properties of DNA». In: *Proceedings of the National Academy of Sciences* 111.35 (Aug. 2014), E3624–E3630. DOI: `10.1073/pnas.1405702111` (cit. on p. 31).

[146] Mary L. Kopka et al. «Ordered water structure around a B-DNA dodecamer». In: *Journal of Molecular Biology* 163.1 (Jan. 1983), pp. 129–146. DOI: `10.1016/0022-2836(83)90033-5` (cit. on p. 32).

[147] Bohdan Schneider, Ketan Patel, and Helen M. Berman. «Hydration of the Phosphate Group in Double-Helical DNA». In: *Biophysical Journal* 75.5 (Nov. 1998), pp. 2422–2434. DOI: `10.1016/s0006-3495(98)77686-6` (cit. on p. 32).

[148] Martin Egli et al. «X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution». In: *Biopolymers* 48.4 (1998), p. 234. DOI: `10.1002/(sici)1097-0282(1998)48:4<234::aid-bip4>3.0.co;2-h` (cit. on p. 32).

[149] Sergio Decherchi et al. «NanoShaper–VMD interface: computing and visualizing surfaces, pockets and channels in molecular systems». In: *Bioinformatics* 35.7 (Aug. 2018). Ed. by Alfonso Valencia, pp. 1241–1243. DOI: `10.1093/bioinformatics/bty761` (cit. on pp. 32, 36, 62, 73).

[150] Matsuyuki Shirota, Takashi Ishida, and Kengo Kinoshita. «Effects of surface-to-volume ratio of proteins on hydrophilic residues: Decrease in occurrence and increase in buried fraction». In: *Protein Science* 17.9 (Sept. 2008), pp. 1596–1602. DOI: `10.1110/ps.035592.108` (cit. on pp. 32, 62).

[151] Artemi Bendandi et al. «The role of histone tails in nucleosome stability: An electrostatic perspective». In: *Computational and Structural Biotechnology Journal* 18 (2020), pp. 2799–2809. DOI: `10.1016/j.csbj.2020.09.034` (cit. on p. 33).

[152] Daniel A. Beard and Tamar Schlick. «Modeling salt-mediated electrostatics of macromolecules: The discrete surface charge optimization algorithm and its application to the nucleosome». In: *Biopolymers* 58.1 (2001), pp. 106–115. DOI: `10.1002/1097-0282(200101)58:1<106::AID-BIP100>3.0.CO;2-\#` (cit. on p. 34).

[153] Alessandro Samuele Patelli. «A sequence-dependent coarse-grain model of B-DNA with explicit description of bases and phosphate groups parametrised from large scale Molecular Dynamics simulations». In: (2019), p. 261. DOI: `10.5075/epfl-thesis-9552` (cit. on p. 34).

[154] Kurt Andresen et al. «Solution Scattering and FRET Studies on Nucleosomes Reveal DNA Unwrapping Effects of H3 and H4 Tail Removal». In: *PLoS ONE* 8.11 (Nov. 2013). Ed. by Martin Fernandez-Zapico, e78587. DOI: `10.1371/journal.pone.0078587` (cit. on pp. 35, 50, 54).

[155] Nils B. Becker and Ralf Everaers. «DNA Nanomechanics in the Nucleosome». In: *Structure* 17.4 (Apr. 2009), pp. 579–589. DOI: `10.1016/j.str.2009.01.013` (cit. on pp. 36, 41, 54).

[156] T. J. Dolinsky et al. «PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations». In: *Nucleic Acids Research* 32.Web Server (July 2004), W665–W667. DOI: `10.1093/nar/gkh381` (cit. on pp. 36, 73).

[157] P. J. A. Cock et al. «Biopython: freely available Python tools for computational molecular biology and bioinformatics». In: *Bioinformatics* 25.11 (Mar. 2009), pp. 1422–1423. DOI: `10.1093/bioinformatics/btp163` (cit. on p. 38).

[158] Rajeswari S. Edayathumangalam et al. «Nucleosomes in Solution Exist as a Mixture of Twist-defect States». In: *Journal of Molecular Biology* 345.1 (Jan. 2005), pp. 103–114. DOI: `10.1016/j.jmb.2004.10.012` (cit. on p. 39).

[159] Jing Jin et al. «Synergistic action of RNA polymerases in overcoming the nucleosomal barrier». In: *Nature Structural & Molecular Biology* 17.6 (May 2010), pp. 745–752. DOI: `10.1038/nsmb.1798` (cit. on pp. 40, 49).

[160] Wakana Iwasaki et al. «Contribution of histone N-terminal tails to the structure and stability of nucleosomes». In: *FEBS Open Bio* 3.1 (Jan. 2013), pp. 363–369. DOI: `10.1016/j.fob.2013.08.007` (cit. on pp. 47, 51).

[161] Hidetoshi Kono, Shun Sakuraba, and Hisashi Ishida. «Free energy profiles for unwrapping the outer superhelical turn of nucleosomal DNA». In: *PLOS Computational Biology* 14.3 (Mar. 2018). Ed. by Anna R. R. Panchenko, e1006024. DOI: `10.1371/journal.pcbi.1006024` (cit. on p. 50).

[162] Brent Brower-Toland et al. «Specific Contributions of Histone Tails and their Acetylation to the Mechanical Stability of Nucleosomes». In: *Journal of Molecular Biology* 346.1 (Feb. 2005), pp. 135–146. DOI: `10.1016/j.jmb.2004.11.056` (cit. on pp. 50, 51).

[163] Hannah S. Tims et al. «Dynamics of Nucleosome Invasion by DNA Binding Proteins». In: *Journal of Molecular Biology* 411.2 (Aug. 2011), pp. 430–448. DOI: `10.1016/j.jmb.2011.05.044` (cit. on p. 51).

[164] H. Ferreira et al. «Histone Tails and the H3 N Helix Regulate Nucleosome Mobility and Stability». In: *Molecular and Cellular Biology* 27.11 (Mar. 2007), pp. 4037–4048. DOI: `10.1128/mcb.02229-06` (cit. on p. 52).

[165] J.-A. Kim et al. «Mutagenesis of pairwise combinations of histone amino-terminal tails reveals functional redundancy in budding yeast». In: *Proceedings of the National Academy of Sciences* 109.15 (Mar. 2012), pp. 5779–5784. DOI: `10.1073/pnas.1203453109` (cit. on p. 52).

[166] Takeru Kameda, Akinori Awazu, and Yuichi Togashi. «Histone Tail Dynamics in Partially Disassembled Nucleosomes During Chromatin Remodeling». In: *Frontiers in Molecular Biosciences* 6 (Nov. 2019). DOI: `10.3389/fmolb.2019.00133` (cit. on p. 52).

[167] D. A. Potoyan and G. A. Papoian. «Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation». In: *Proceedings of the National Academy of Sciences* 109.44 (Sept. 2012), pp. 17857–17862. DOI: `10.1073/pnas.1201805109` (cit. on p. 54).

[168] Xiaoying Wang et al. «Acetylation Increases the $\alpha$-Helical Content of the Histone Tails of the Nucleosome». In: *Journal of Biological Chemistry* 275.45 (Aug. 2000), pp. 35013–35020. DOI: `10.1074/jbc.m004998200` (cit. on p. 54).

[169] J.D Anderson, P.T Lowary, and J Widom. «Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites 1 1Edited by R. Ebright». In: *Journal of Molecular Biology* 307.4 (Apr. 2001), pp. 977–985. DOI: `10.1006/jmbi.2001.4528` (cit. on p. 54).

[170] K.J. Polach, P.T. Lowary, and J. Widom. «Effects of core histone tail domains on the equilibrium constants for dynamic dna site accessibility in nucleosomes». In: *Journal of Molecular Biology* 298.2 (Apr. 2000), pp. 211–223. doi: `10.1006/jmbi.2000.3644` (cit. on p. 54).

[171] Elizaveta Forbes and Andrew Chryss. «Fundamentals of Clays Surface and Colloid Science, and Rheology». In: *Clays in the Minerals Processing Value Chain.* Ed. by Markus Gräfe et al. Cambridge University Press, 2017, pp. 81–110. doi: `10.1017/9781316661888.004` (cit. on p. 60).

[172] Sourav Bhattacharjee. «DLS and zeta potential – What they are and what they are not?» In: *Journal of Controlled Release* 235 (Aug. 2016), pp. 337–351. doi: `10.1016/j.jconrel.2016.06.017` (cit. on p. 59).

[173] Gregory V. Lowry et al. «Guidance to improve the scientific value of zeta-potential measurements in nanoEHS». In: *Environmental Science: Nano* 3.5 (2016), pp. 953–965. doi: `10.1039/c6en00136j` (cit. on p. 59).

[174] S. Mangenot et al. «Interactions between isolated nucleosome core particles: A tail-bridging effect?» In: *The European Physical Journal E* 7.3 (Mar. 2002), pp. 221–231. doi: `10.1140/epje/i200101151` (cit. on p. 63).

[175] Daniel A. Beard and Tamar Schlick. «Modeling salt-mediated electrostatics of macromolecules: The discrete surface charge optimization algorithm and its application to the nucleosome». In: *Biopolymers* 58.1 (2001), pp. 106–115. doi: `https://doi.org/10.1002/1097-0282(200101)58:1<106::AID-BIP100>3.0.CO;2-\#`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0282%28200101%2958%3A1%3C106%3A%3AAID-BIP100%3E3.0.CO%3B2-%23` (cit. on p. 68).

[176] Elena Bichoutskaia et al. «Electrostatic analysis of the interactions between charged particles of dielectric materials». In: *The Journal of Chemical Physics* 133.2 (July 2010), p. 024105. doi: `10.1063/1.3457157` (cit. on p. 68).

[177] Alexey Onufriev, Donald Bashford, and David A. Case. «Modification of the Generalized Born Model Suitable for Macromolecules». In: *The Journal of Physical Chemistry B* 104.15 (Apr. 2000), pp. 3712–3720. doi: `10.1021/jp994072s` (cit. on p. 68).

[178] Alexey Onufriev, Donald Bashford, and David A. Case. «Exploring protein native states and large-scale conformational changes with a modified generalized born model». In: *Proteins: Structure, Function, and Bioinformatics* 55.2 (Mar. 2004), pp. 383–394. doi: `10.1002/prot.20033` (cit. on p. 68).

[179] Grigori Sigalov, Peter Scheffel, and Alexey Onufriev. «Incorporating variable dielectric environments into the generalized Born model». In: *The Journal of Chemical Physics* 122.9 (Mar. 2005), p. 094511. doi: `10.1063/1.1857811` (cit. on p. 68).

[180] Negin Forouzesh, Saeed Izadi, and Alexey V. Onufriev. «Grid-Based Surface Generalized Born Model for Calculation of Electrostatic Binding Free Energies». In: *Journal of Chemical Information and Modeling* 57.10 (Oct. 2017), pp. 2505–2513. DOI: `10.1021/acs.jcim.7b00192` (cit. on p. 69).

[181] John C. Gordon, Andrew T. Fenley, and Alexey Onufriev. «An analytical approach to computing biomolecular electrostatic potential. II. Validation and applications». In: *The Journal of Chemical Physics* 129.7 (Aug. 2008), p. 075102. DOI: `10.1063/1.2956499` (cit. on p. 69).

[182] Ivan N. Derbenev et al. «Electrostatic interactions between spheroidal dielectric particles». In: *The Journal of Chemical Physics* 152.2 (Jan. 2020), p. 024121. DOI: `10.1063/1.5129756` (cit. on pp. 69, 70).

[183] Ivan N. Derbenev et al. «Electrostatic interactions between charged dielectric particles in an electrolyte solution: constant potential boundary conditions». In: *Soft Matter* 14.26 (2018), pp. 5480–5487. DOI: `10.1039/c8sm01068d` (cit. on pp. 69, 70).

[184] Anatoly V. Filippov. «Electrostatic Interaction of Two Point Charges in Equilibrium Plasmas within the Debye Approximation». In: *Contributions to Plasma Physics* 56.5 (Feb. 2016), pp. 380–390. DOI: `10.1002/ctpp.201500122` (cit. on pp. 69, 70).

[185] A. V. Filippov and I. N. Derbenev. «Effect of the size of charged spherical macroparticles on their electrostatic interaction in an equilibrium plasma». In: *Journal of Experimental and Theoretical Physics* 123.6 (Dec. 2016), pp. 1099–1109. DOI: `10.1134/s106377611611008x` (cit. on pp. 69–71).

[186] Michael E. Fisher, Yan Levin, and Xiaojun Li. «The interaction of ions in an ionic medium». In: *The Journal of Chemical Physics* 101.3 (Aug. 1994), pp. 2273–2282. DOI: `10.1063/1.467668` (cit. on pp. 69, 71, 85).

[187] Cyril Dominguez, Rolf Boelens, and Alexandre M. J. J. Bonvin. «HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information». In: *Journal of the American Chemical Society* 125.7 (Feb. 2003), pp. 1731–1737. DOI: `10.1021/ja026939x` (cit. on p. 69).

[188] J. D. Jackson. *Classical Electrodynamics, 3rd Ed.* John Wiley & Sons Ltd., 1999 (cit. on pp. 70, 85, 86).

[189] G. N. Watson. *A Treatise on the Theory of Bessel Functions.* Cambridge University Press, 1966 (cit. on p. 72).

[190] E. W. Hobson. *The Theory of Spherical and Ellipsoidal Harmonics.* Cambridge University Press, 1931 (cit. on p. 72).

[191] Sergio Decherchi et al. «The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning». In: *Nature Communications* 6.1 (Jan. 2015). DOI: `10.1038/ncomms7155` (cit. on pp. 73, 114).

[192] T. Hamelryck and B. Manderick. «PDB file parser and structure class implemented in Python». In: *Bioinformatics* 19.17 (Nov. 2003), pp. 2308–2310. DOI: `10.1093/bioinformatics/btg299` (cit. on p. 74).

[193] M. Johnson et al. «NCBI BLAST: a better web interface». In: *Nucleic Acids Research* 36.Web Server (May 2008), W5–W9. DOI: `10.1093/nar/gkn201` (cit. on p. 75).

[194] W. Kabsch. «A solution for the best rotation to relate two sets of vectors». In: *Acta Crystallographica Section A* 32.5 (Sept. 1976), pp. 922–923. DOI: `10.1107/s0567739476001873` (cit. on p. 75).

[195] Jian Qin. «Charge polarization near dielectric interfaces and the multiple-scattering formalism». In: *Soft Matter* 15.10 (2019), pp. 2125–2134. DOI: `10.1039/c8sm02196a` (cit. on p. 85).

[196] D. J. Griffiths. *Introduction to Electrodynamics, 4rd Ed.* Pearson, 2013 (cit. on pp. 85, 86).

# Appendix A: Supplementary Figures of Chapter 3

## Figure Index

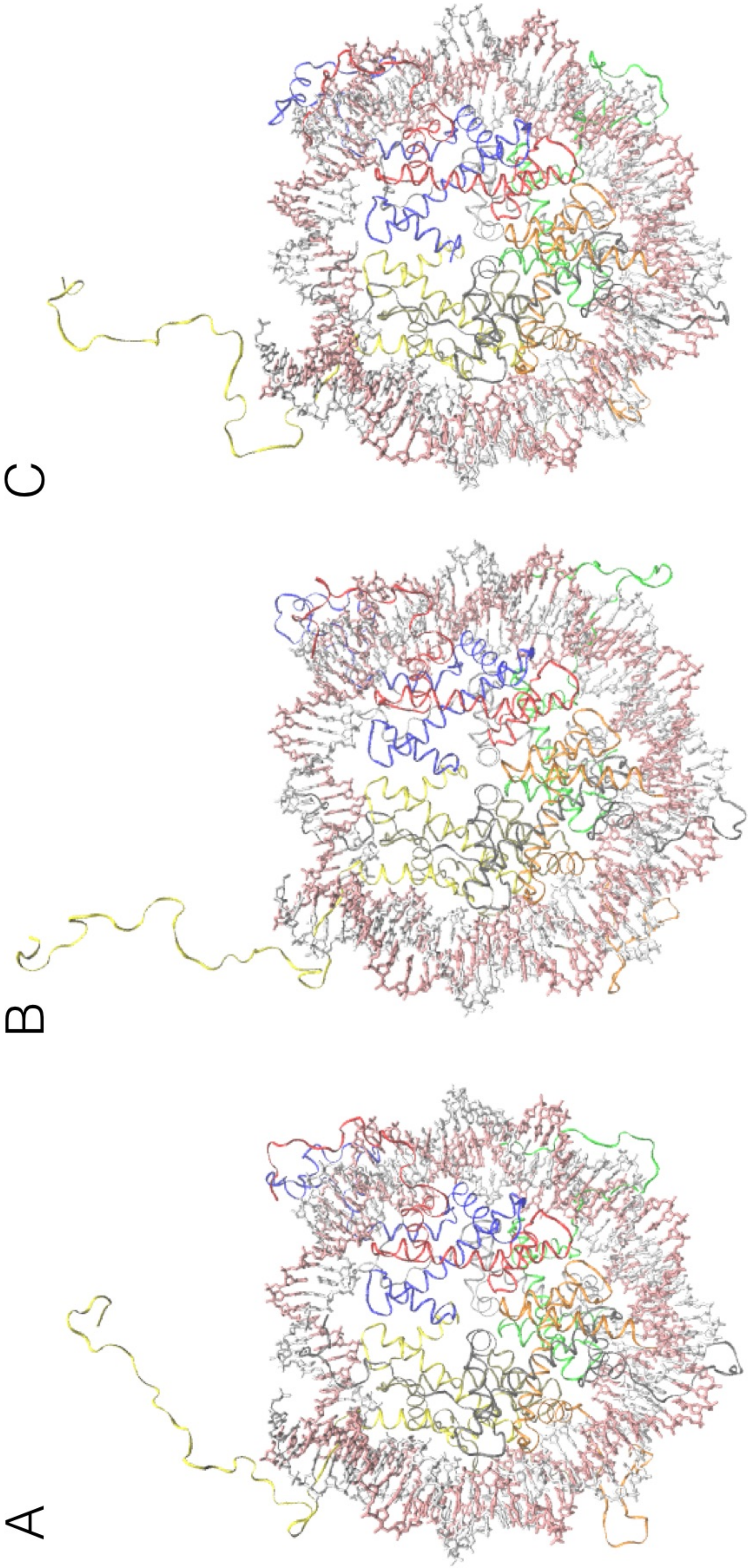
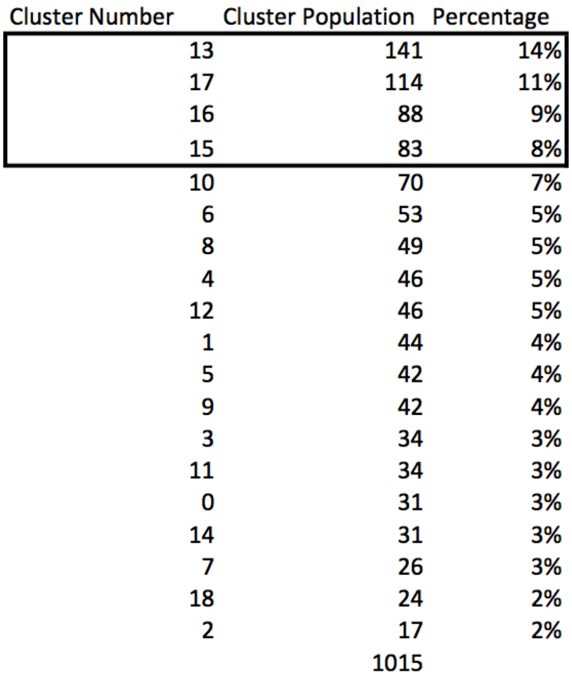


Figure 1: Molecular structures referred to in the text A: medoid15 B: medoid16 C: medoid17.



| Cluster Number | Cluster Population | Percentage |
|---|---|---|
| 13 | 141 | 14% |
| 17 | 114 | 11% |
| 16 | 88 | 9% |
| 15 | 83 | 8% |
| 10 | 70 | 7% |
| 6 | 53 | 5% |
| 8 | 49 | 5% |
| 4 | 46 | 5% |
| 12 | 46 | 5% |
| 1 | 44 | 4% |
| 5 | 42 | 4% |
| 9 | 42 | 4% |
| 3 | 34 | 3% |
| 11 | 34 | 3% |
| 0 | 31 | 3% |
| 14 | 31 | 3% |
| 7 | 26 | 3% |
| 18 | 24 | 2% |
| 2 | 17 | 2% |
|  | 1015 |  |

Figure 2: Cluster populations and percentage of structures belonging to each cluster. [191]

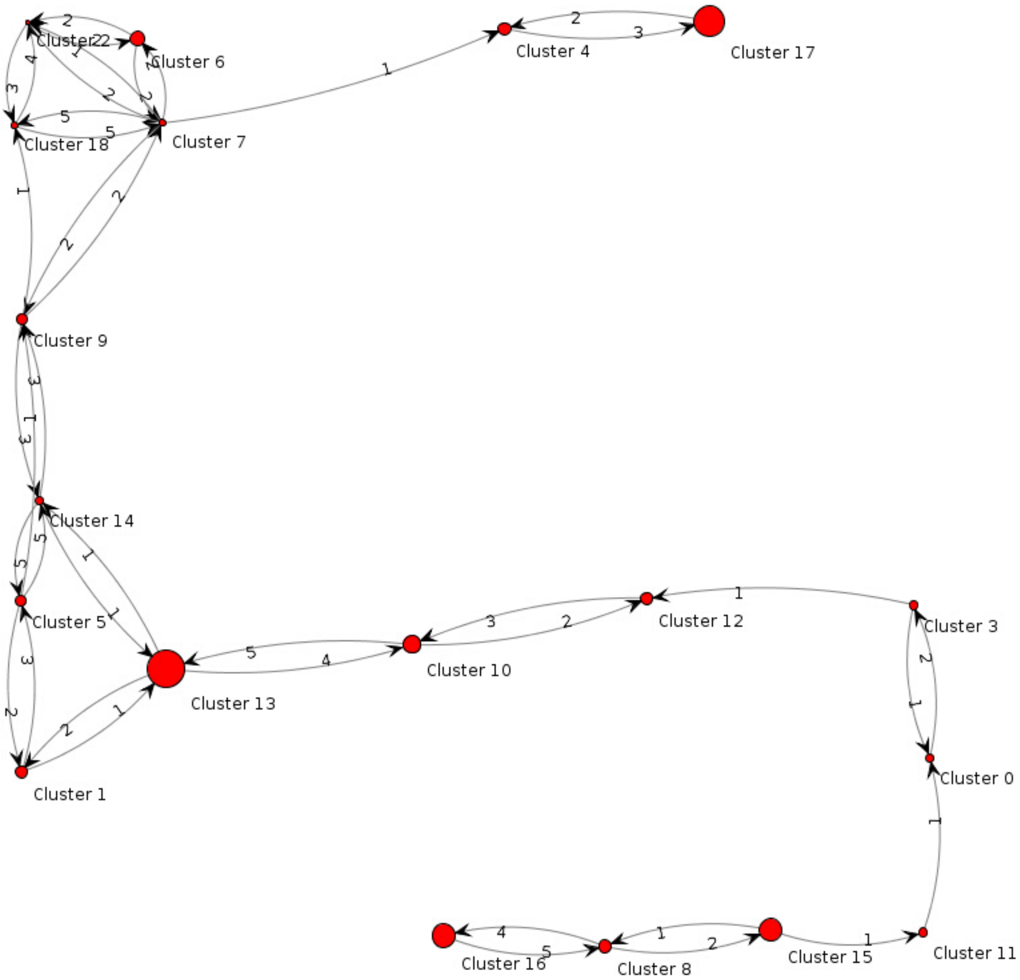


Figure 3: Cluster population resulting from k-medoids clustering on a full-atom MD trajectory of the NCP (1µs, 10.000 frames, by Shaytan et al [137]), using the BiKi Life Sciences software.

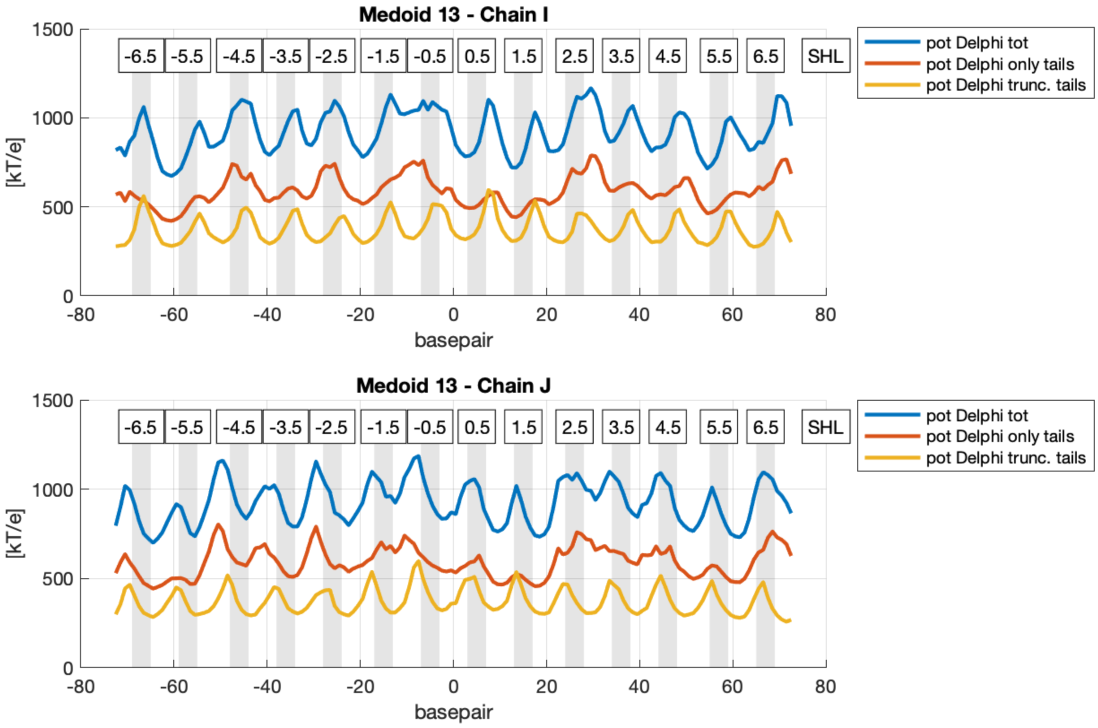


Figure 4: Electrostatic Potential calculated on phosphates of both DNA chains of medoid13. Blue line indicates potential of full structure, red line indicates potential of "only tails" structure, and orange line indicates potential of "no tails" structure.

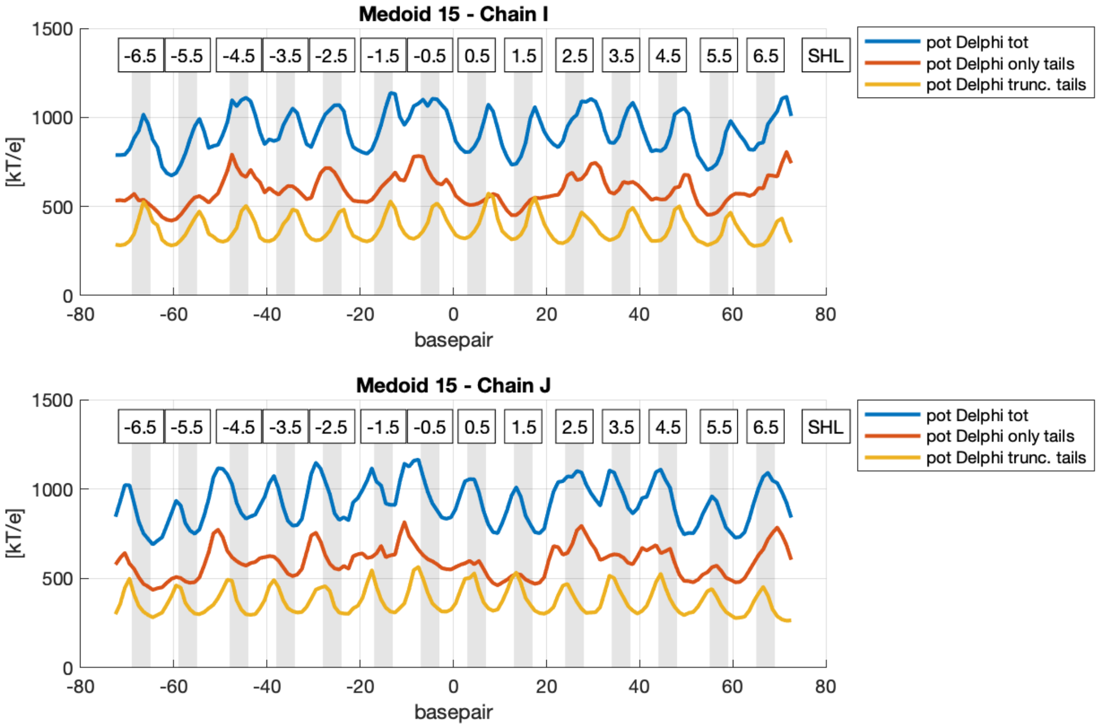


Figure 5: Electrostatic Potential calculated on phosphates of both DNA chains of medoid15. Blue line indicates potential of full structure, red line indicates potential of "only tails" structure, and orange line indicates potential of "no tails" structure.

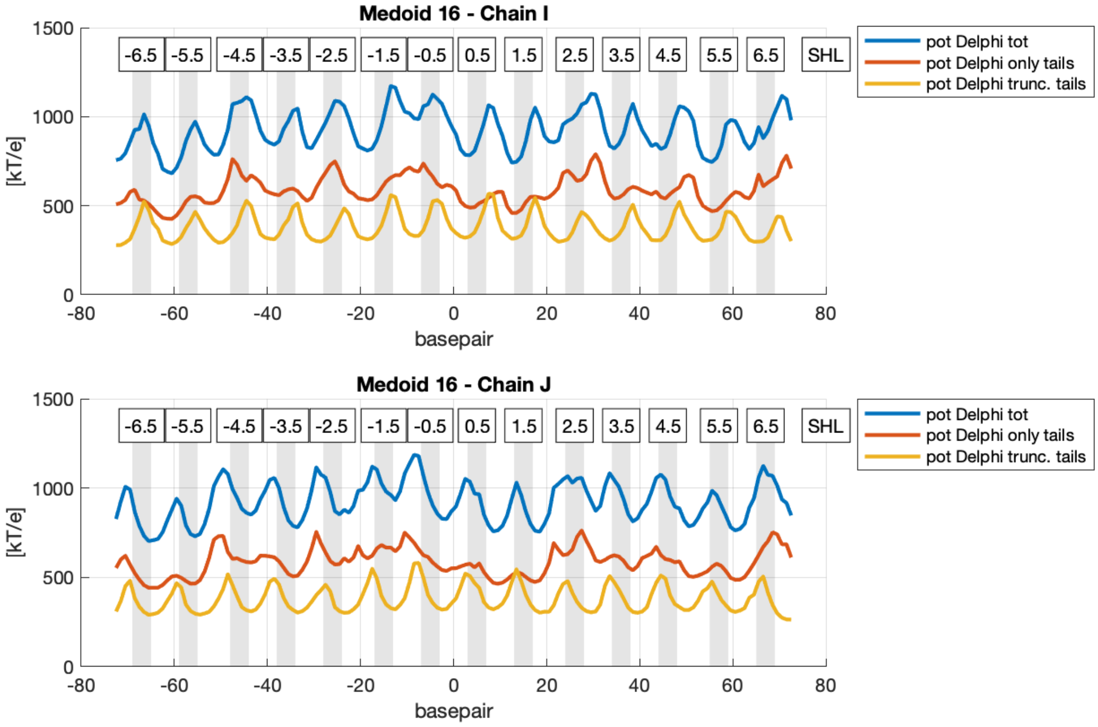


Figure 6: Electrostatic Potential calculated on phosphates of both DNA chains of medoid16. Blue line indicates potential of full structure, red line indicates potential of "only tails" structure, and orange line indicates potential of "no tails" structure.

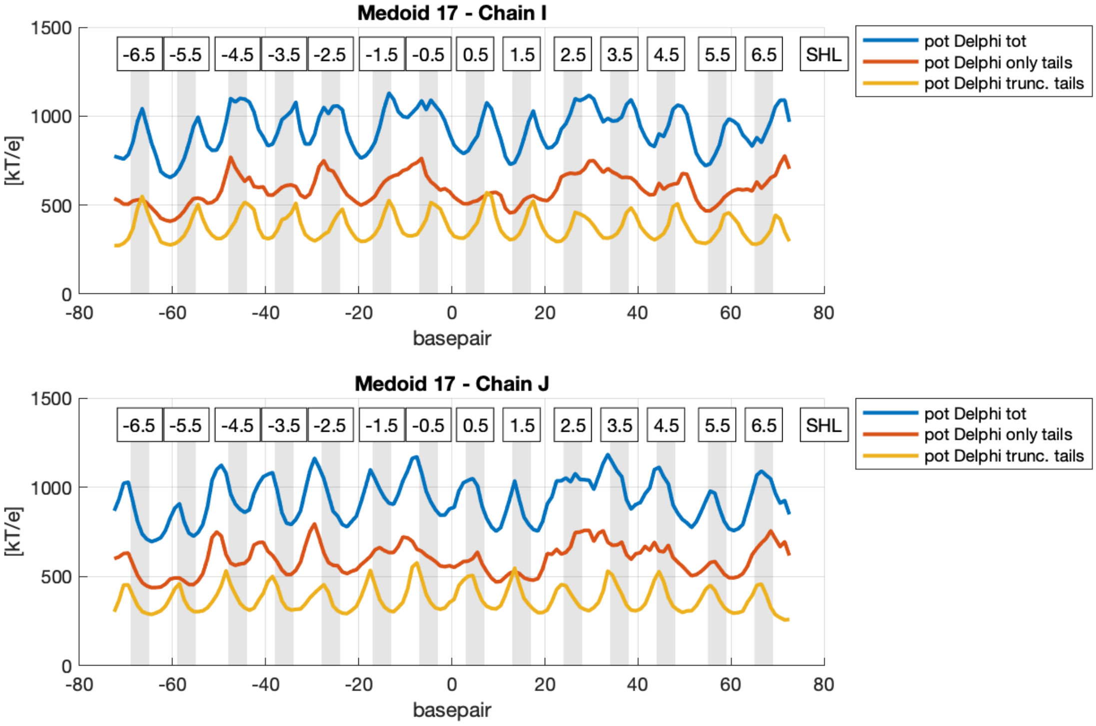


Figure 7: Electrostatic Potential calculated on phosphates of both DNA chains of medoid17. Blue line indicates potential of full structure, red line indicates potential of "only tails" structure, and orange line indicates potential of "no tails" structure.

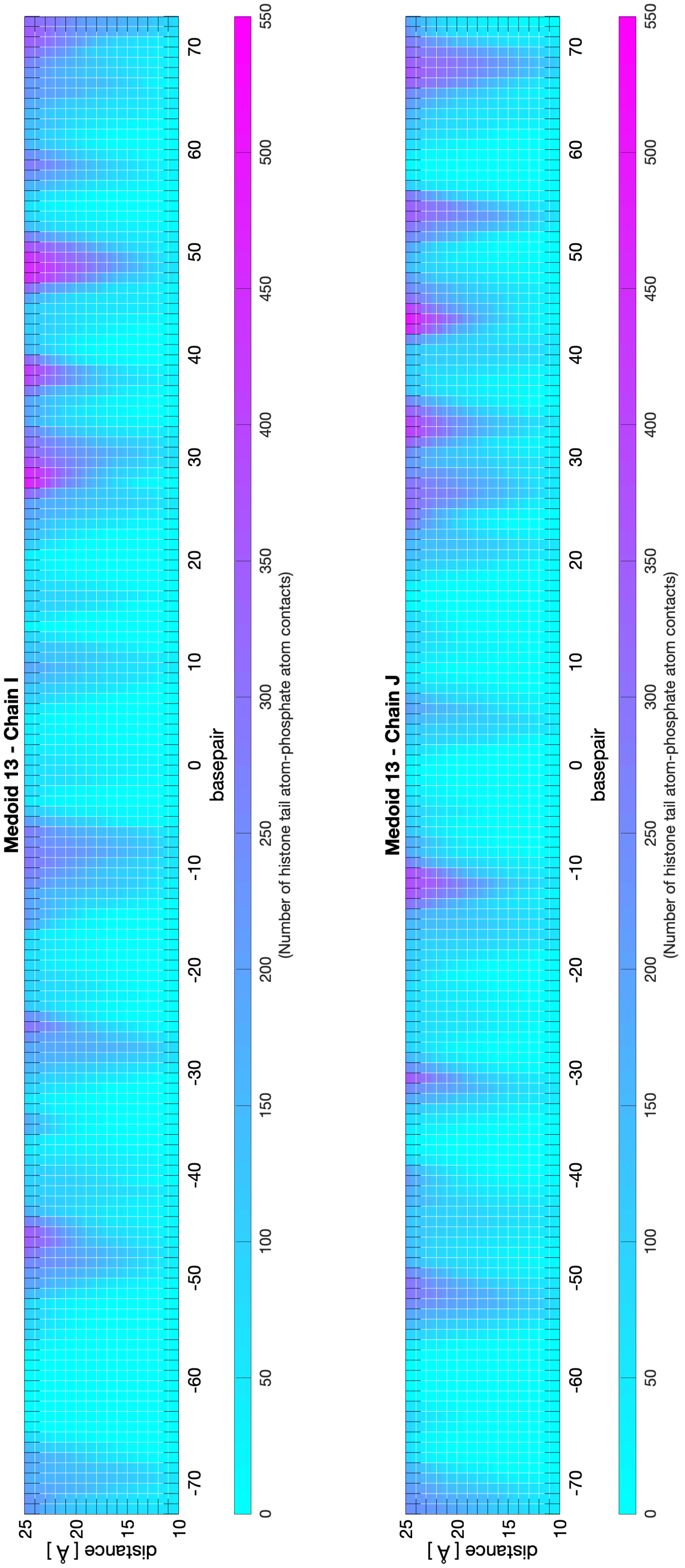


Figure 8: Medoid13: Number of histone tail atom - phosphate atom contacts, within a range of distances from 10 to 25Å

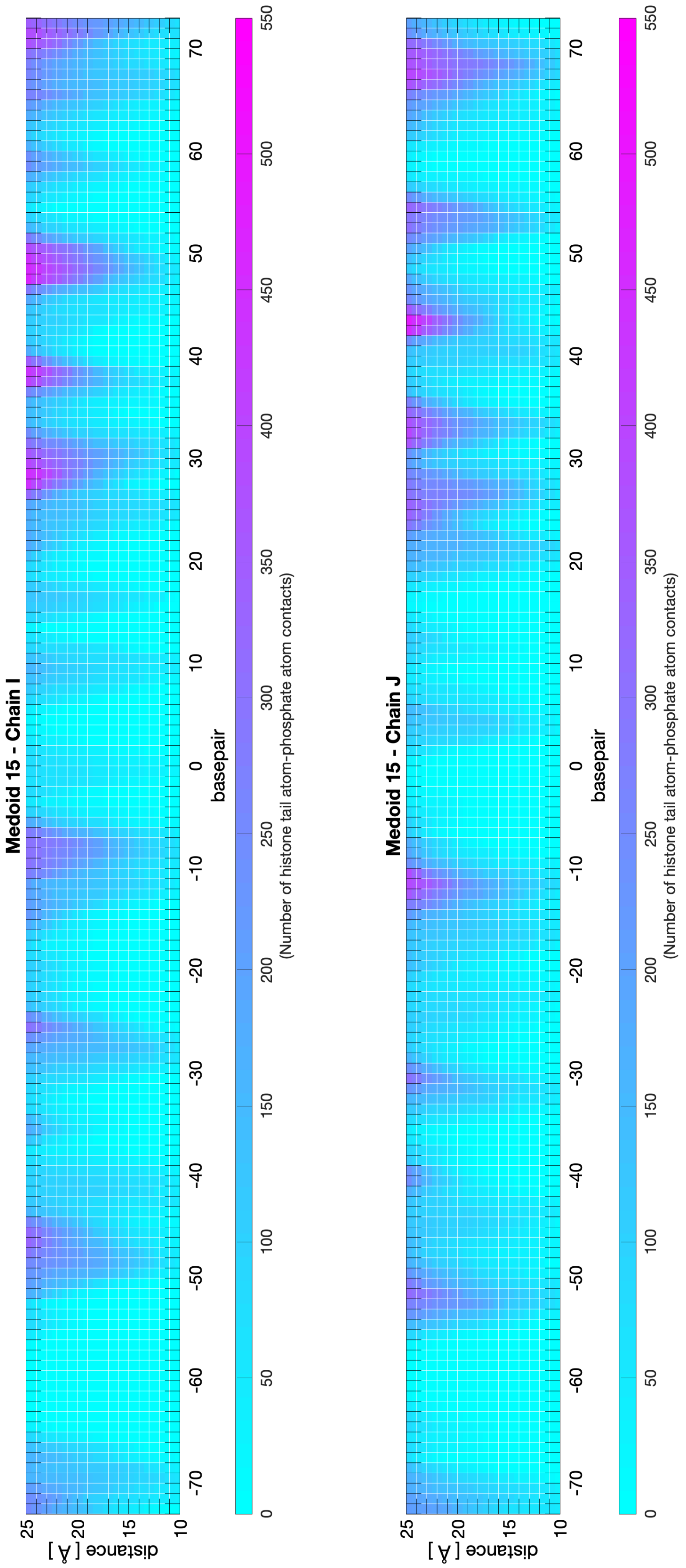


Figure 9: Medoid15: Number of histone tail atom - phosphate atom contacts, within a range of distances from 10 to 25Å

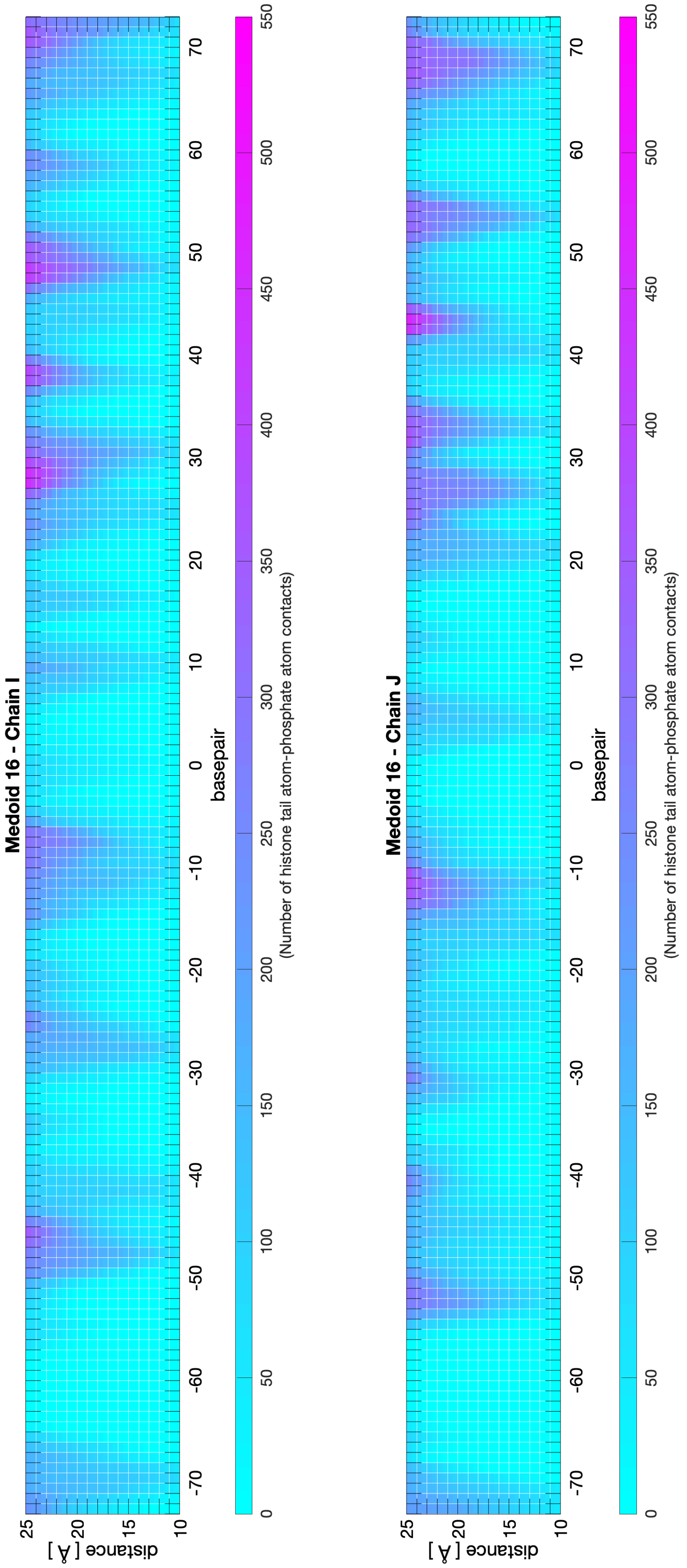


Figure 10: Medoid16: Number of histone tail atom - phosphate atom contacts, within a range of distances from 10 to 25Å

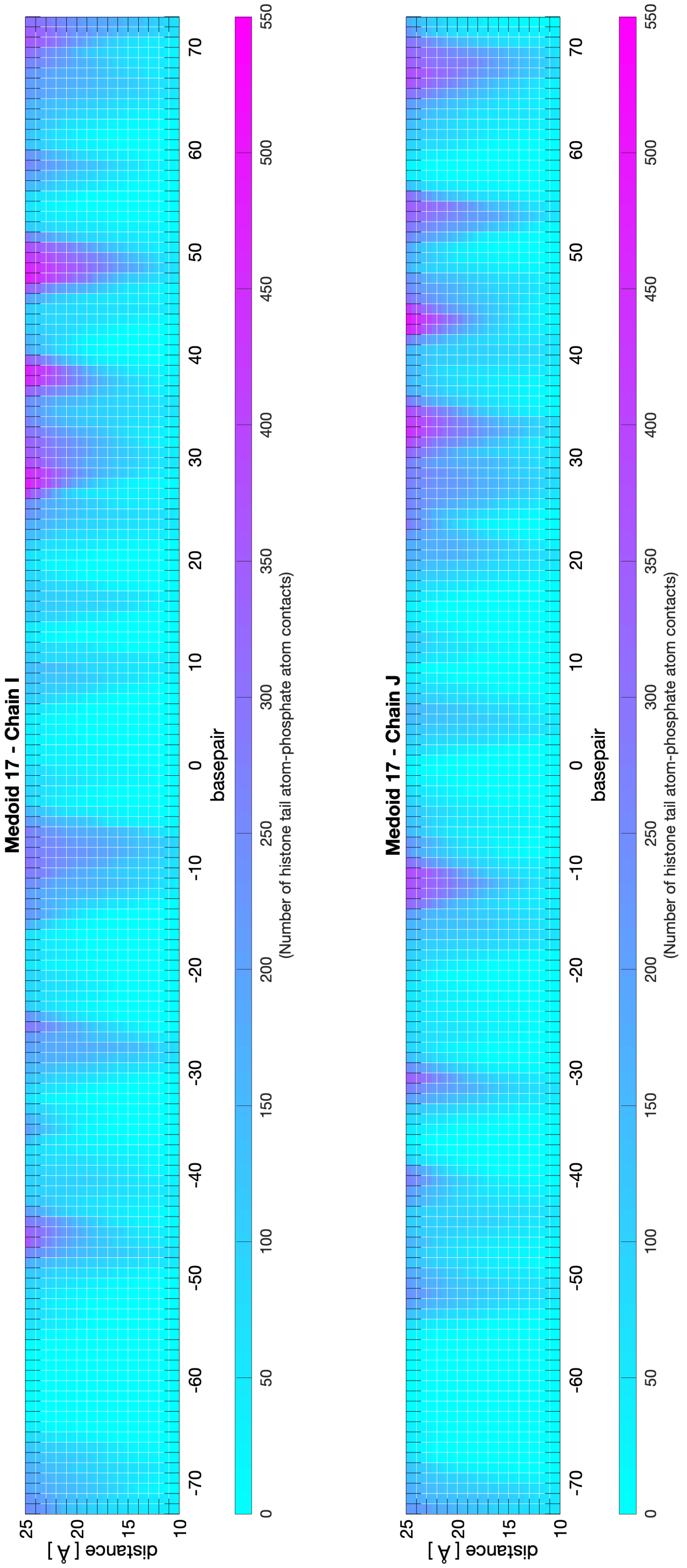


Figure 11: Medoid17: Number of histone tail atom - phosphate atom contacts, within a range of distances from 10 to 25Å

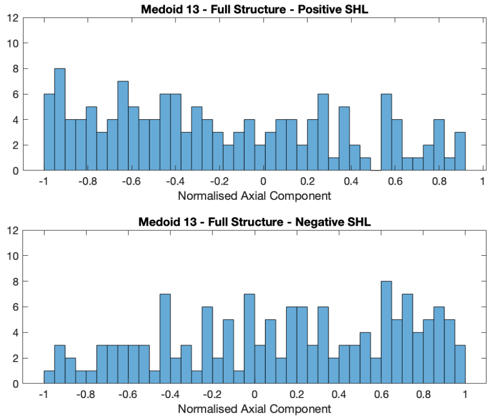


Figure 24: Histogram distribution of axial component of electrostatic force on phosphates of medoid13 (normalised): full structure

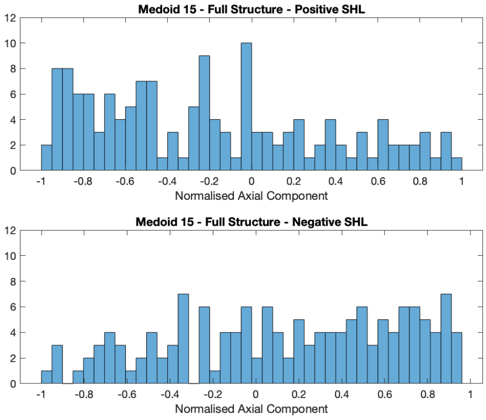


Figure 25: Histogram distribution of axial component of electrostatic force on phosphates of medoid15 (normalised): full structure

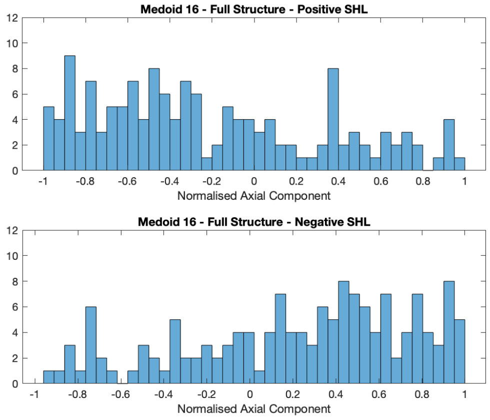


Figure 26: Histogram distribution of axial component of electrostatic force on phosphates of medoid16 (normalised): full structure

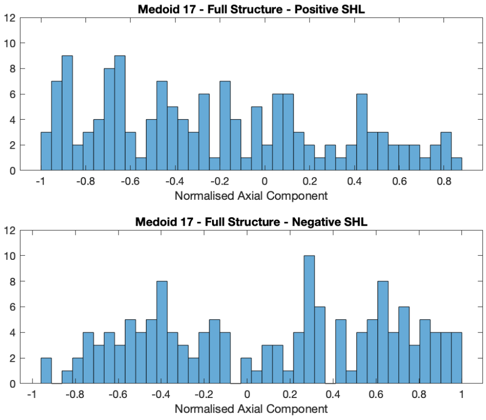


Figure 27: Histogram distribution of axial component of electrostatic force on phosphates of medoid17 (normalised): full structure

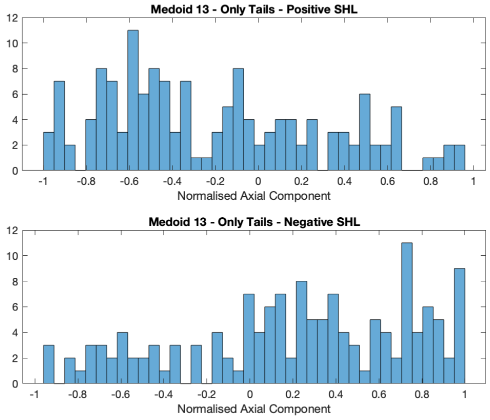


Figure 28: Histogram distribution of axial component of electrostatic force on phosphates of medoid13 (normalised): "only tails" structure

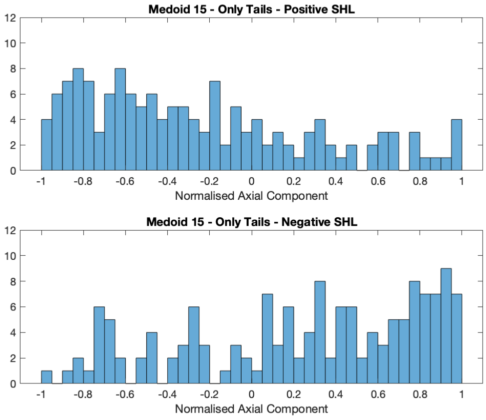


Figure 29: Histogram distribution of axial component of electrostatic force on phosphates of medoid15 (normalised): "only tails" structure

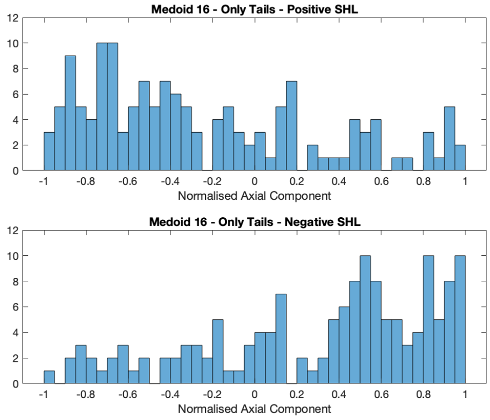


Figure 30: Histogram distribution of axial component of electrostatic force on phosphates of medoid16 (normalised): "only tails" structure

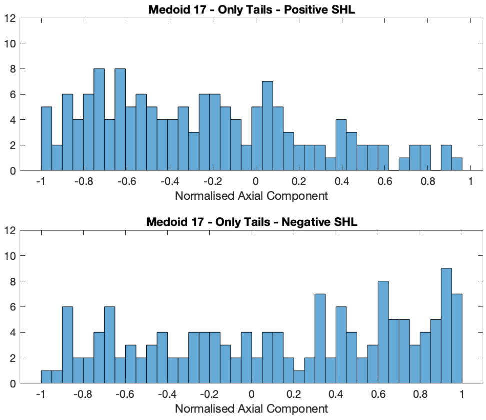


Figure 31: Histogram distribution of axial component of electrostatic force on phosphates of medoid17 (normalised): "only tails" structure

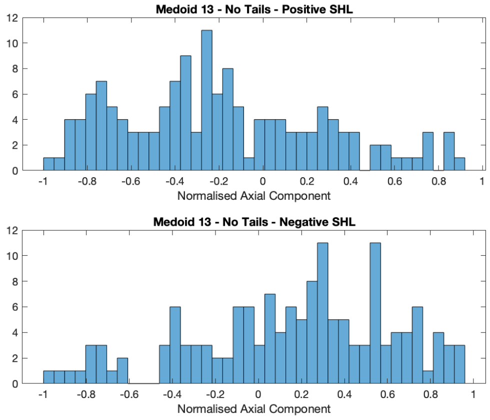


Figure 32: Histogram distribution of axial component of electrostatic force on phosphates of medoid13 (normalised): "no tails" structure

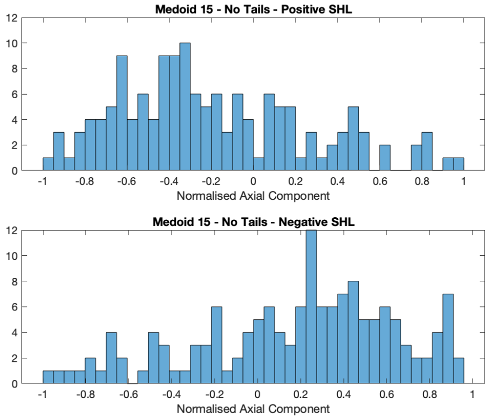


Figure 33: Histogram distribution of axial component of electrostatic force on phosphates of medoid15 (normalised): "no tails" structure

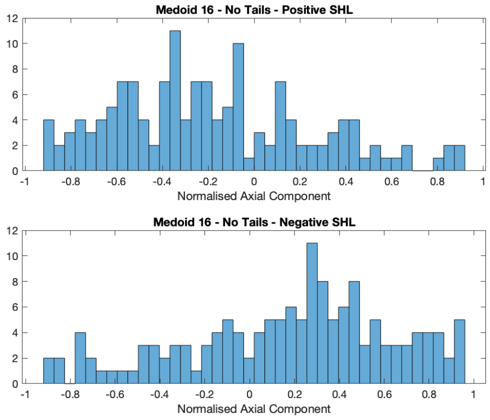


Figure 34: Histogram distribution of axial component of electrostatic force on phosphates of medoid16 (normalised): "no tails" structure

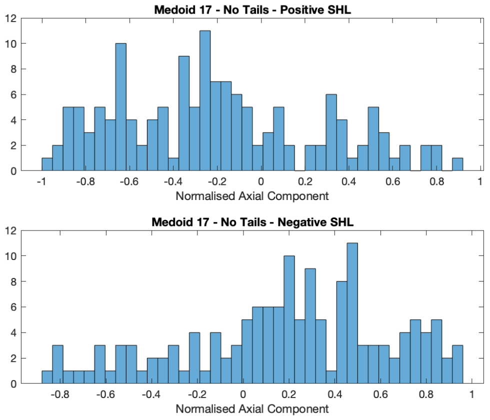
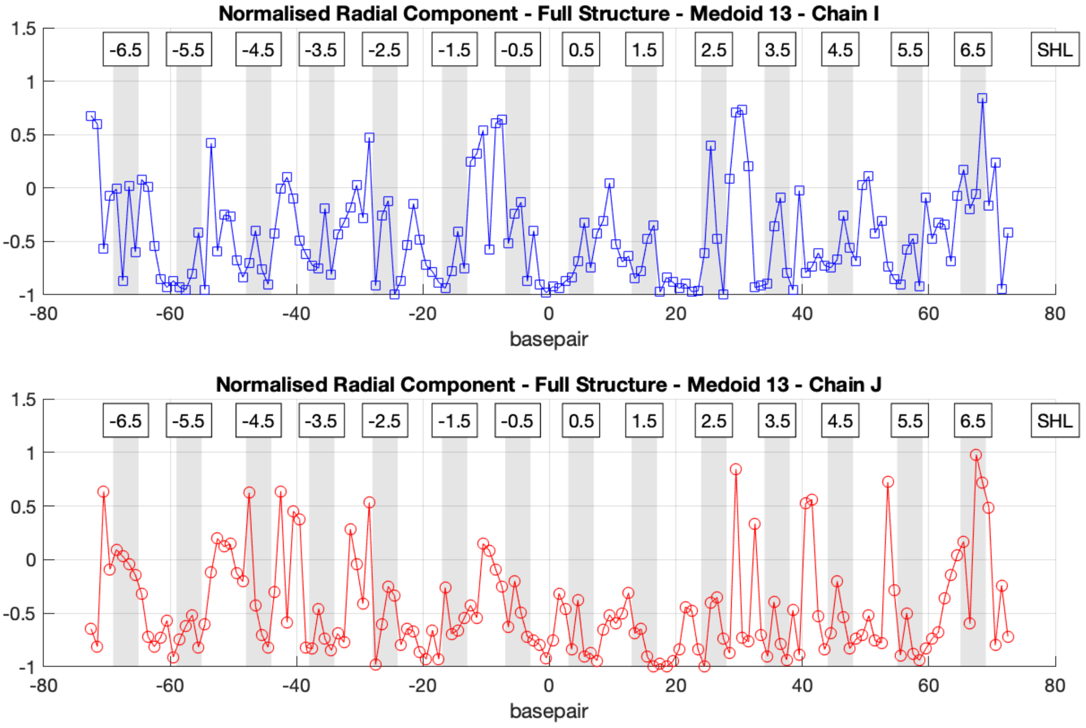


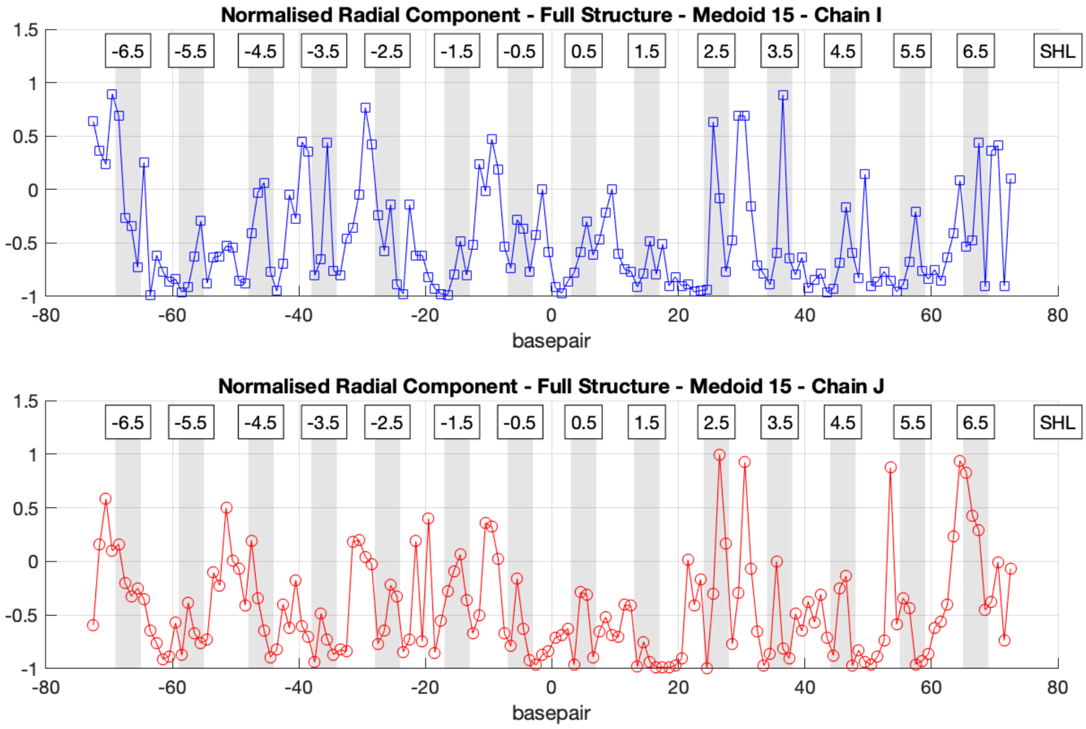
Figure 35: Histogram distribution of axial component of electrostatic force on phosphates of medoid17 (normalised): "no tails" structure



Normalised Radial Component - Full Structure - Medoid 13 - Chain I

Normalised Radial Component - Full Structure - Medoid 13 - Chain J

Figure 12: Radial component of electrostatic force on phosphates of medoid13 (normalised): full structure



Normalised Radial Component - Full Structure - Medoid 15 - Chain I

Normalised Radial Component - Full Structure - Medoid 15 - Chain J

Figure 13: Radial component of electrostatic force on phosphates of medoid15 (normalised): full structure

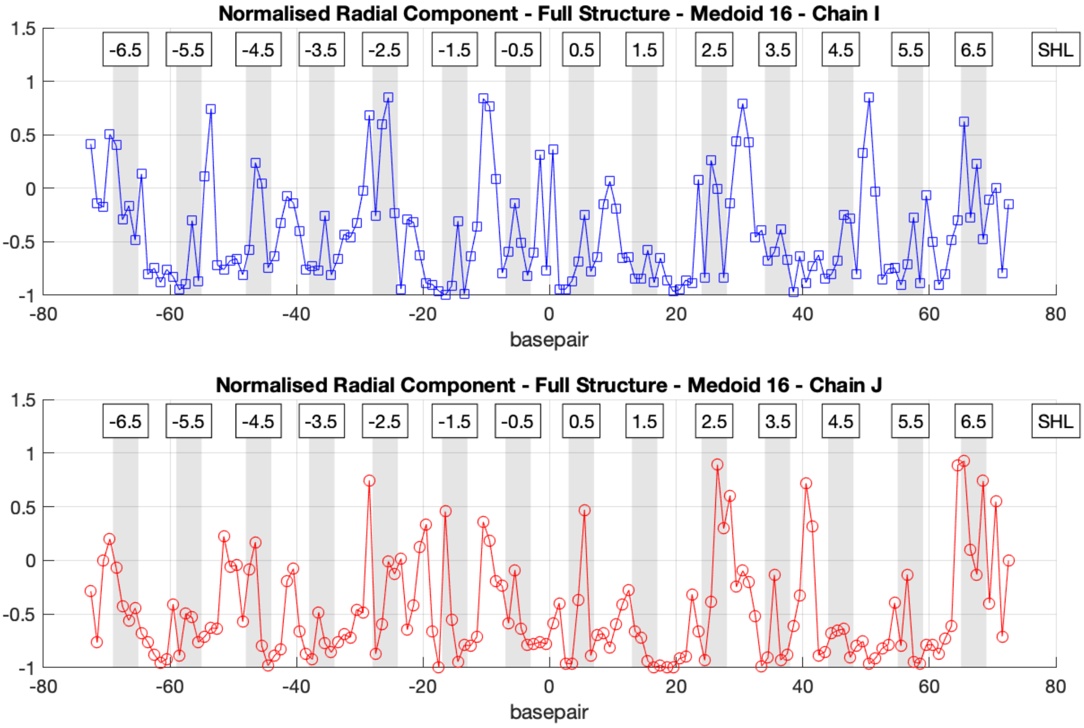


Figure 14: Radial component of electrostatic force on phosphates of medoid16 (normalised): full structure

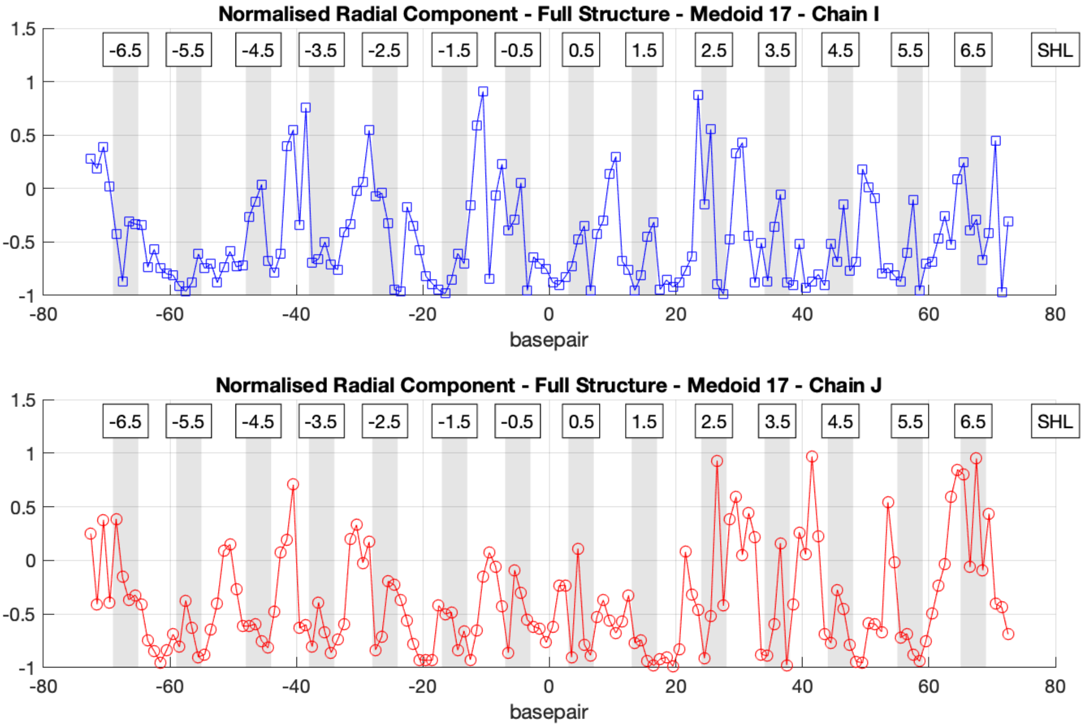
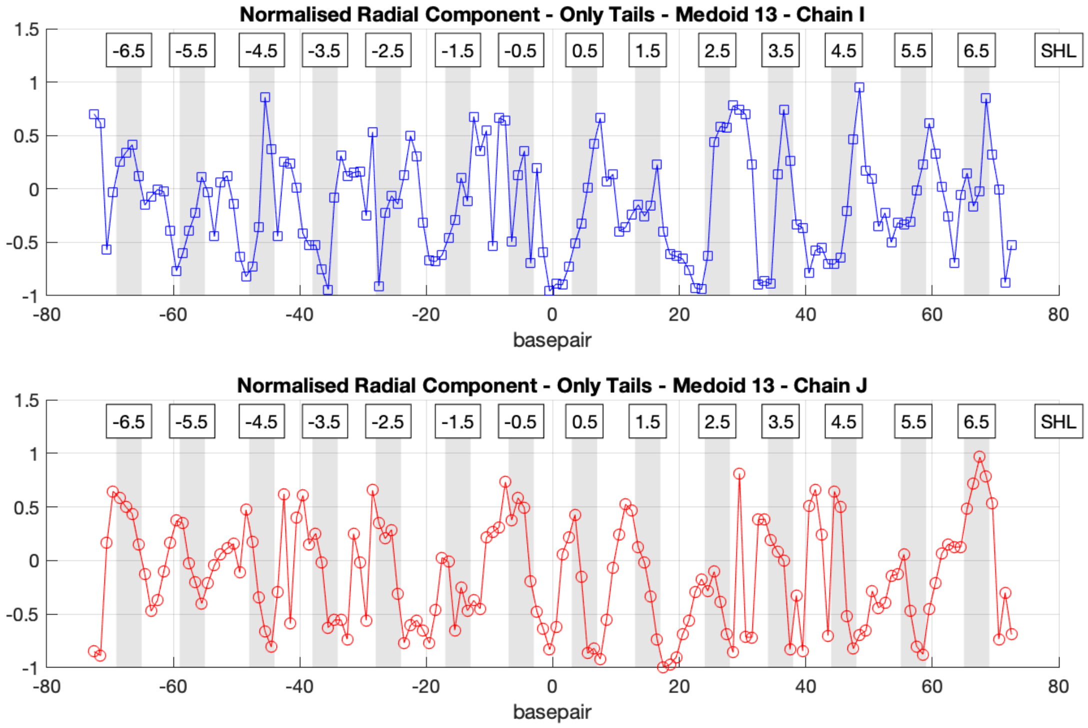


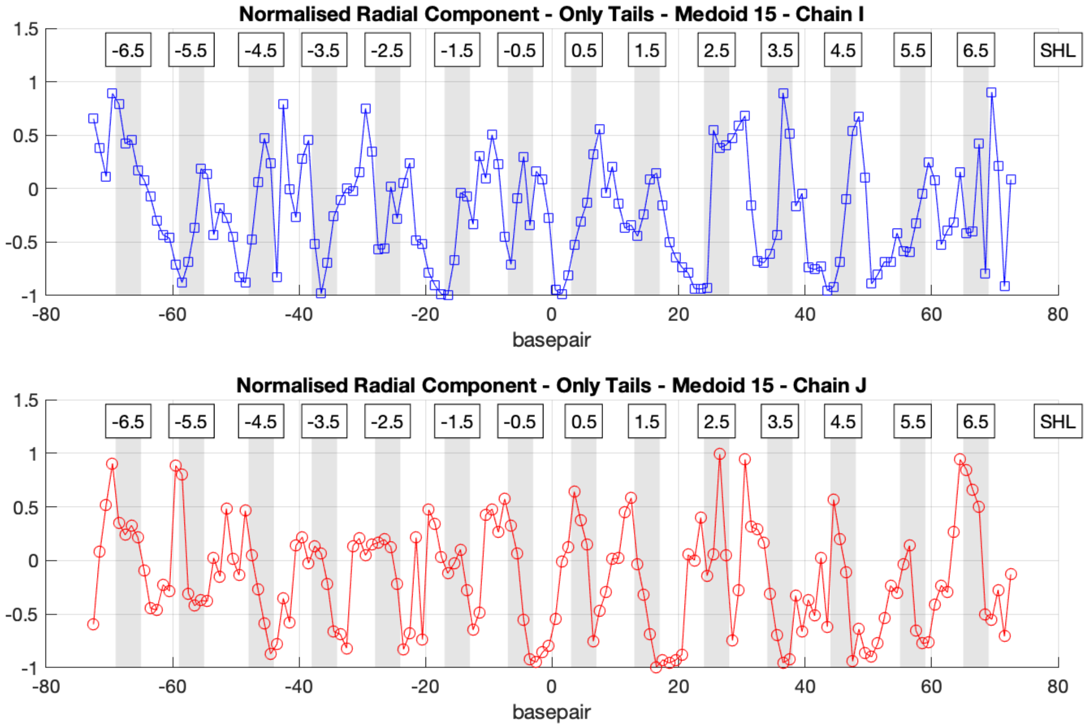
Figure 15: Radial component of electrostatic force on phosphates of medoid17 (normalised): full structure



Normalised Radial Component - Only Tails - Medoid 13 - Chain I

Normalised Radial Component - Only Tails - Medoid 13 - Chain J

Figure 16: Radial component of electrostatic force on phosphates of medoid13 (normalised): "only tails" structure



Normalised Radial Component - Only Tails - Medoid 15 - Chain I

Normalised Radial Component - Only Tails - Medoid 15 - Chain J

Figure 17: Radial component of electrostatic force on phosphates of medoid15 (normalised): "only tails" structure

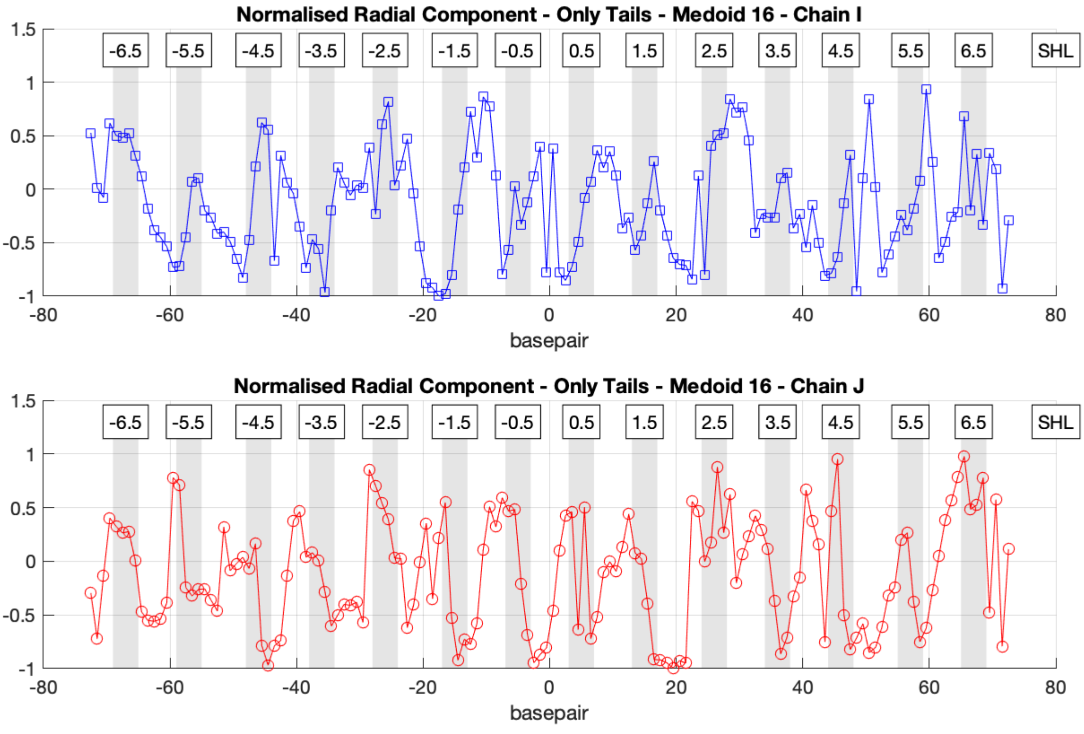


Figure 18: Radial component of electrostatic force on phosphates of medoid16 (normalised): "only tails" structure

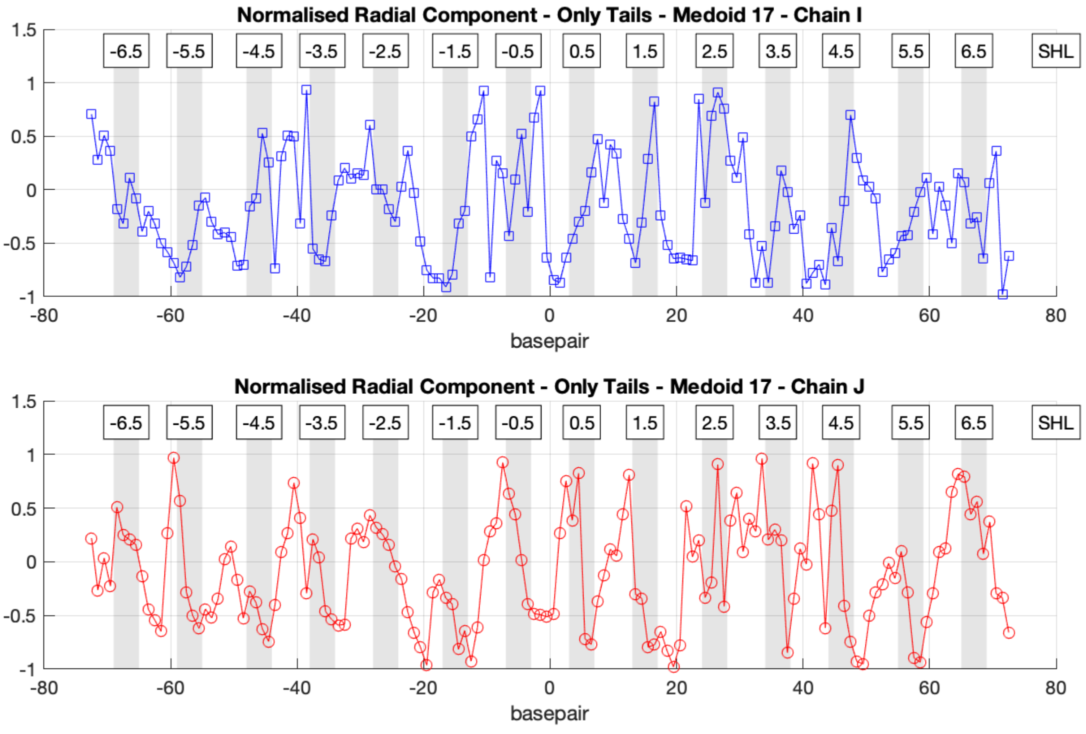


Figure 19: Radial component of electrostatic force on phosphates of medoid17 (normalised): "only tails" structure

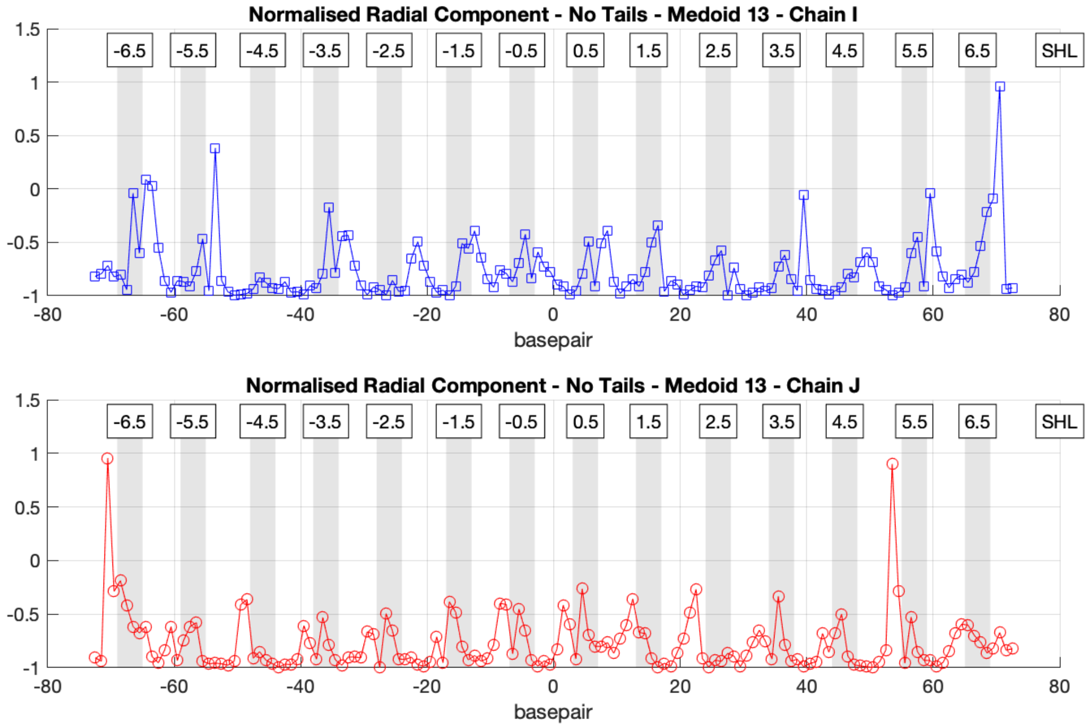


Figure 20: Radial component of electrostatic force on phosphates of medoid13 (normalised): "no tails" structure

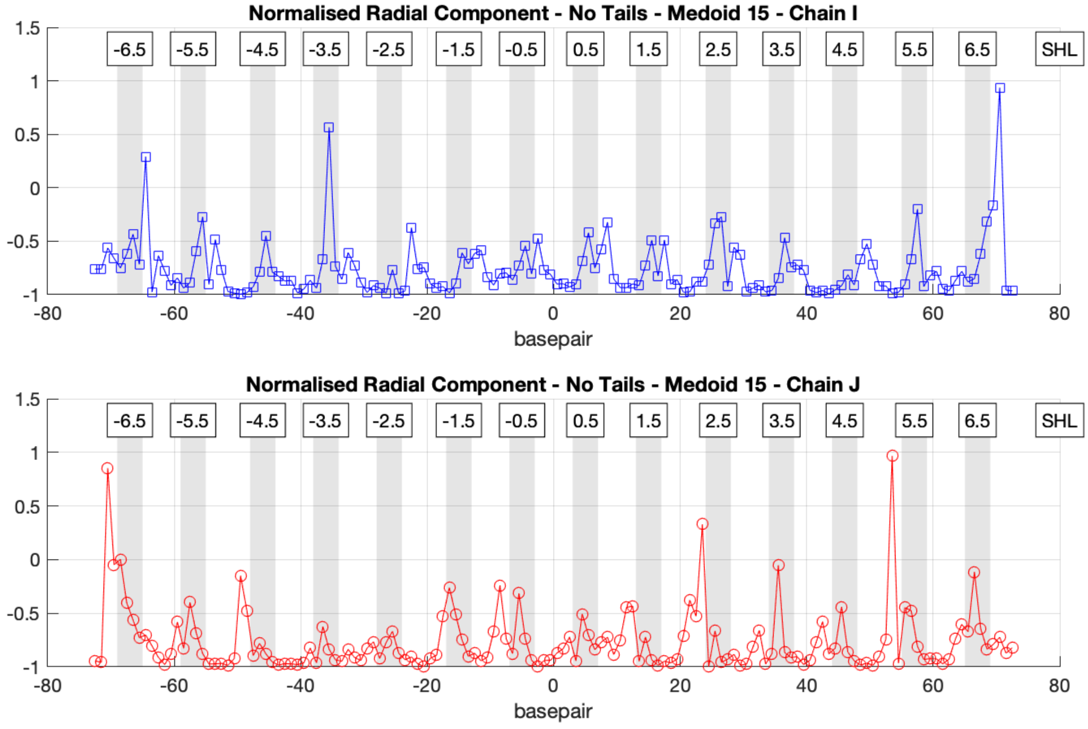


Figure 21: Radial component of electrostatic force on phosphates of medoid15 (normalised): "no tails" structure

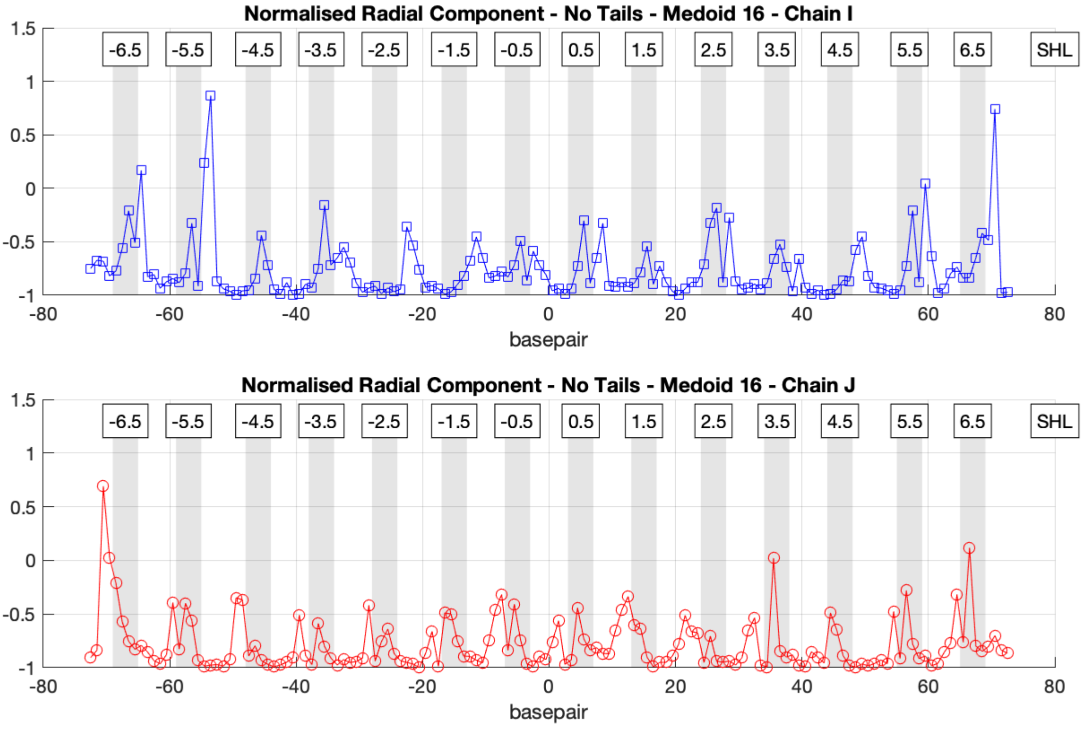


Figure 22: Radial component of electrostatic force on phosphates of medoid16 (normalised): "no tails" structure

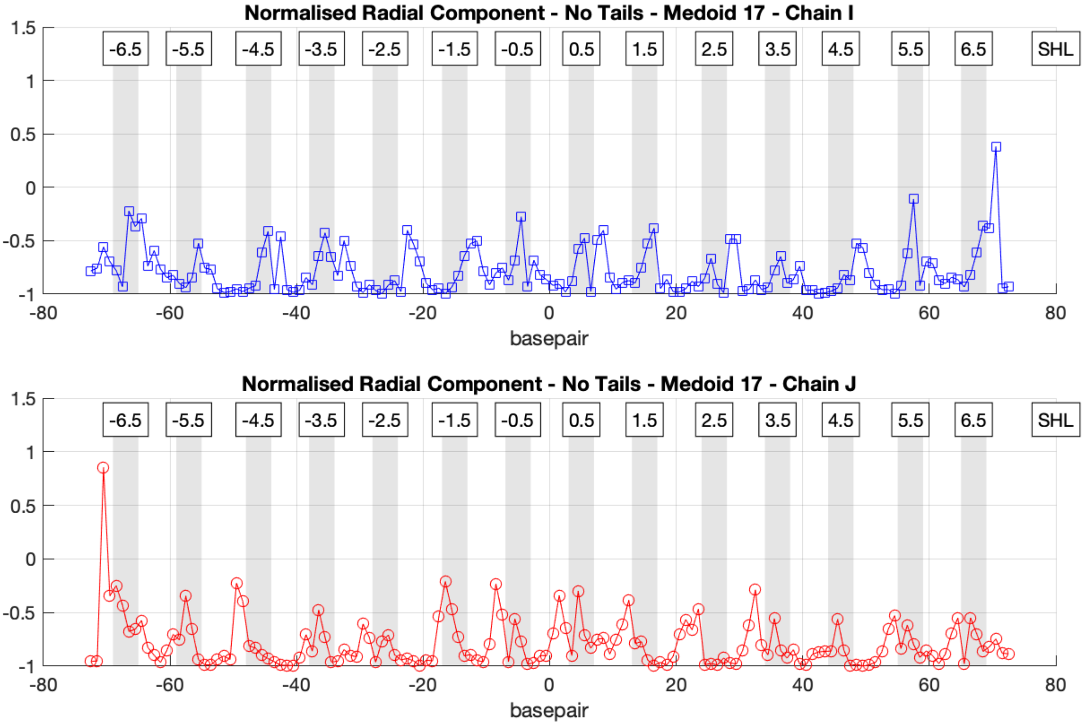


Figure 23: Radial component of electrostatic force on phosphates of medoid17 (normalised): "no tails" structure

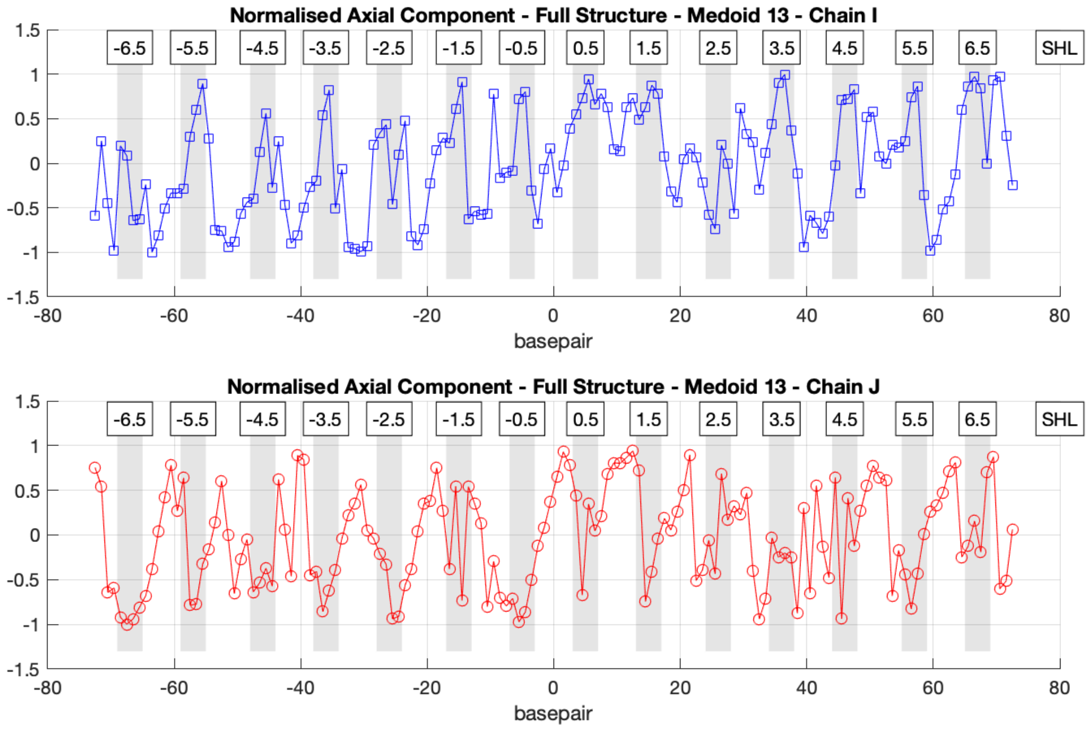


Figure 36: Axial component of electrostatic force on phosphates of medoid13 (normalised): full structure

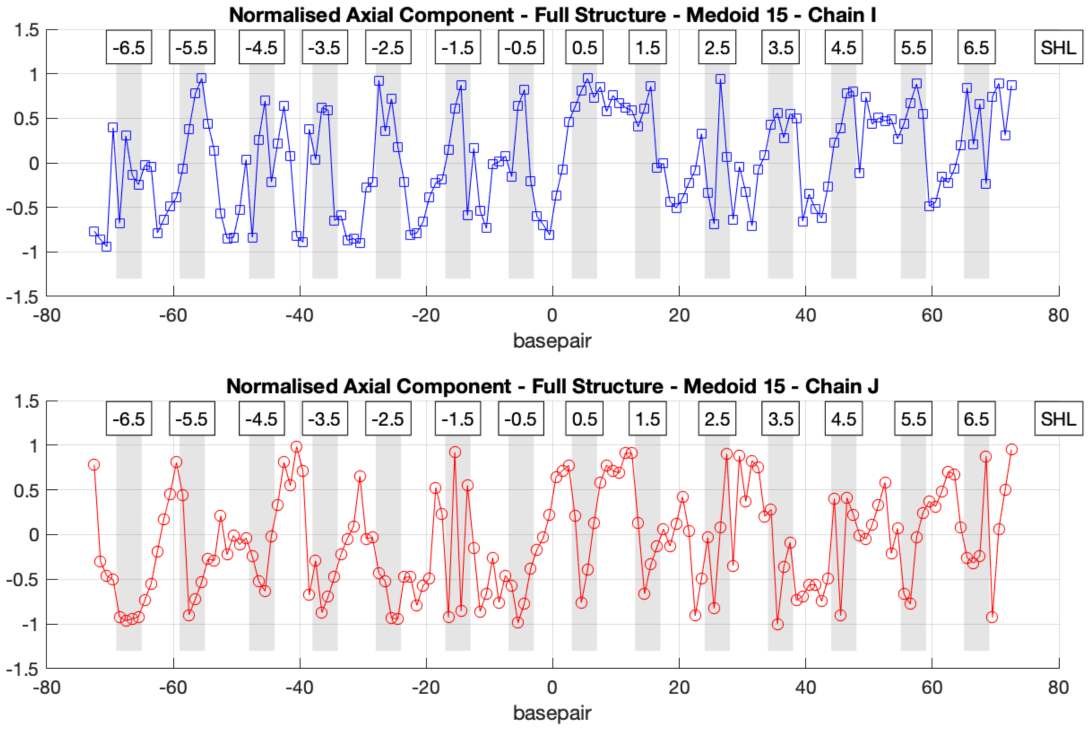


Figure 37: Axial component of electrostatic force on phosphates of medoid15 (normalised): full structure

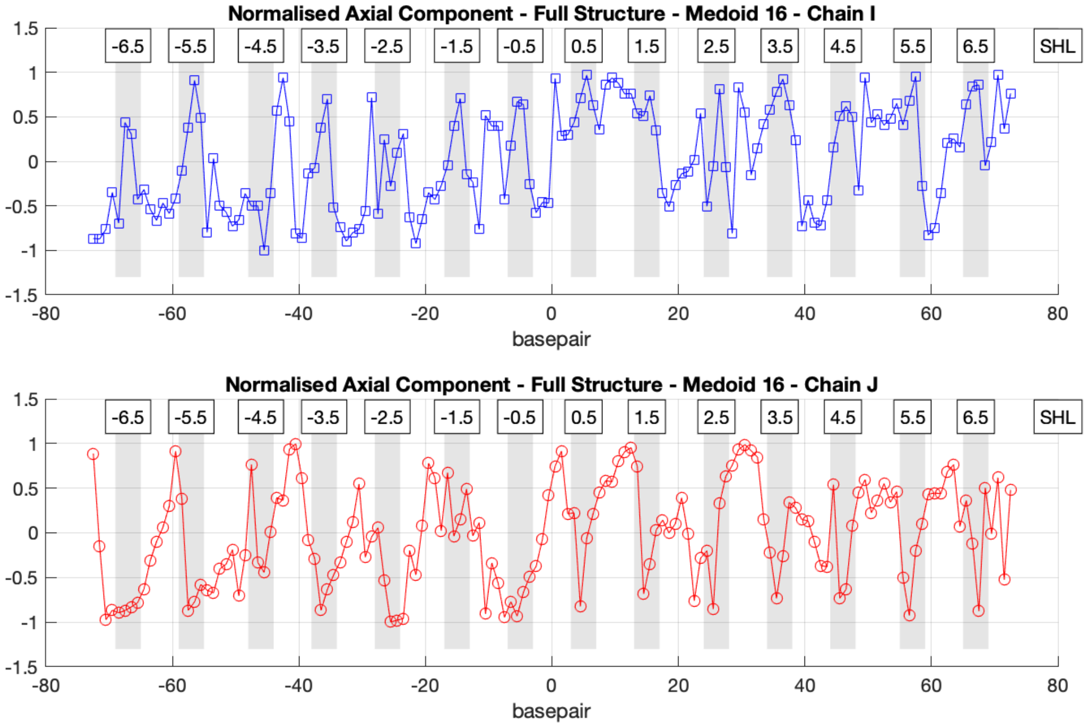


Figure 38: Axial component of electrostatic force on phosphates of medoid16 (normalised): full structure

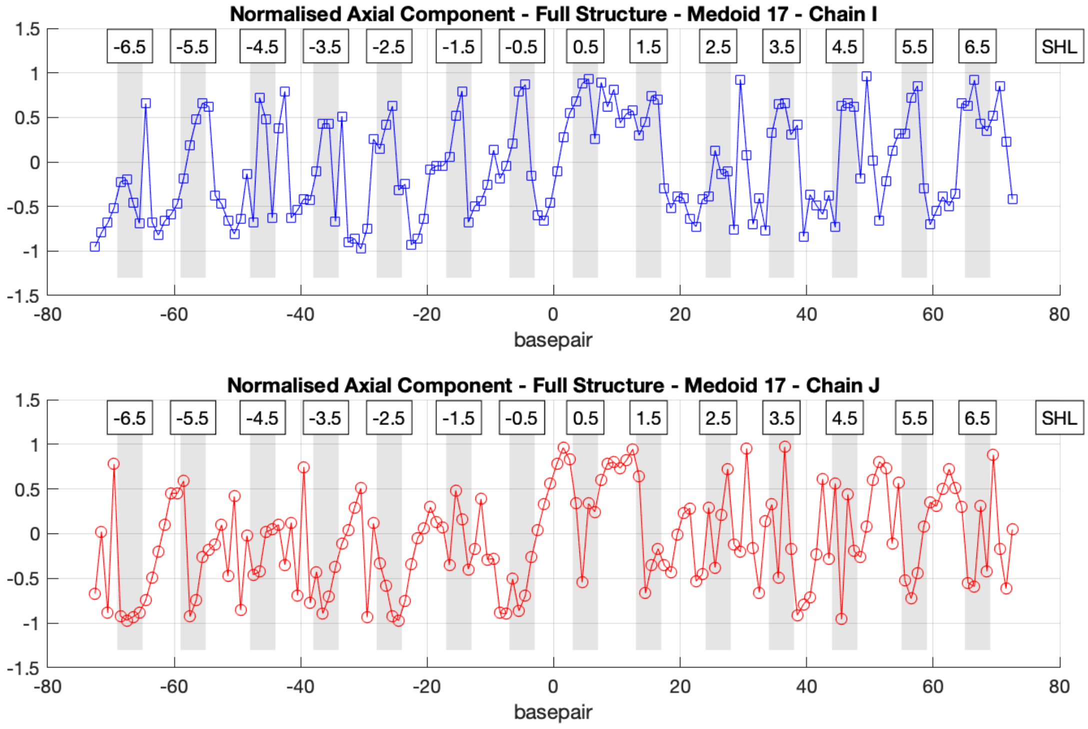
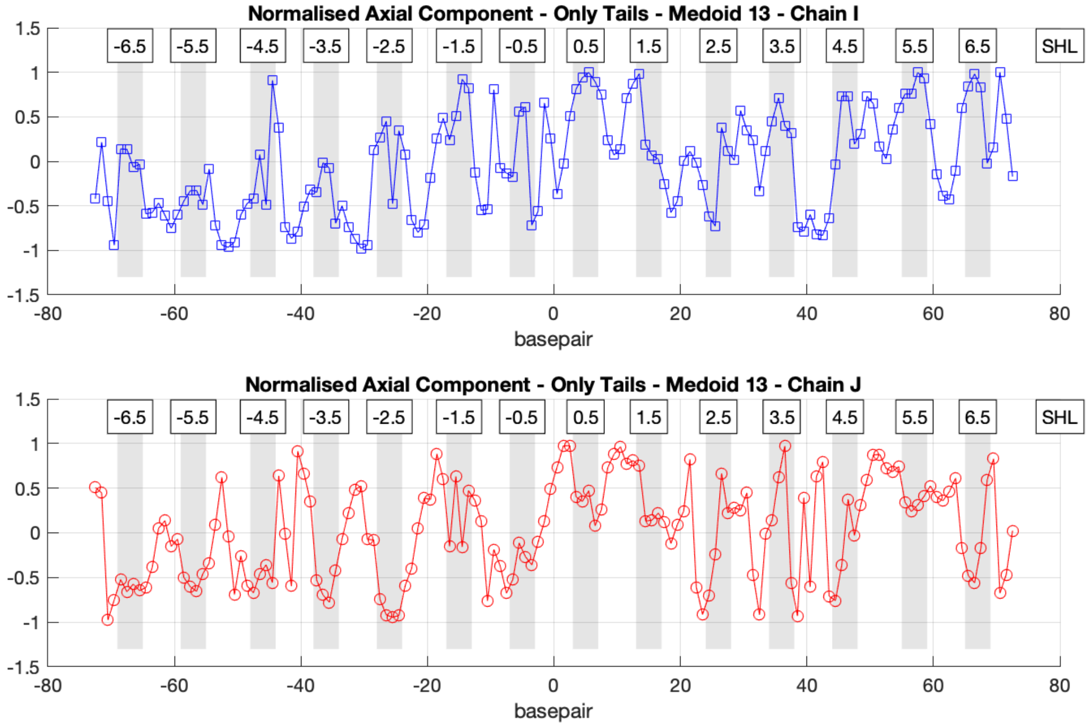
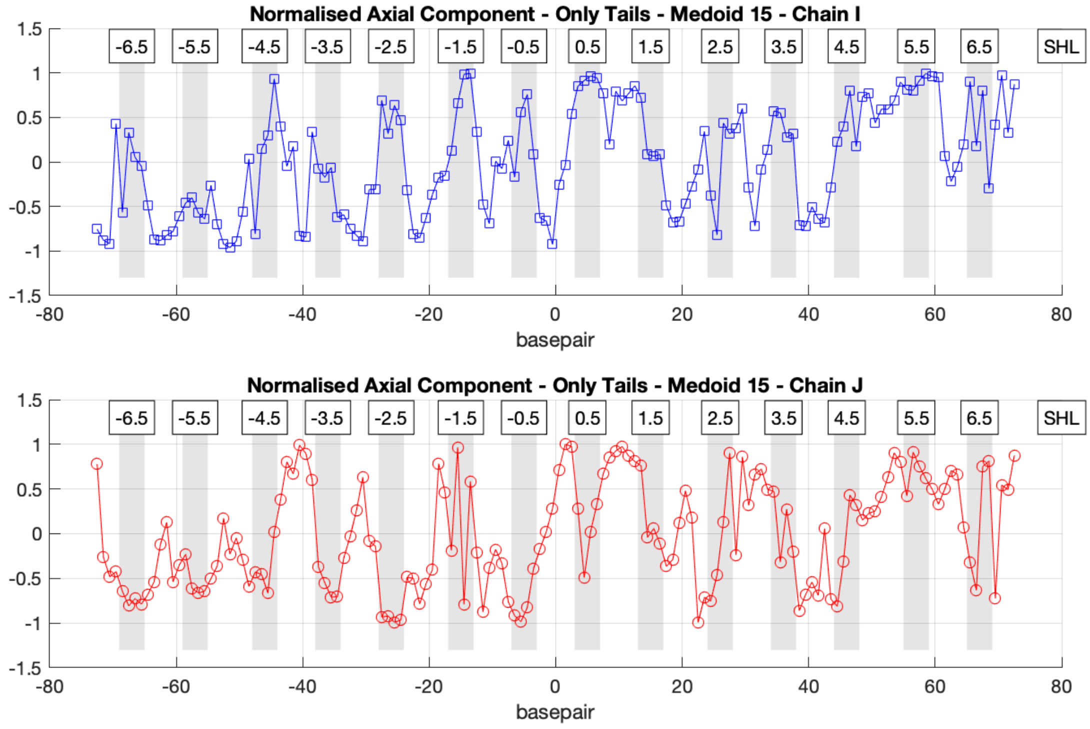


Figure 39: Axial component of electrostatic force on phosphates of medoid17 (normalised): full structure


Normalised Axial Component - Only Tails - Medoid 13 - Chain I

Normalised Axial Component - Only Tails - Medoid 13 - Chain J

Figure 40: Axial component of electrostatic force on phosphates of medoid13 (normalised): "only tails" structure


Normalised Axial Component - Only Tails - Medoid 15 - Chain I

Normalised Axial Component - Only Tails - Medoid 15 - Chain J

Figure 41: Axial component of electrostatic force on phosphates of medoid15 (normalised): "only tails" structure

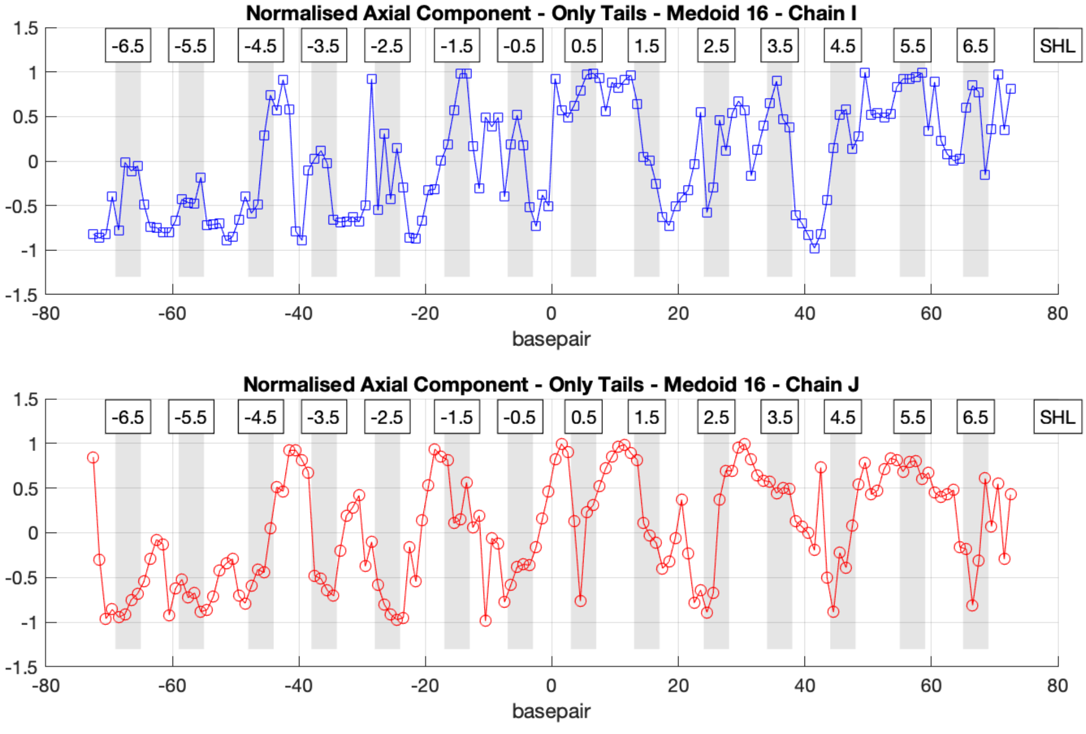


Figure 42: Axial component of electrostatic force on phosphates of medoid16 (normalised): "only tails" structure

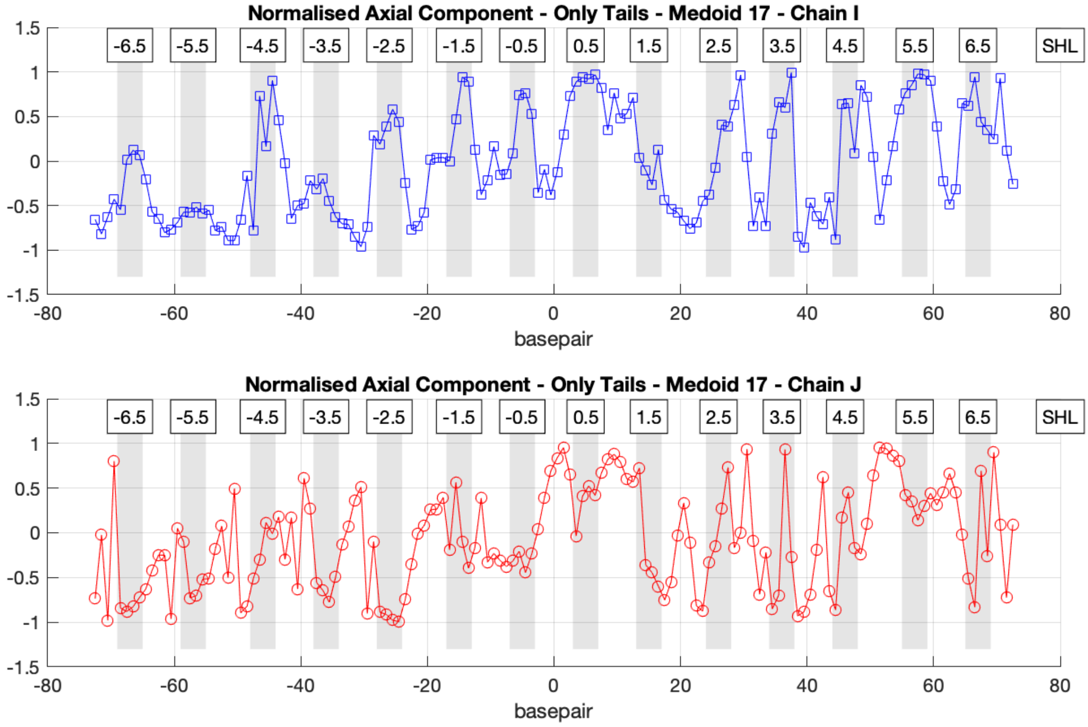


Figure 43: Axial component of electrostatic force on phosphates of medoid17 (normalised): "only tails" structure

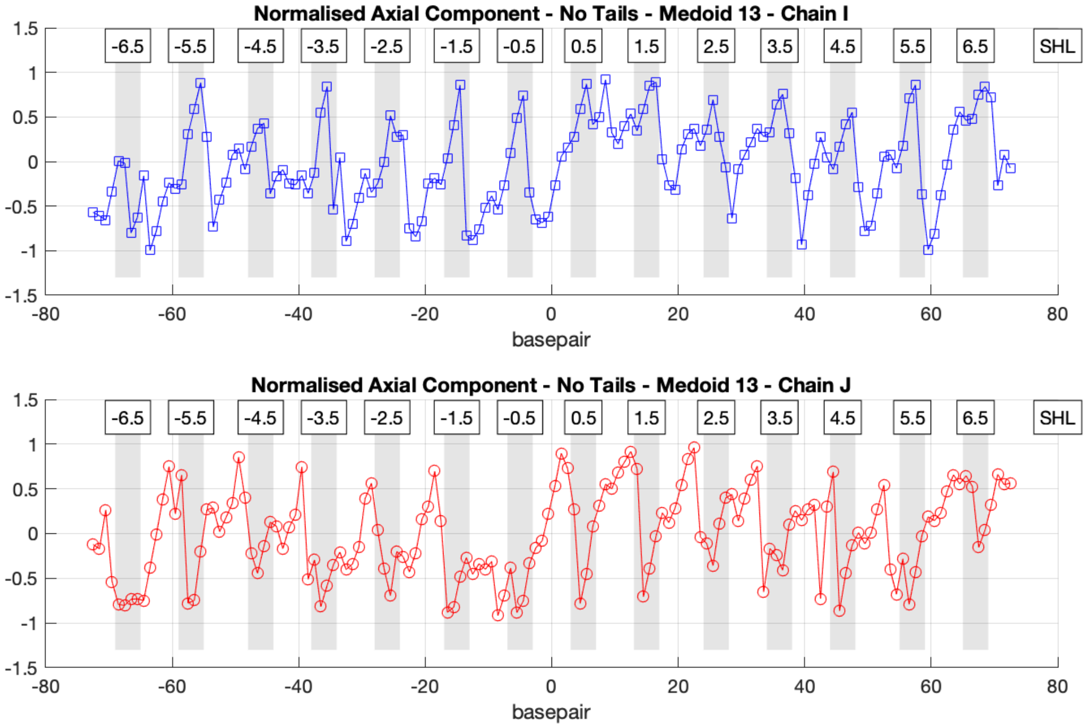


Figure 44: Axial component of electrostatic force on phosphates of medoid13 (normalised): "no tails" structure

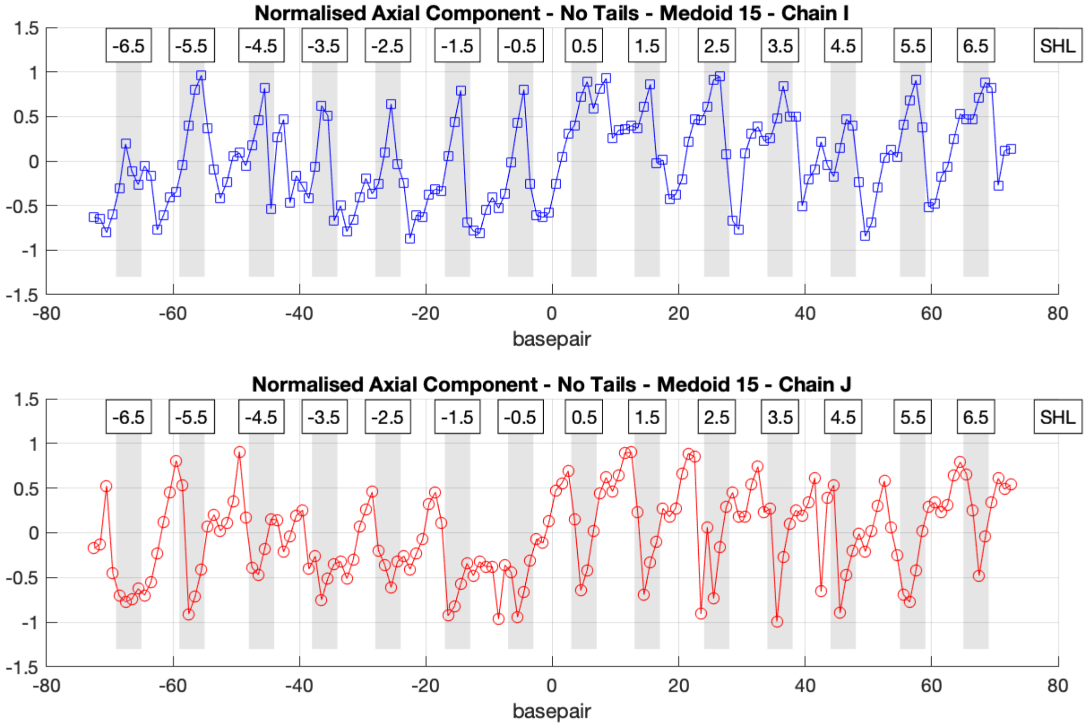


Figure 45: Axial component of electrostatic force on phosphates of medoid15 (normalised): "no tails" structure

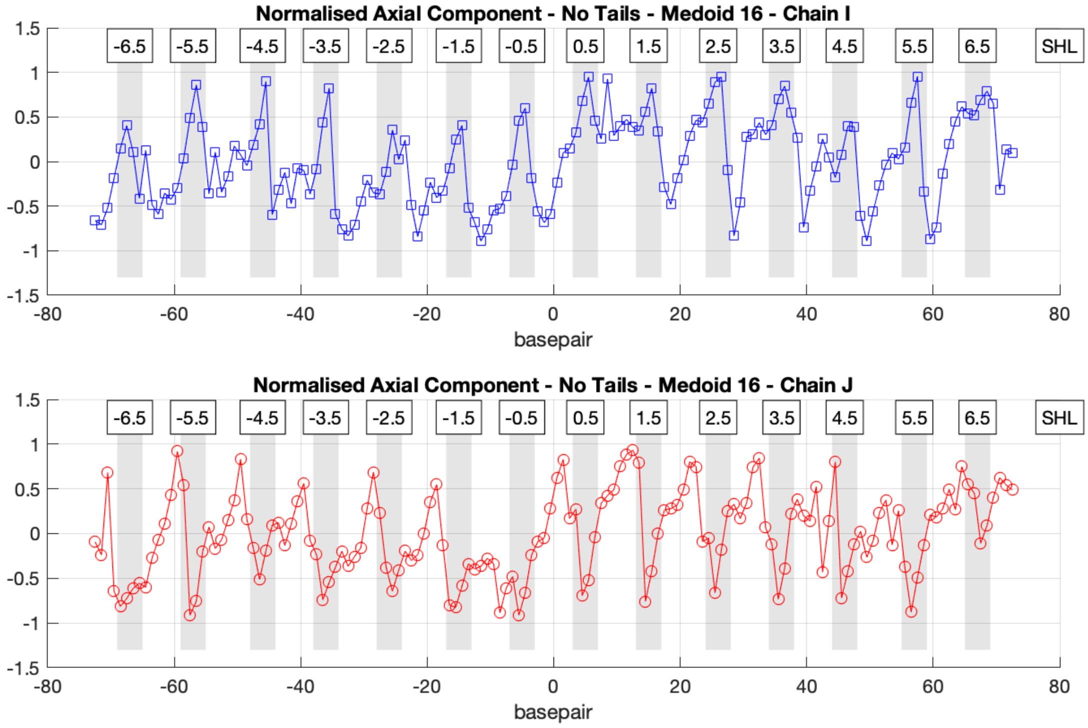


Figure 46: Axial component of electrostatic force on phosphates of medoid16 (normalised): "no tails" structure

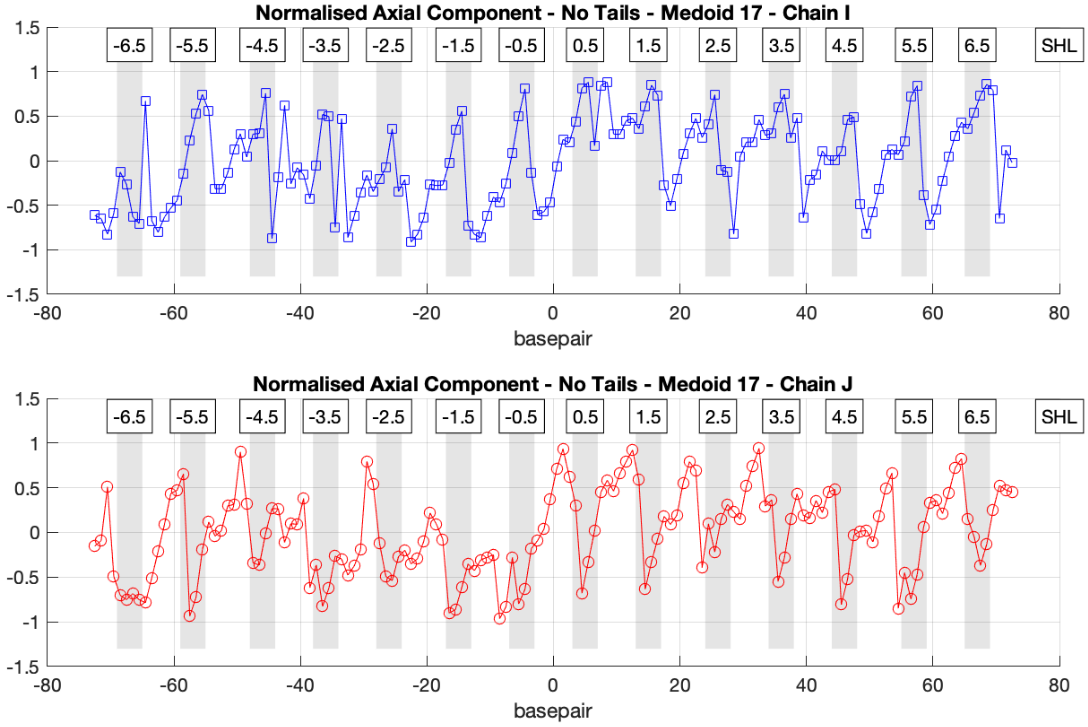


Figure 47: Axial component of electrostatic force on phosphates of medoid17 (normalised): "no tails" structure

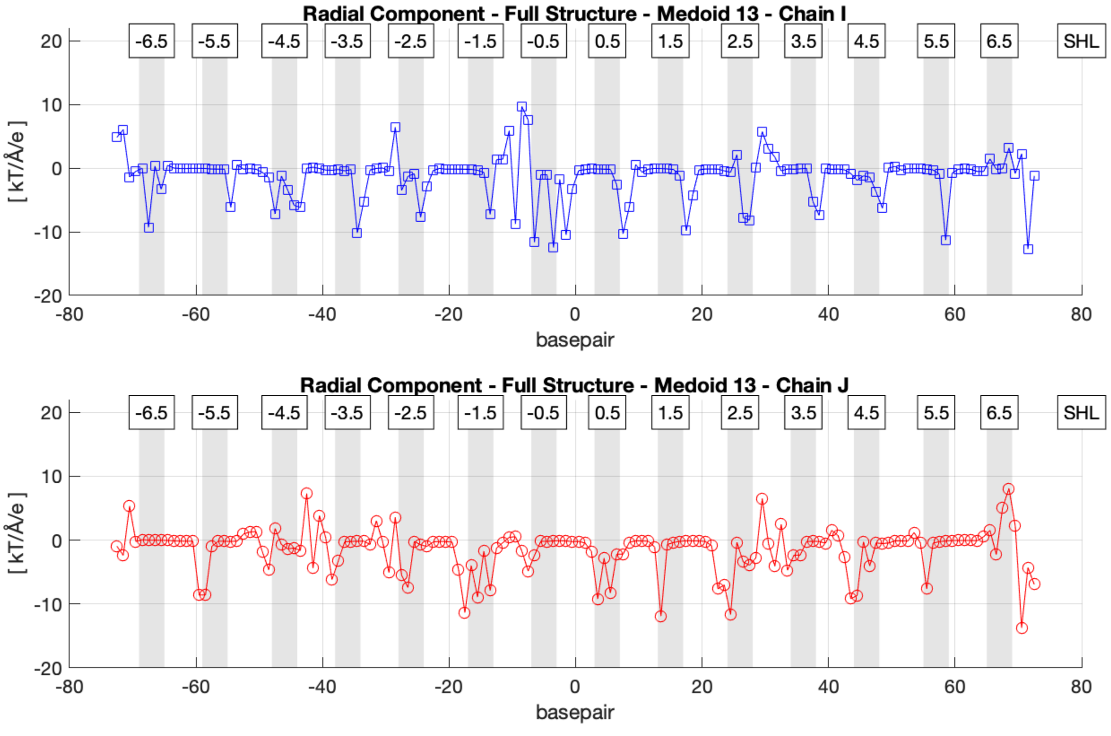


Figure 48: Radial component of electrostatic force on phosphates of medoid13 (non-normalised): full structure

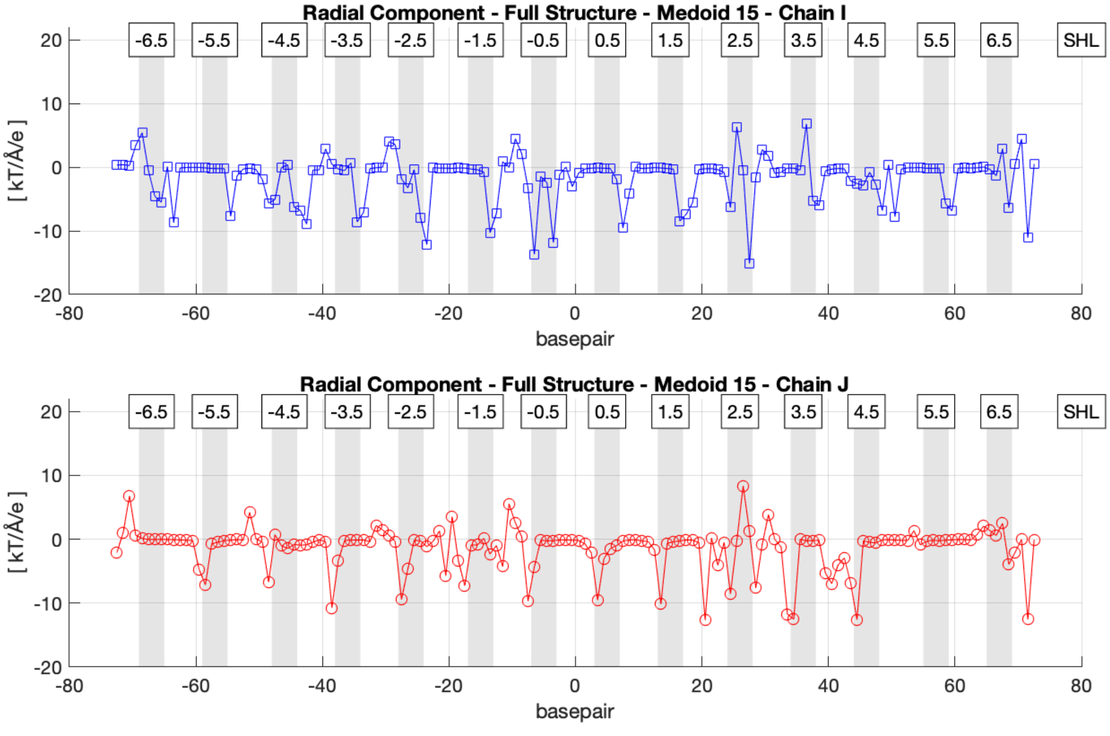
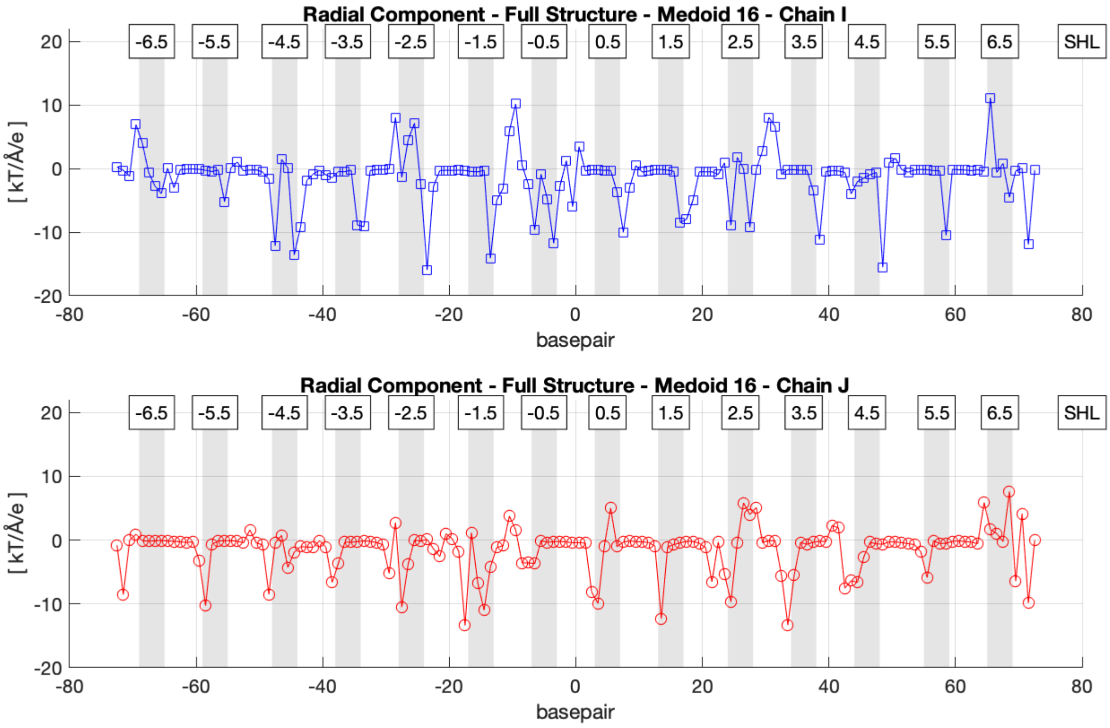


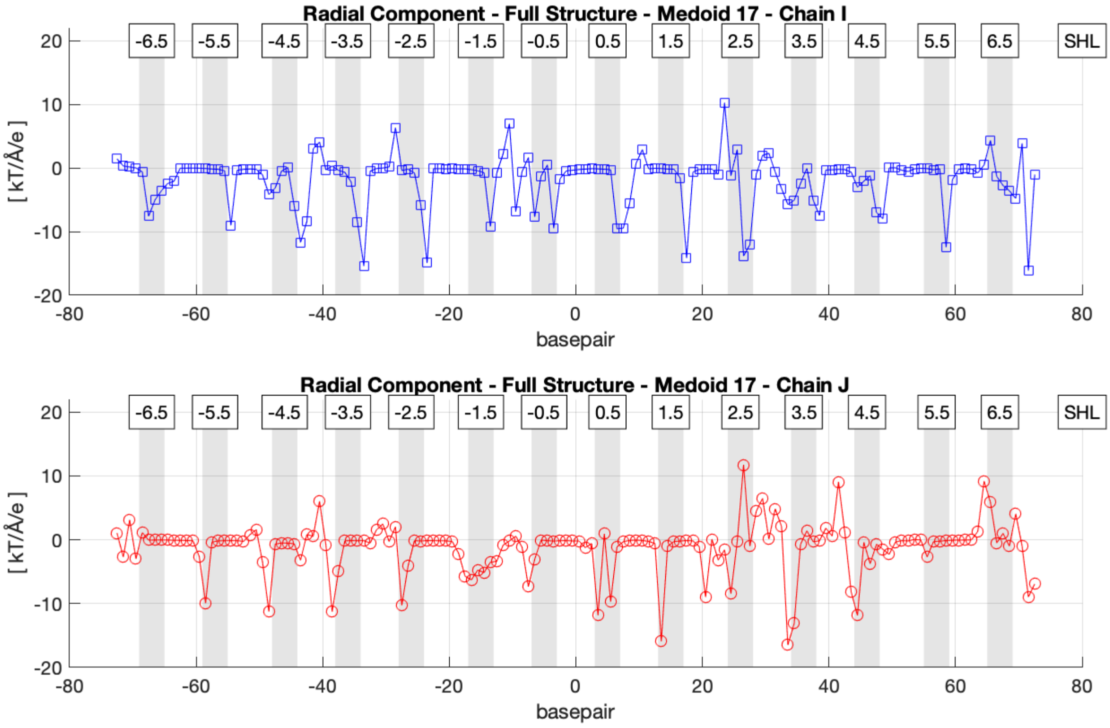
Figure 49: Radial component of electrostatic force on phosphates of medoid15 (non-normalised): full structure



Radial Component - Full Structure - Medoid 16 - Chain I

Radial Component - Full Structure - Medoid 16 - Chain J

Figure 50: Radial component of electrostatic force on phosphates of medoid16 (non-normalised): full structure



Radial Component - Full Structure - Medoid 17 - Chain I

Radial Component - Full Structure - Medoid 17 - Chain J

Figure 51: Radial component of electrostatic force on phosphates of medoid17 (non-normalised): full structure

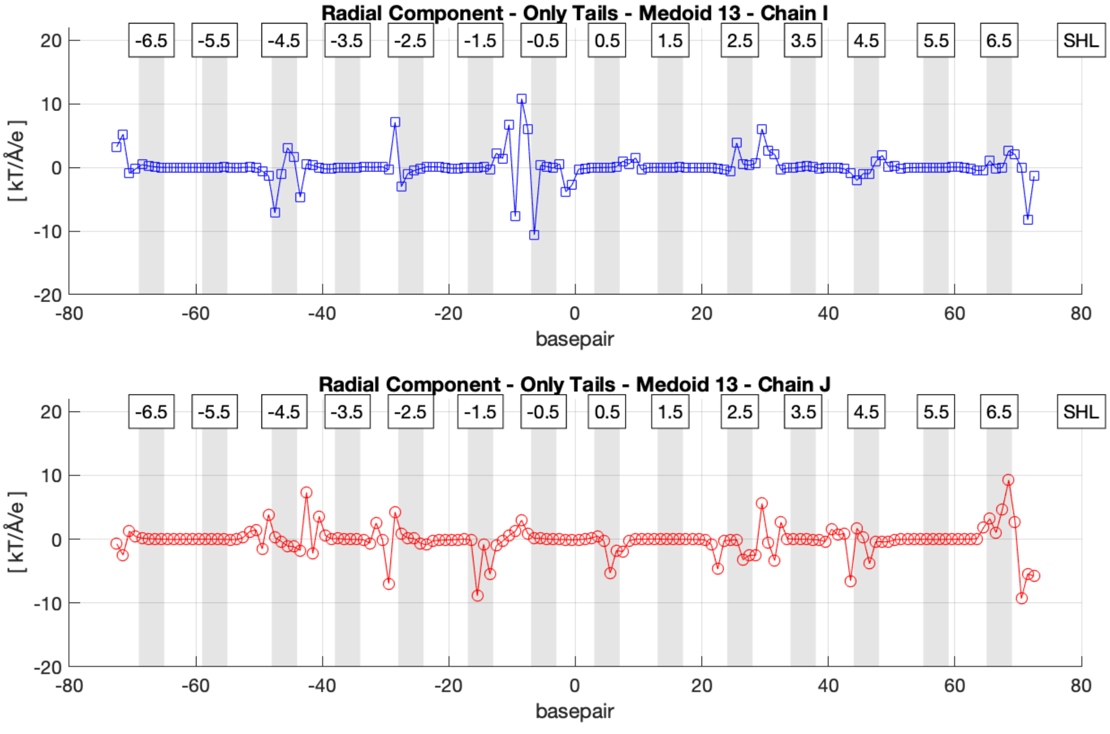


Figure 52: Radial component of electrostatic force on phosphates of medoid13 (non-normalised): "only tails" structure

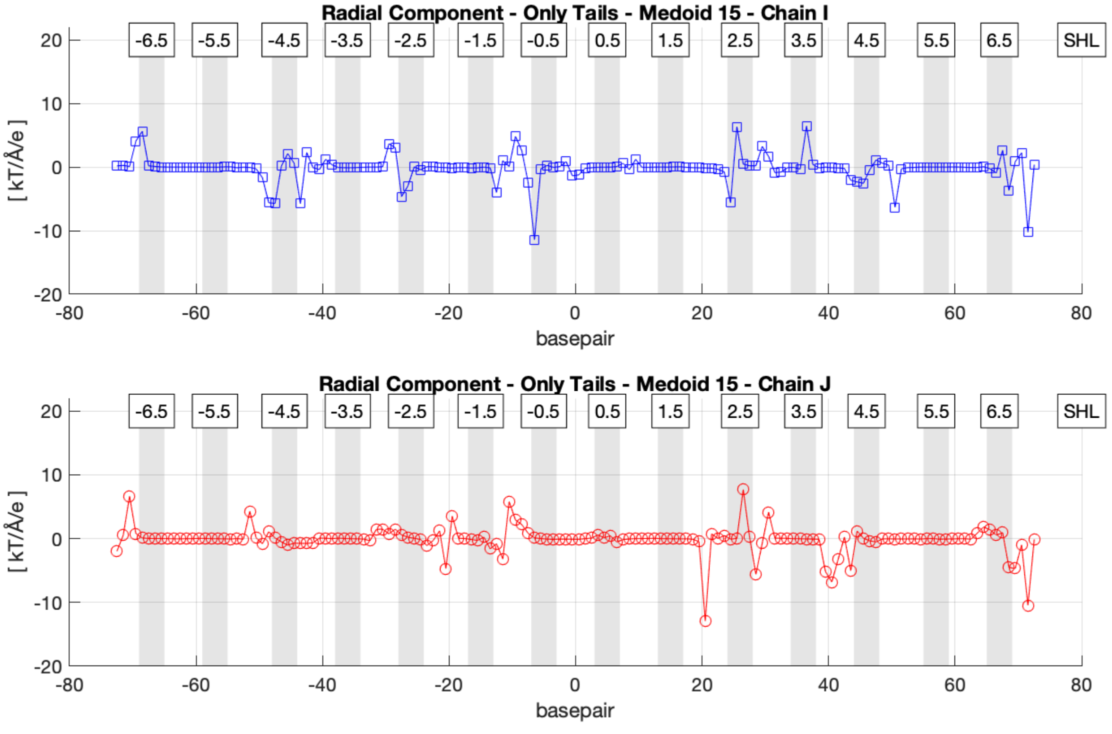


Figure 53: Radial component of electrostatic force on phosphates of medoid15 (non-normalised): "only tails" structure

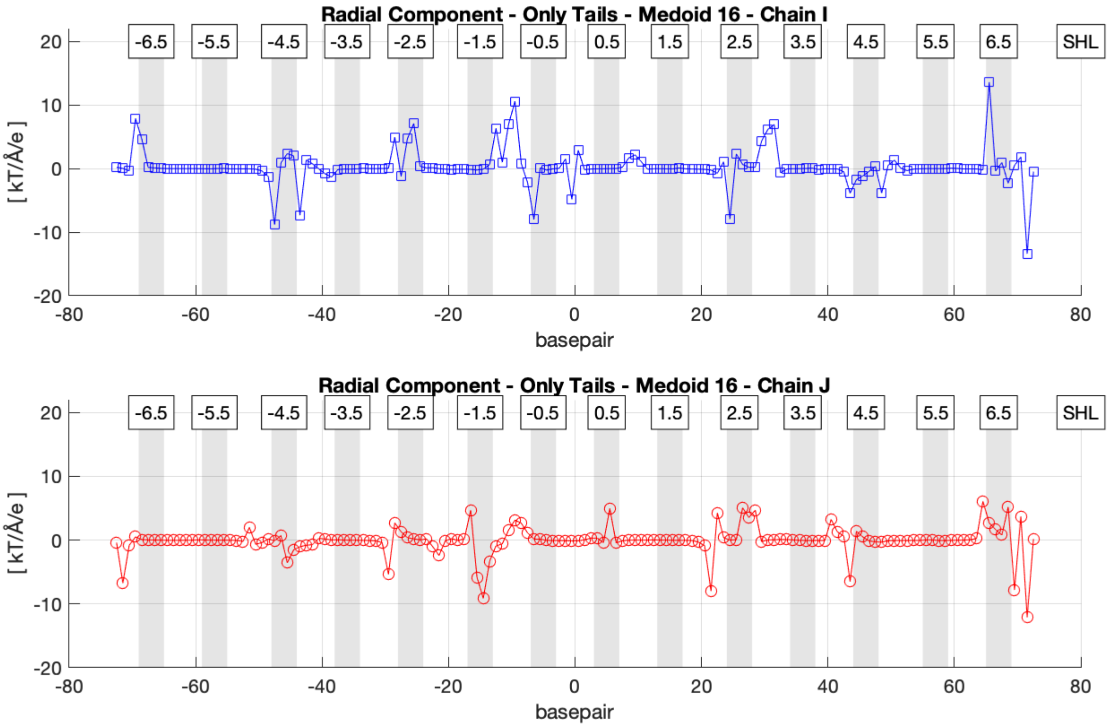


Figure 54: Radial component of electrostatic force on phosphates of medoid16 (non-normalised): "only tails" structure

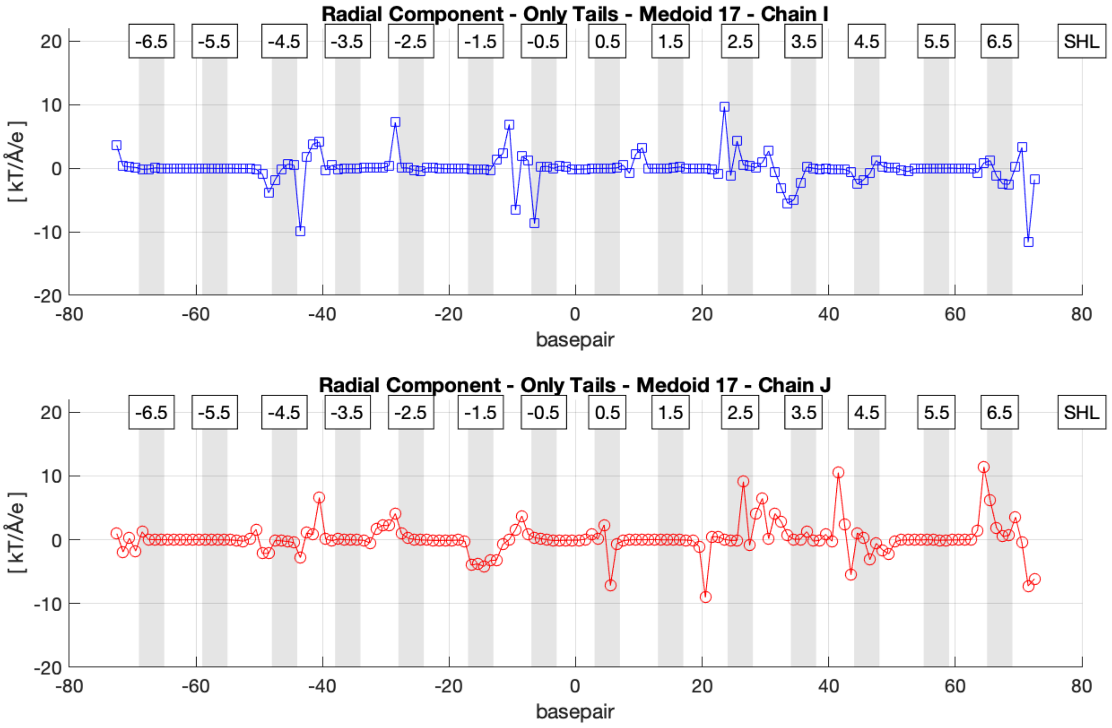


Figure 55: Radial component of electrostatic force on phosphates of medoid17 (non-normalised): "only tails" structure

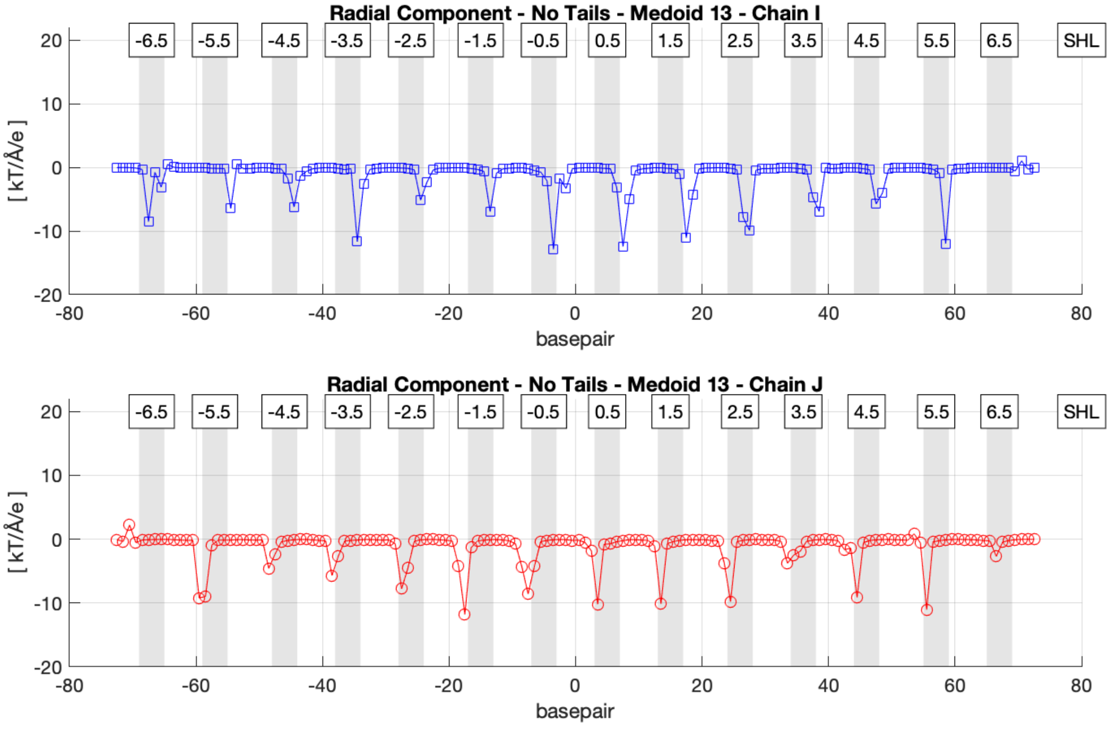


Figure 56: Radial component of electrostatic force on phosphates of medoid13 (non-normalised): "no tails" structure

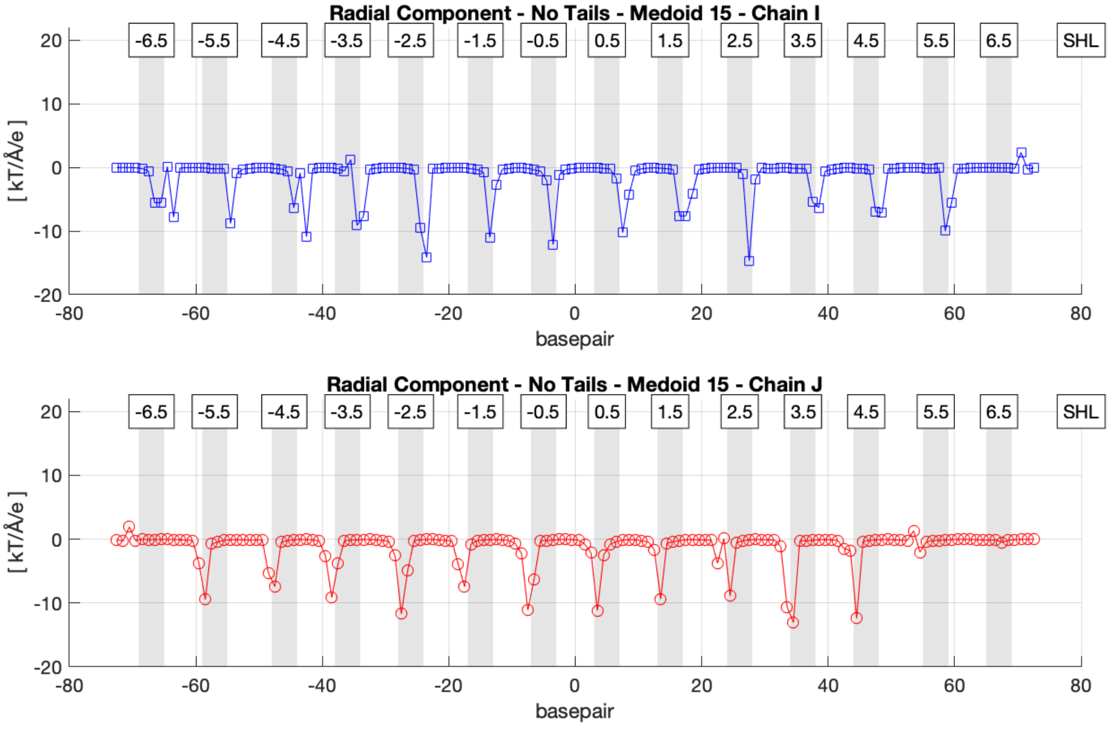


Figure 57: Radial component of electrostatic force on phosphates of medoid15 (non-normalised): "no tails" structure

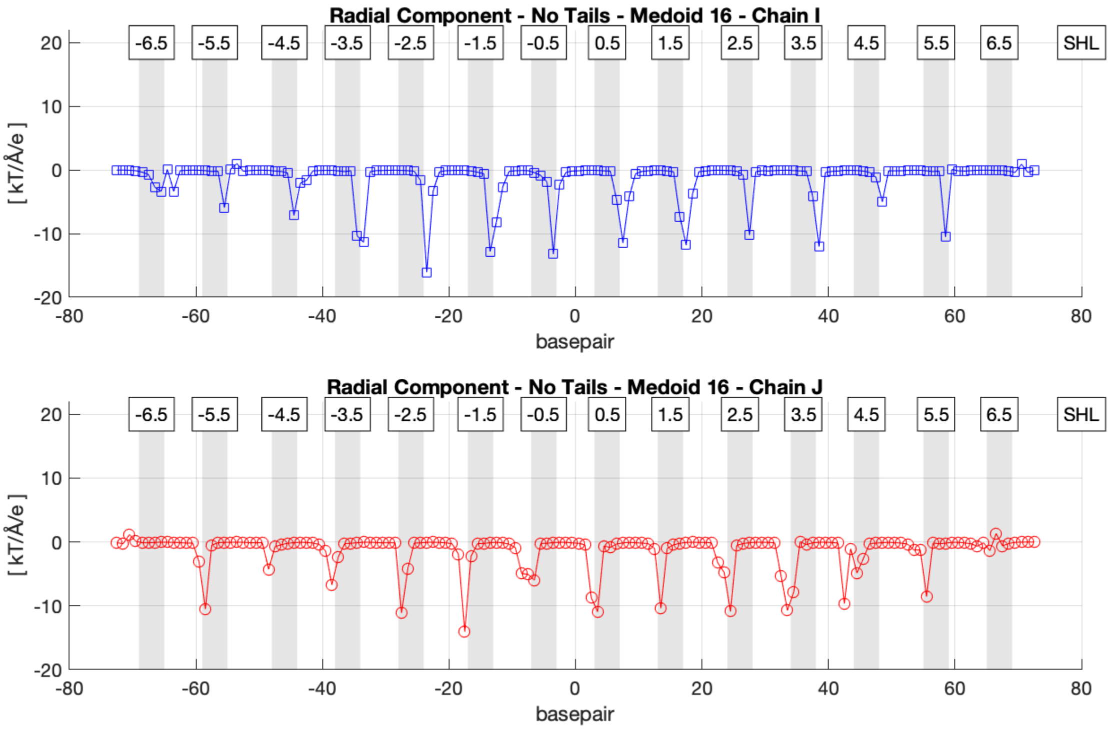


Figure 58: Radial component of electrostatic force on phosphates of medoid16 (non-normalised): "no tails" structure

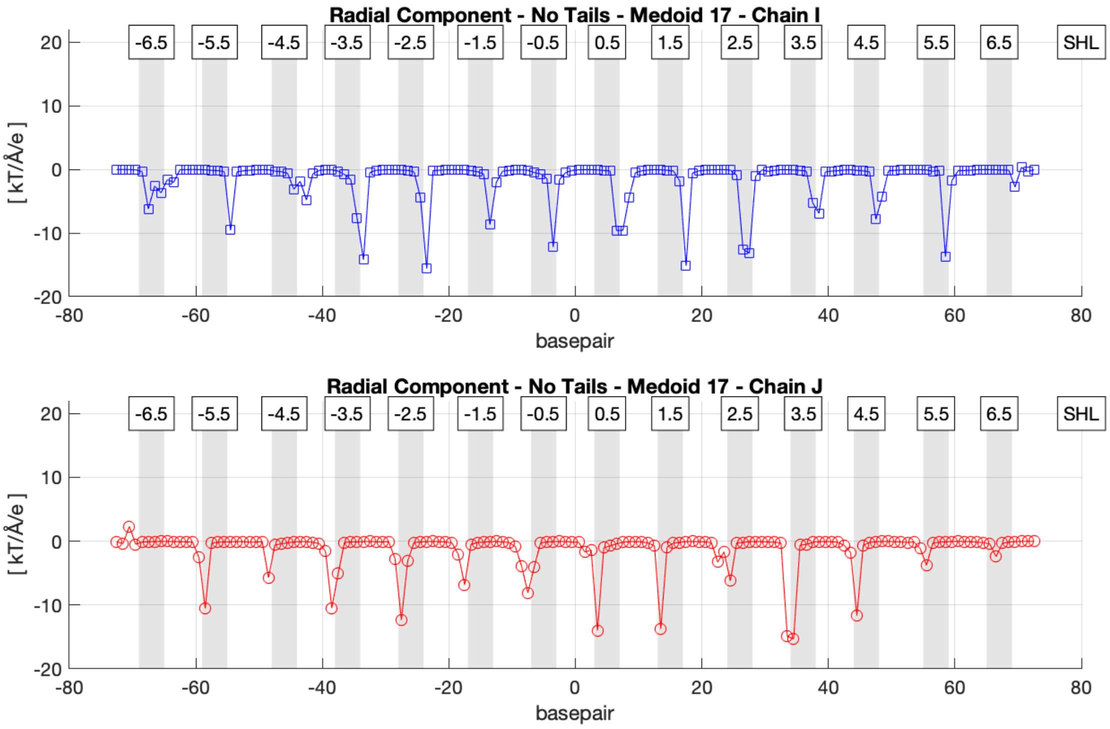


Figure 59: Radial component of electrostatic force on phosphates of medoid17 (non-normalised): "no tails" structure

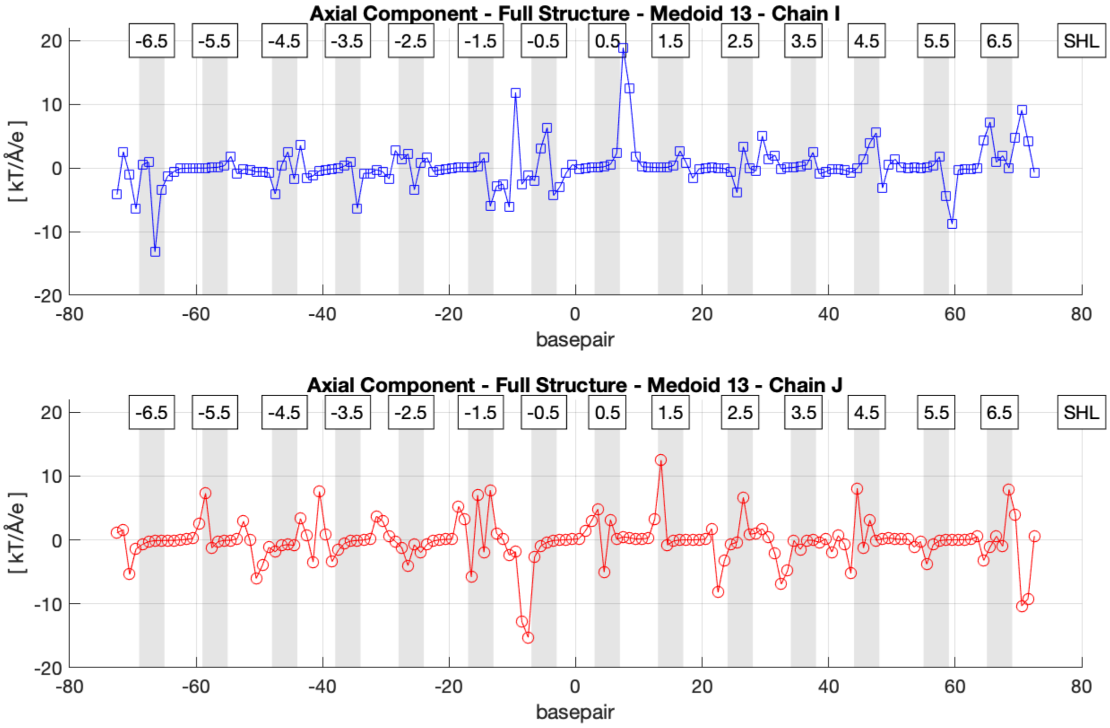


Figure 60: Axial component of electrostatic force on phosphates of medoid13 (non-normalised): full structure

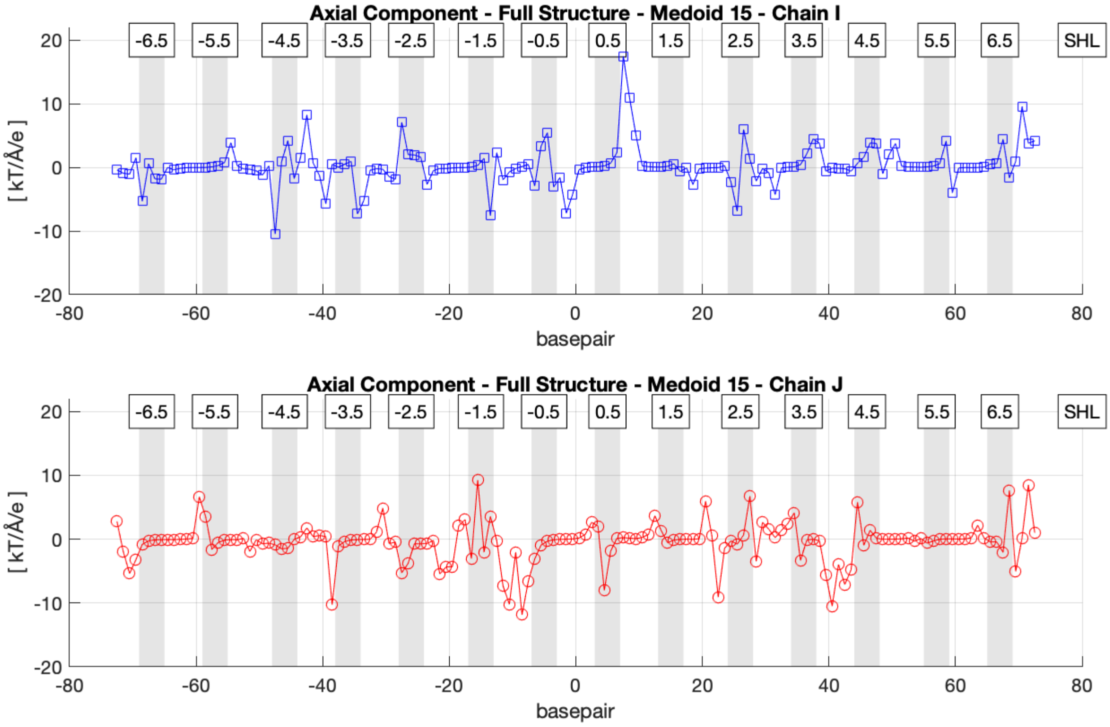


Figure 61: Axial component of electrostatic force on phosphates of medoid15 (non-normalised): full structure

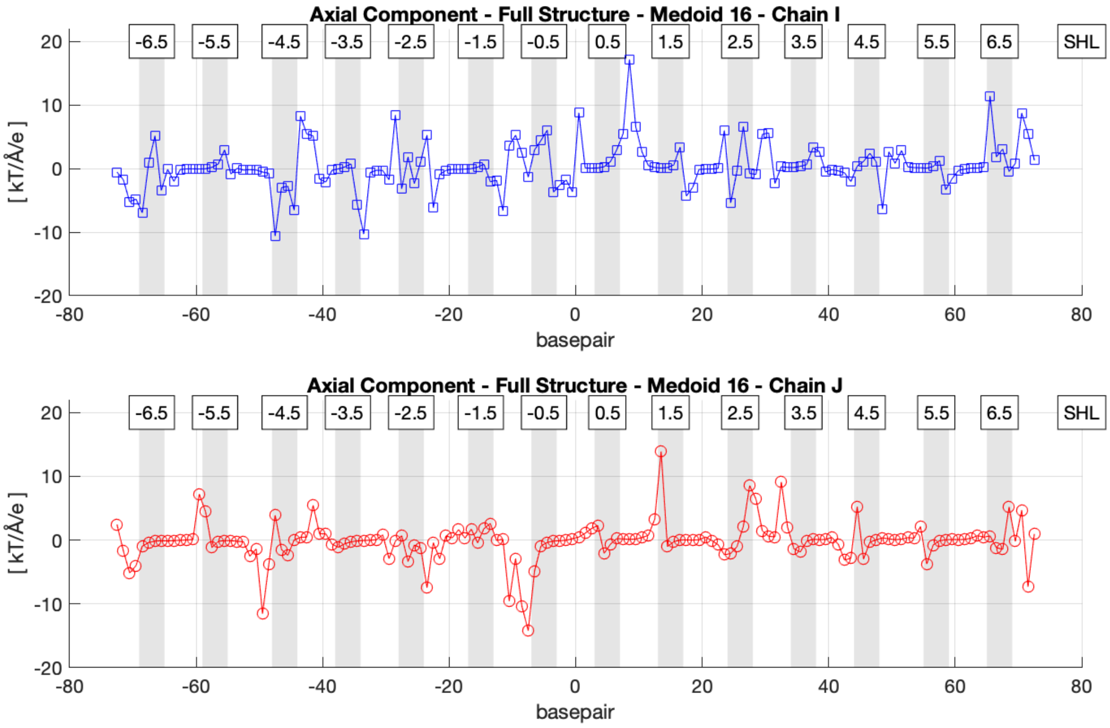


Figure 62: Axial component of electrostatic force on phosphates of medoid16 (non-normalised): full structure

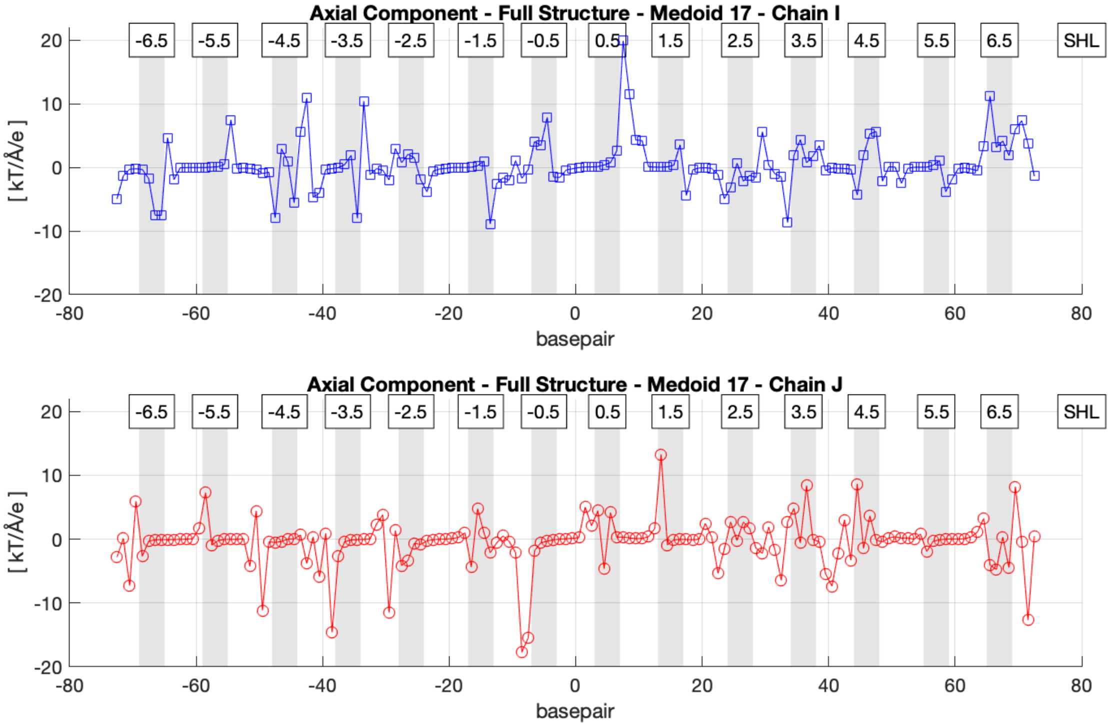


Figure 63: Axial component of electrostatic force on phosphates of medoid17 (non-normalised): full structure

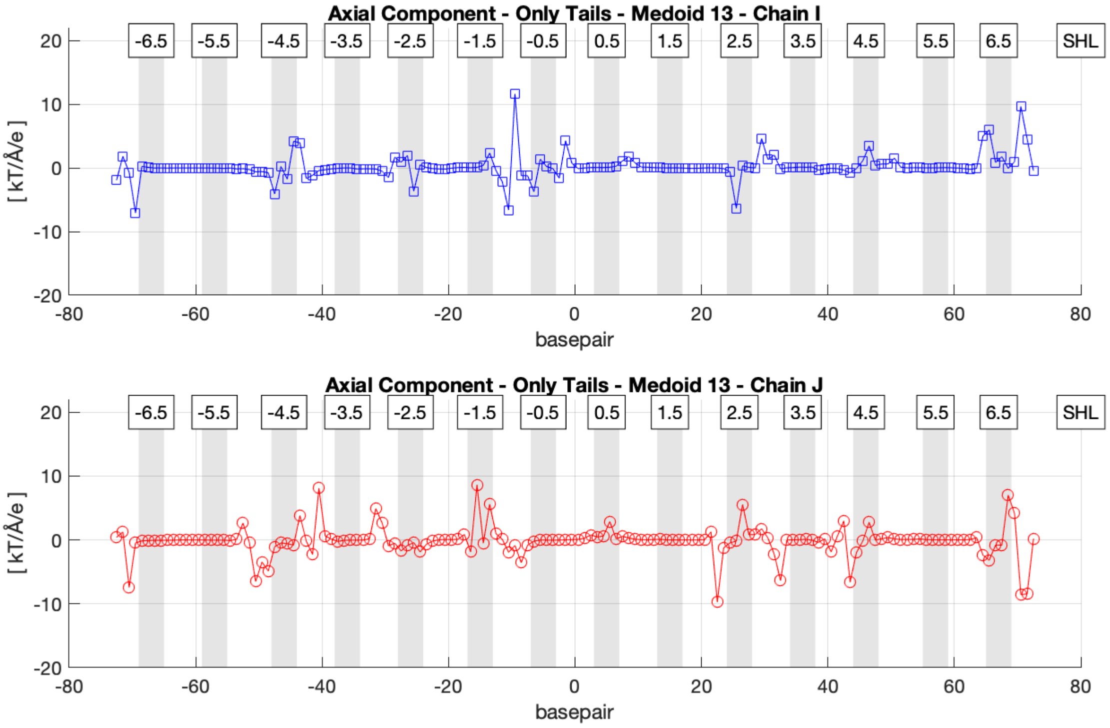


Figure 64: Axial component of electrostatic force on phosphates of medoid13 (non-normalised): "only tails" structure

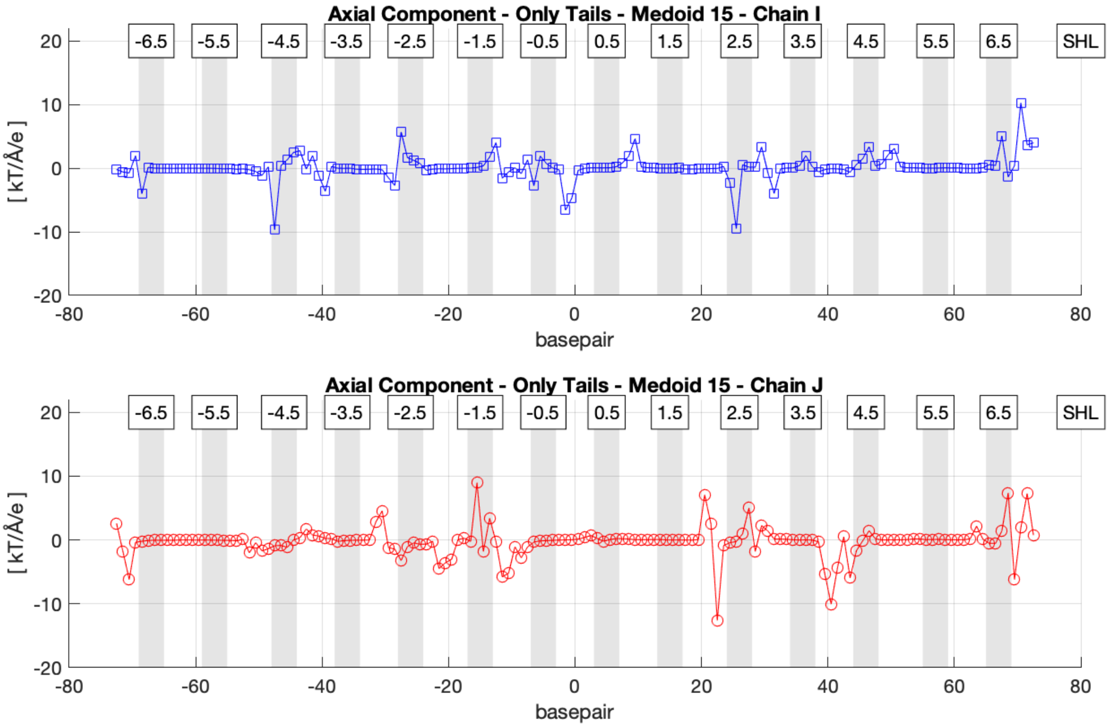


Figure 65: Axial component of electrostatic force on phosphates of medoid15 (non-normalised): "only tails" structure

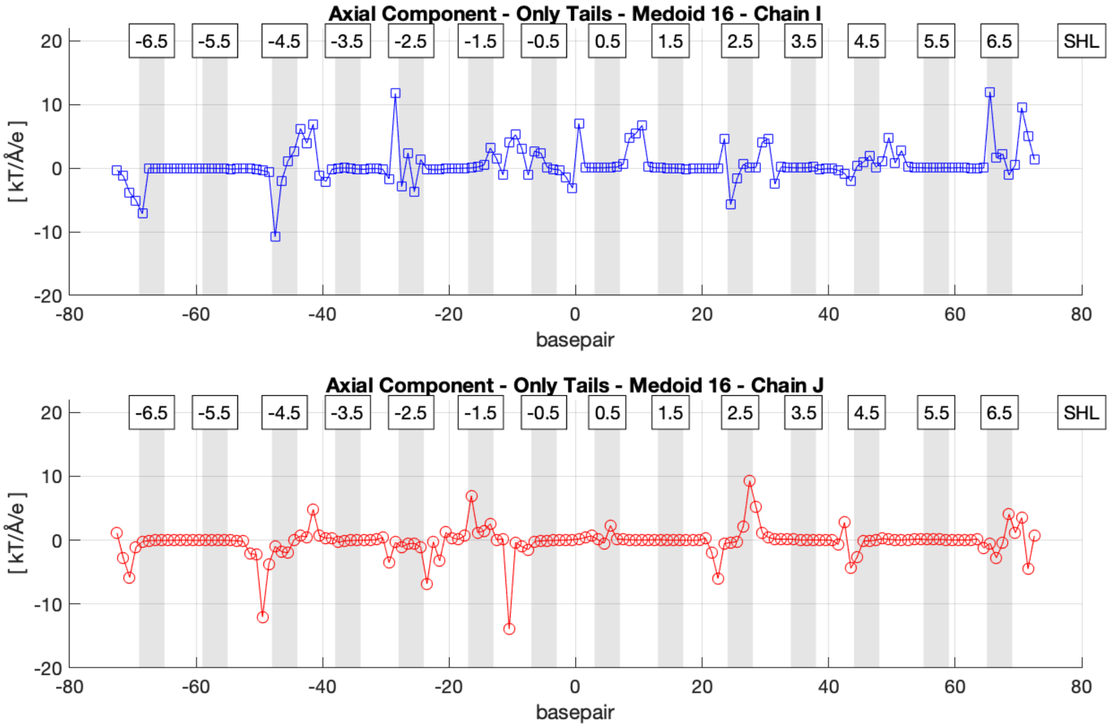


Figure 66: Axial component of electrostatic force on phosphates of medoid16 (non-normalised): "only tails" structure

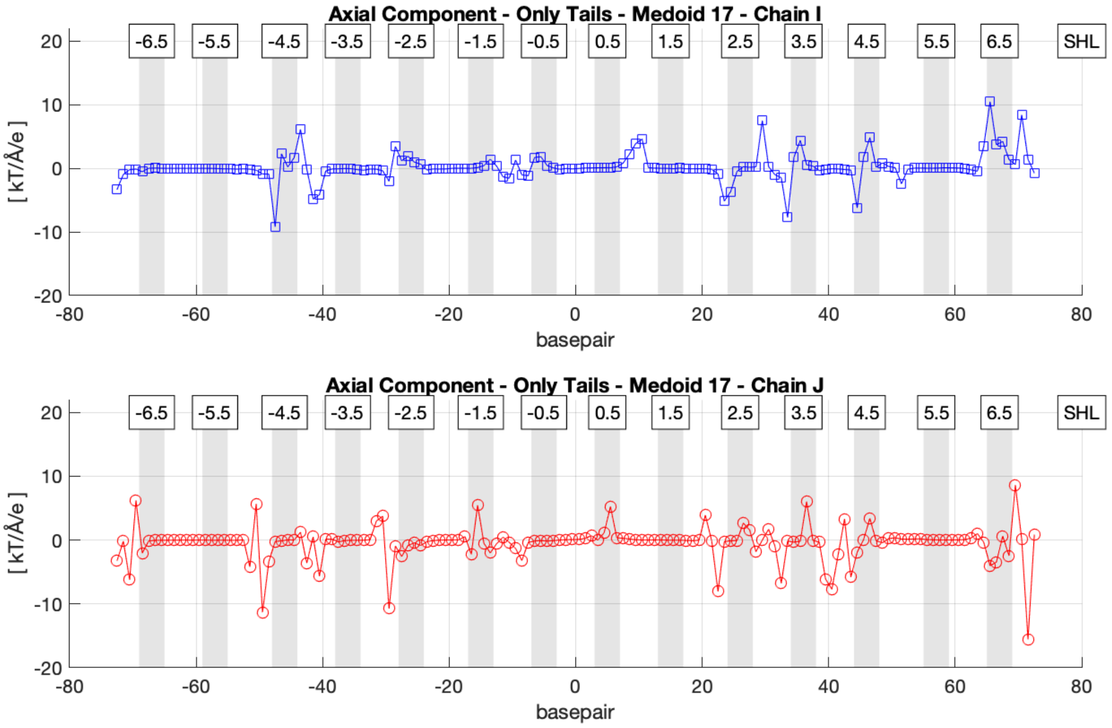


Figure 67: Axial component of electrostatic force on phosphates of medoid17 (non-normalised): "only tails" structure

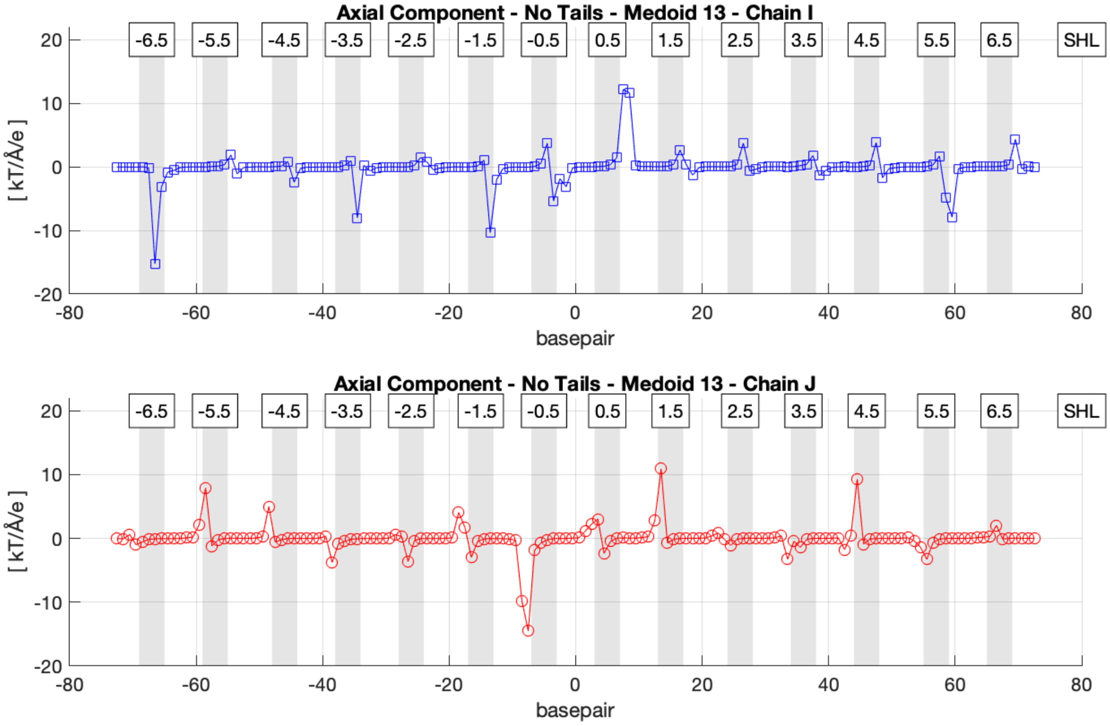


Figure 68: Axial component of electrostatic force on phosphates of medoid13 (non-normalised): "no tails" structure

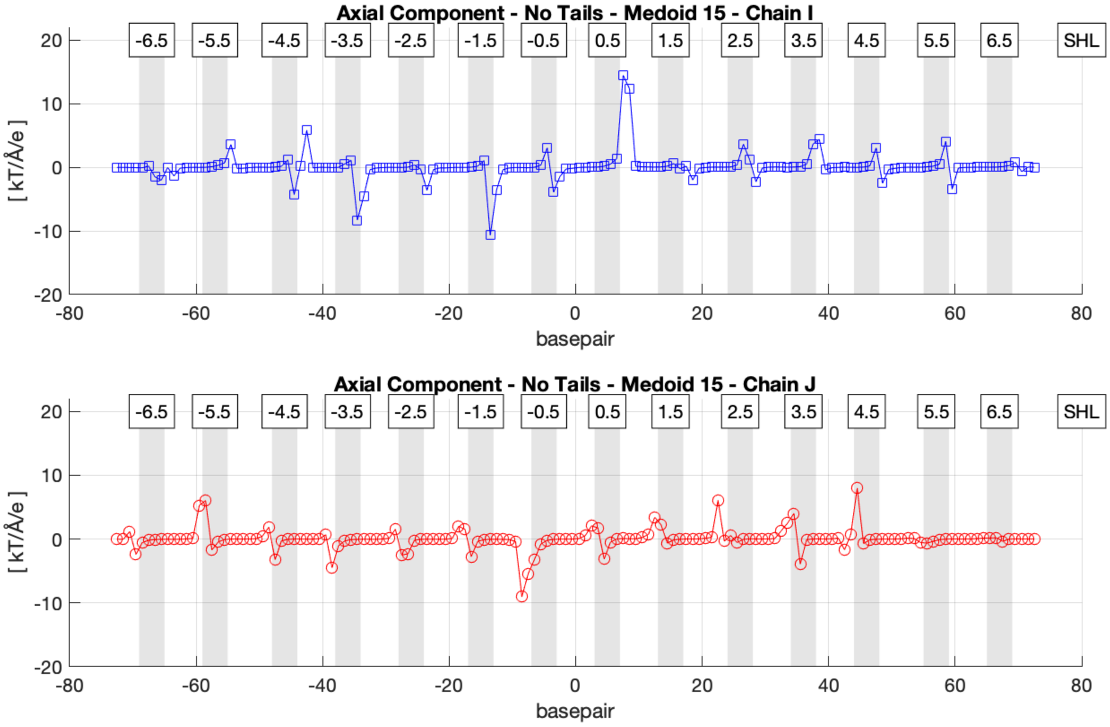
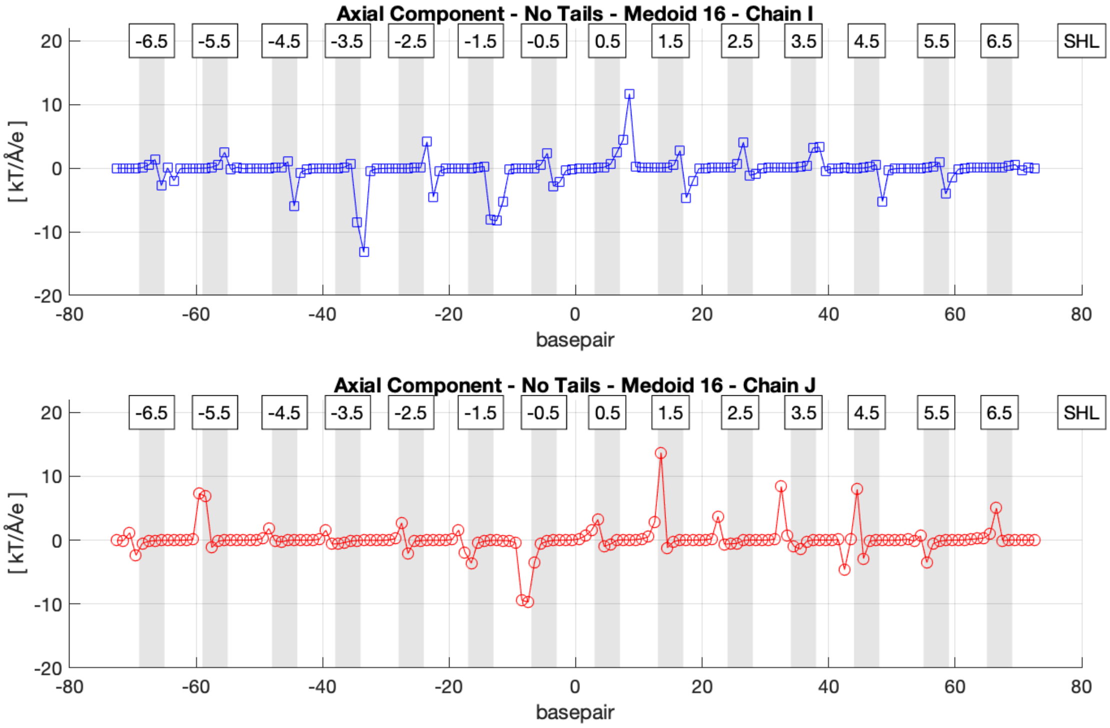
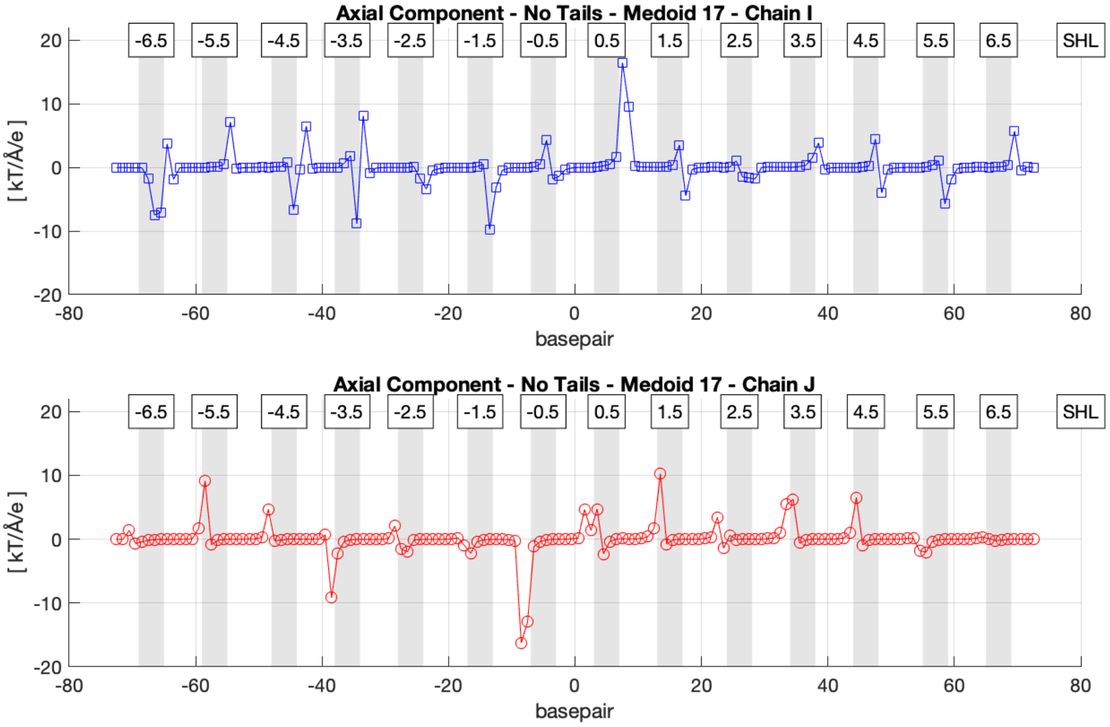


Figure 69: Axial component of electrostatic force on phosphates of medoid15 (non-normalised): "no tails" structure


Axial Component - No Tails - Medoid 16 - Chain I

Axial Component - No Tails - Medoid 16 - Chain J

Figure 70: Axial component of electrostatic force on phosphates of medoid16 (non-normalised): "no tails" structure


Axial Component - No Tails - Medoid 17 - Chain I

Axial Component - No Tails - Medoid 17 - Chain J

Figure 71: Axial component of electrostatic force on phosphates of medoid17 (non-normalised): "no tails" structure

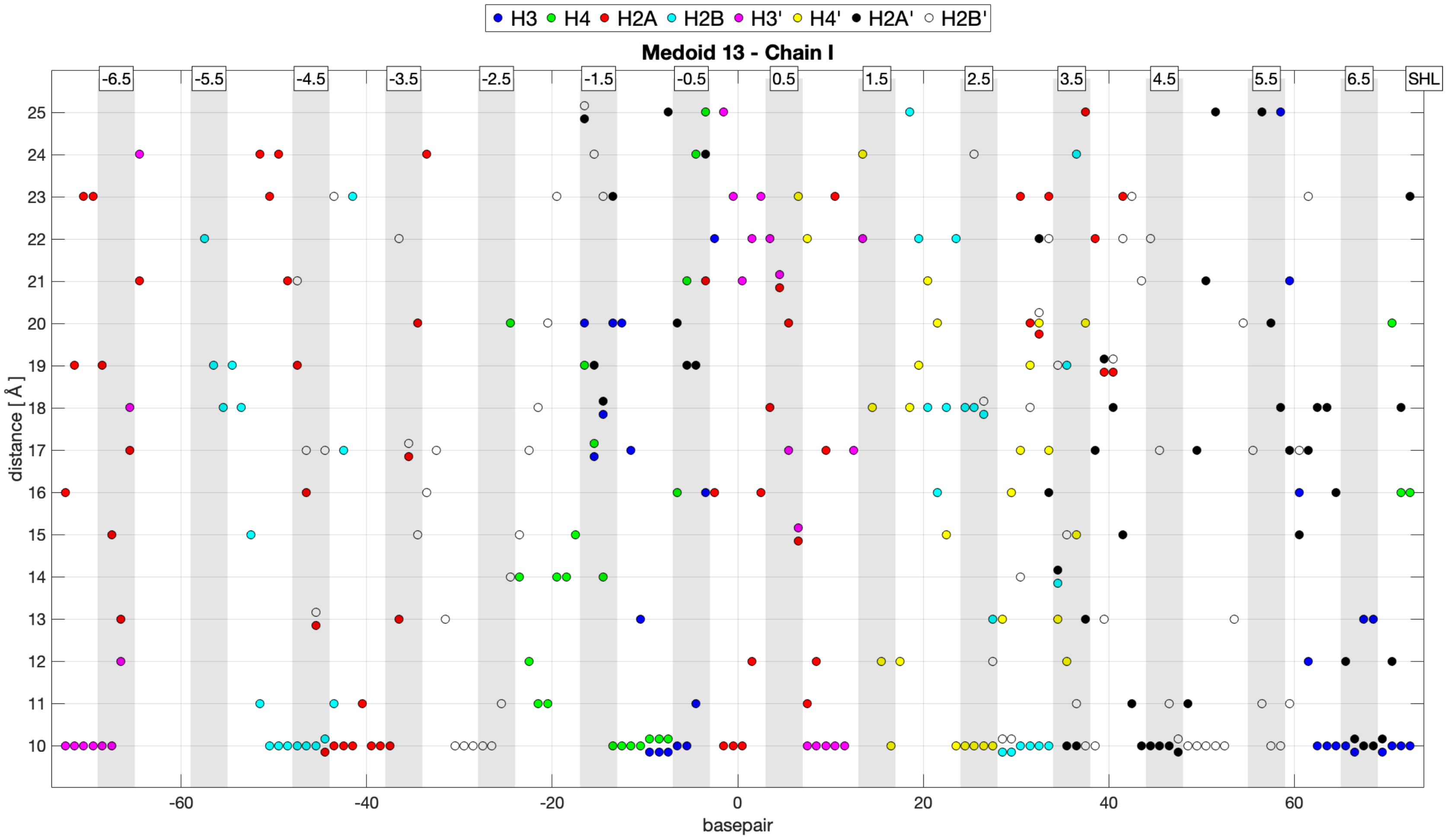


Figure 72: Medoid 13: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain I. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

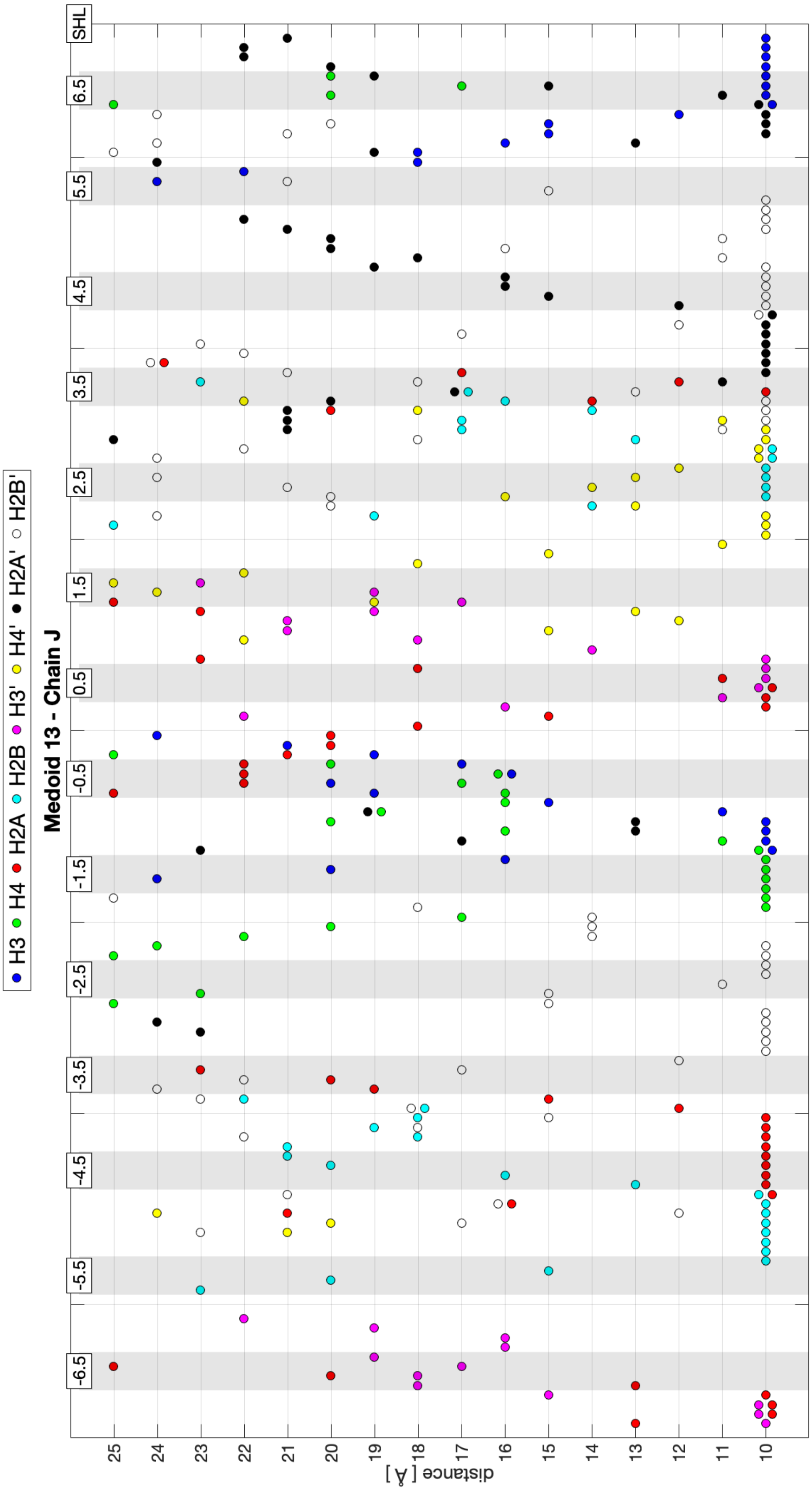


Figure 73: Medoid 13: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain J. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

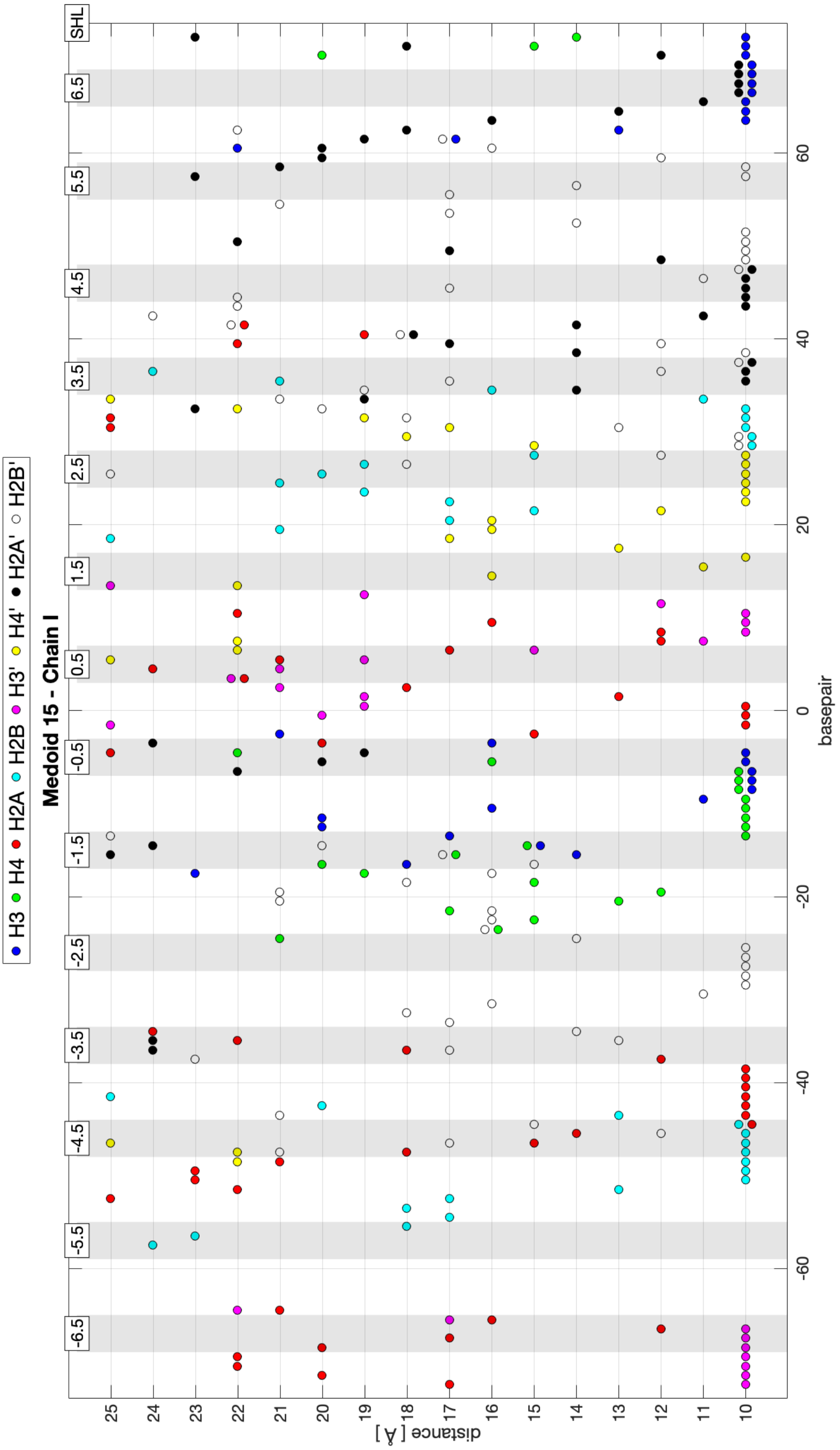


Figure 74: Medoid 15: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain I. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

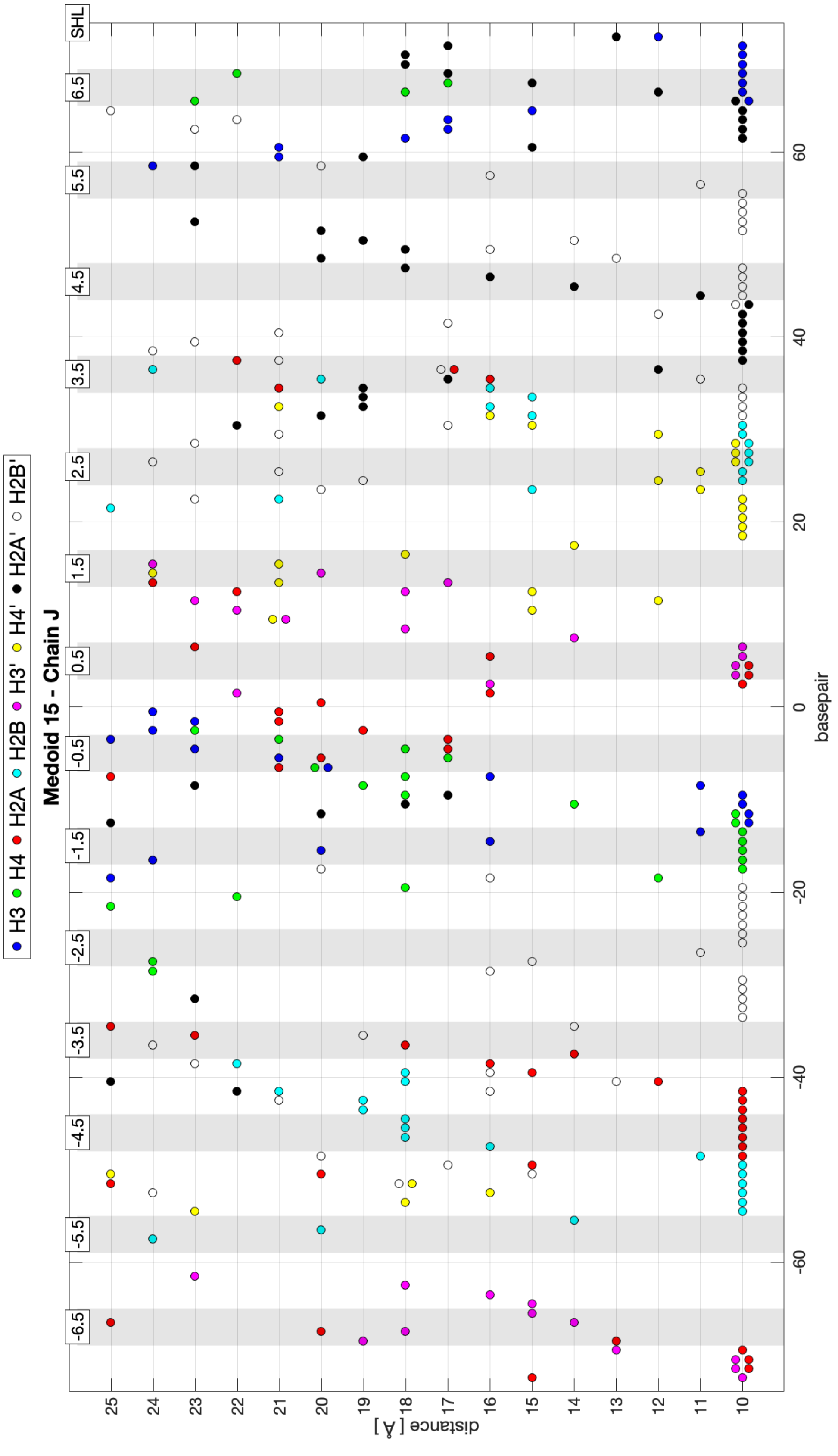


Figure 75: Medoid 15: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain J. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

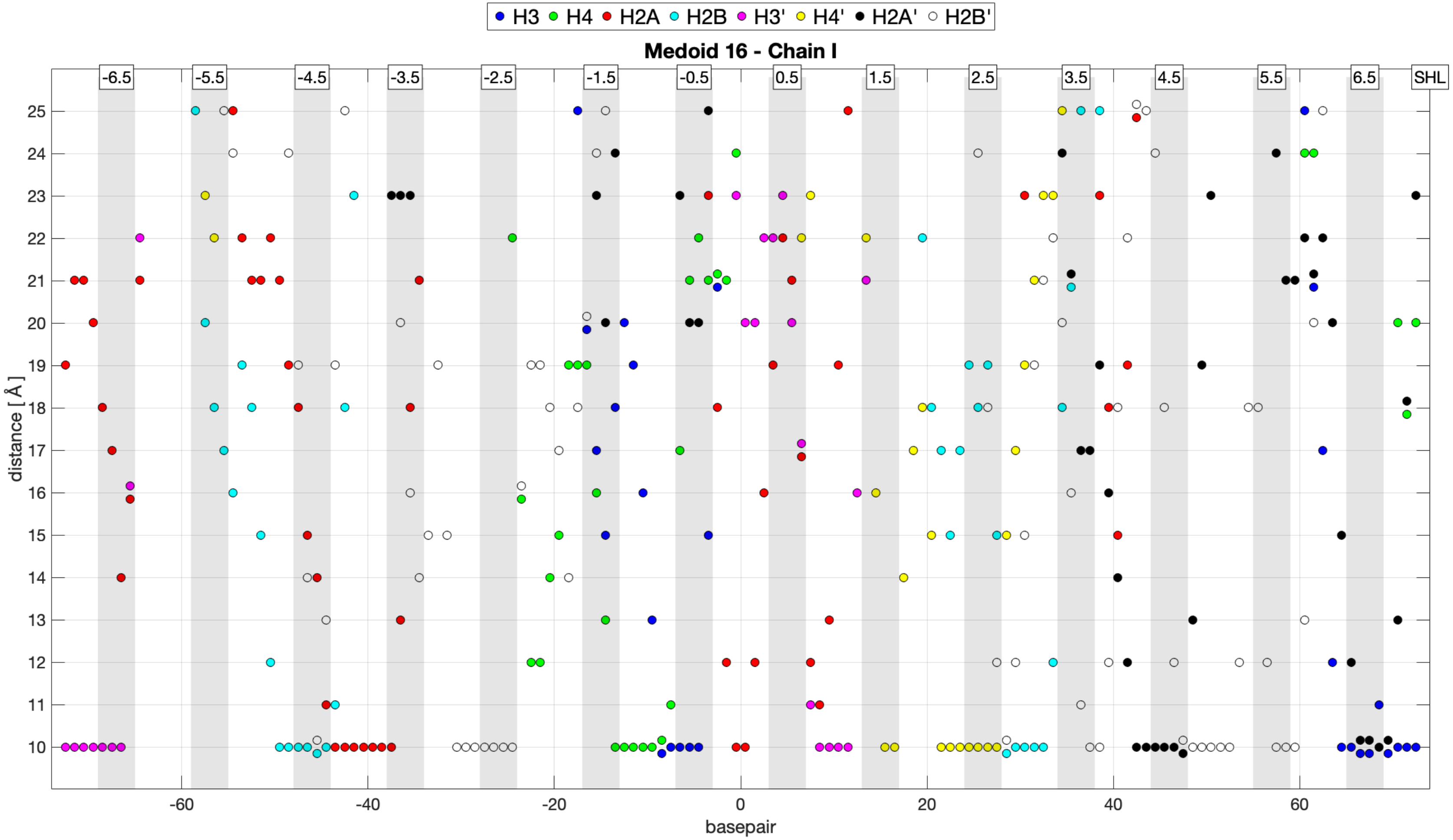


Figure 76: Medoid 16: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain I. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

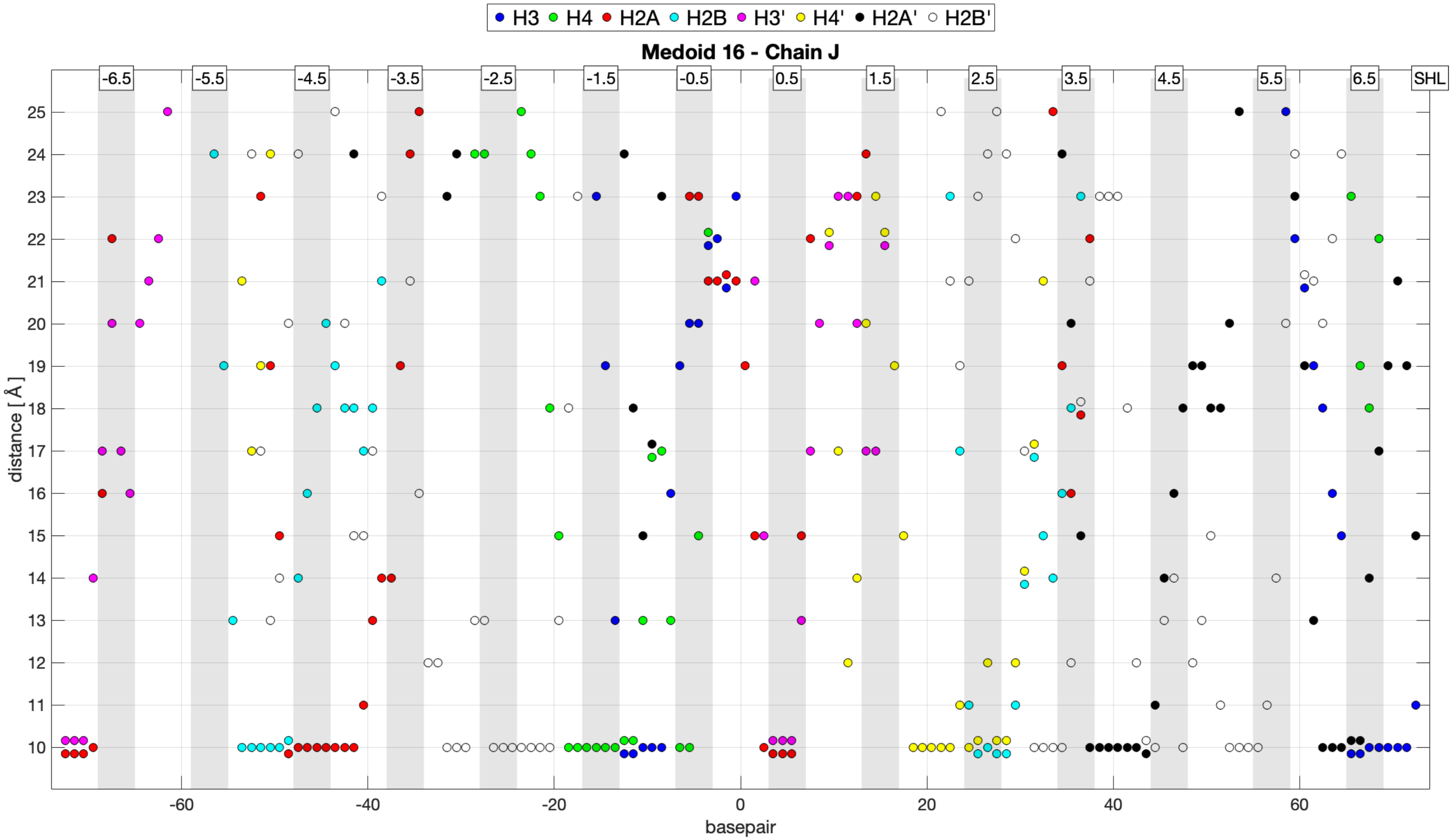


Figure 77: Medoid 16: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain J. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

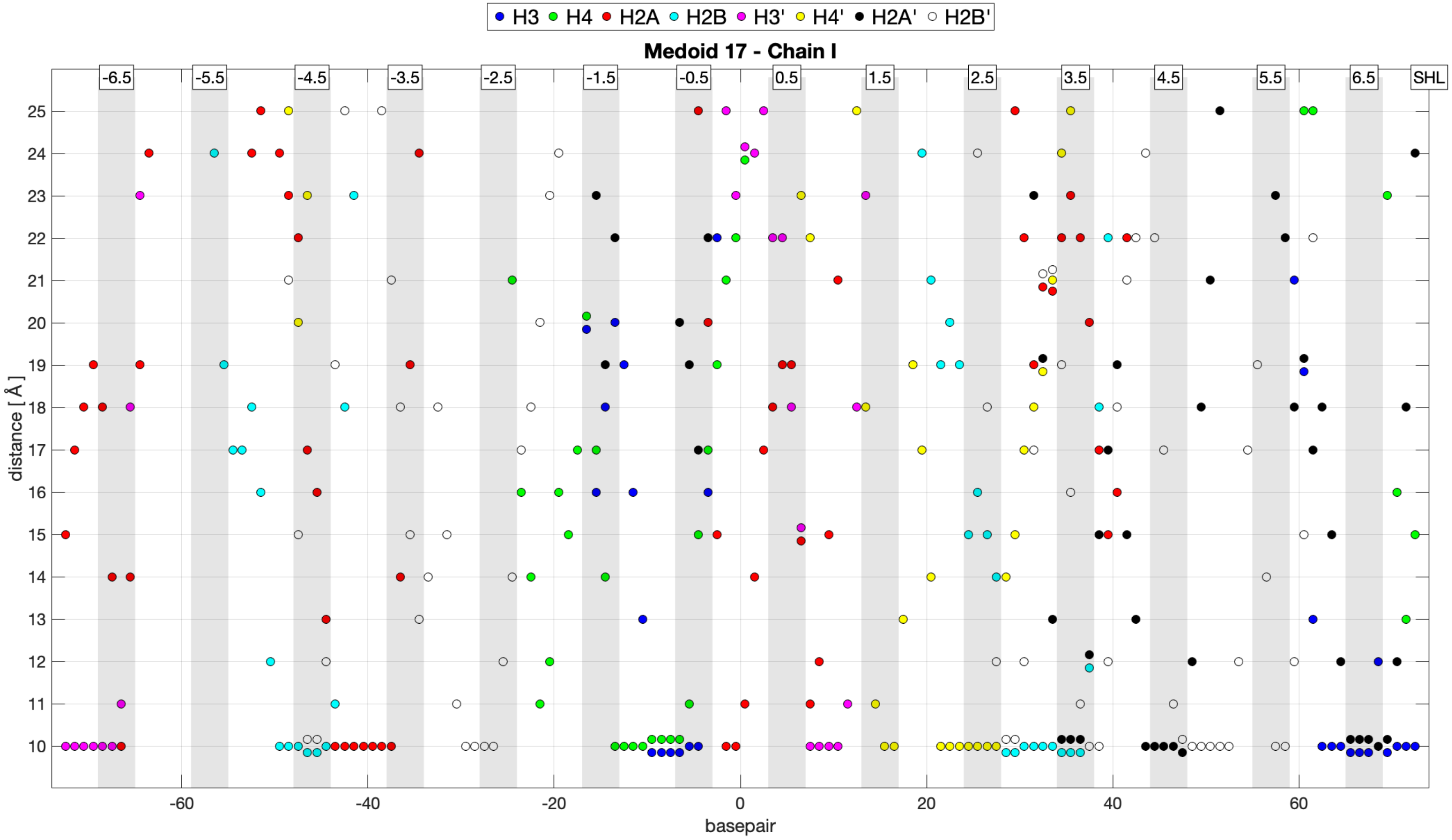


Figure 78: Medoid 17: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain I. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

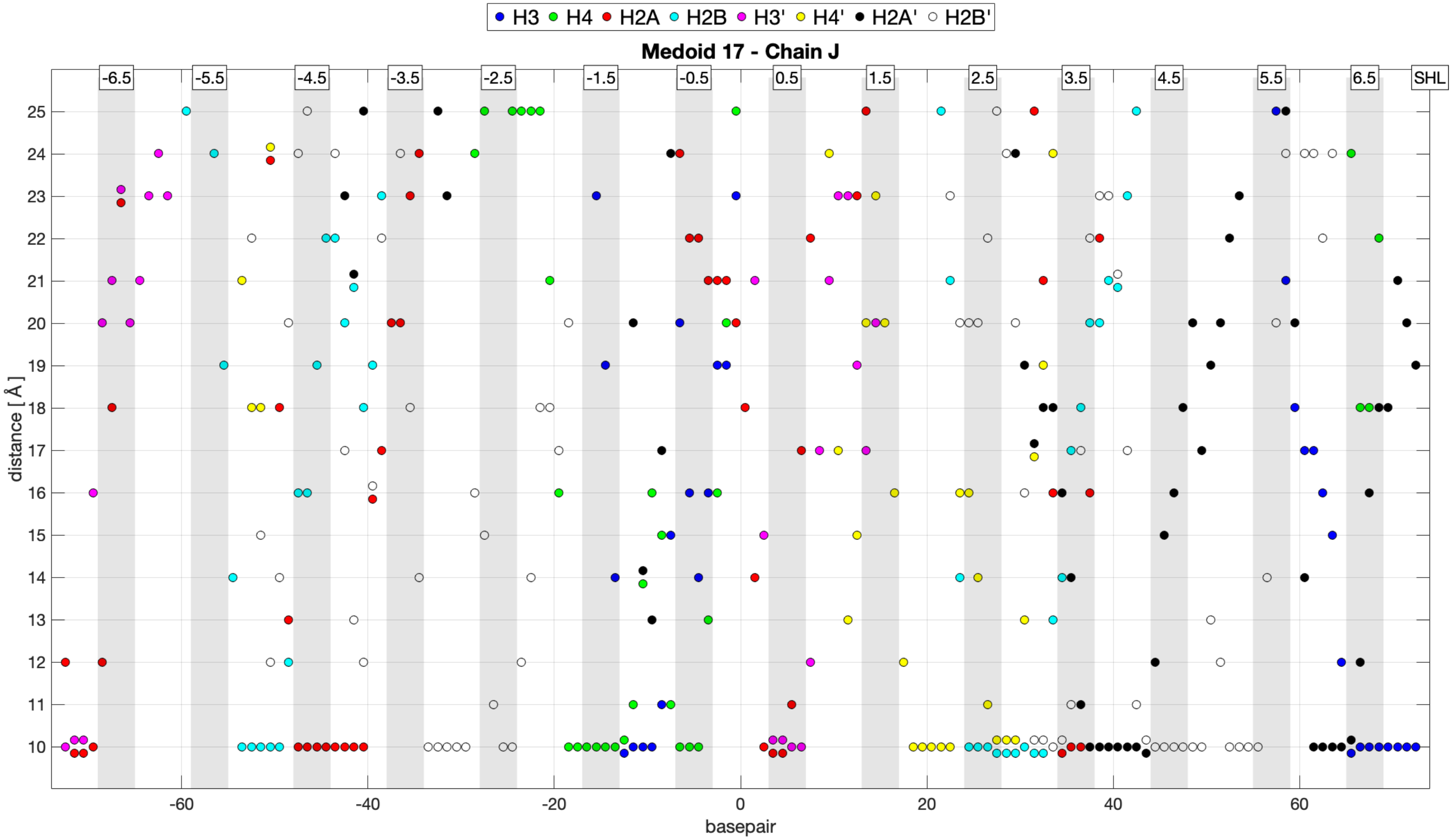


Figure 79: Medoid 17: Plot indicating the threshold in which each histone tail comes into contact with each phosphate of chain J. Thresholds vary from a distance of 10 to 25Å. Histone tails are indicated by different colours.

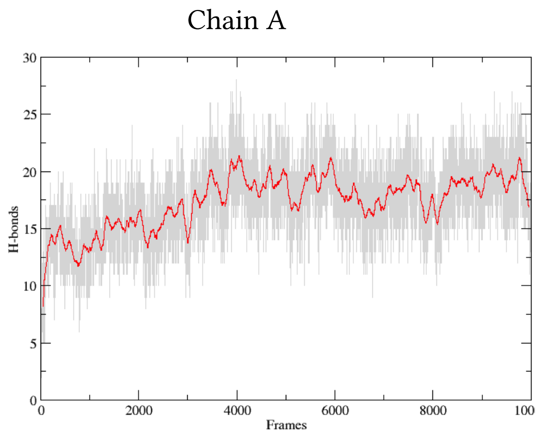


Figure 80: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain A (histone H3) during the 1$\mu$s MD trajectory.

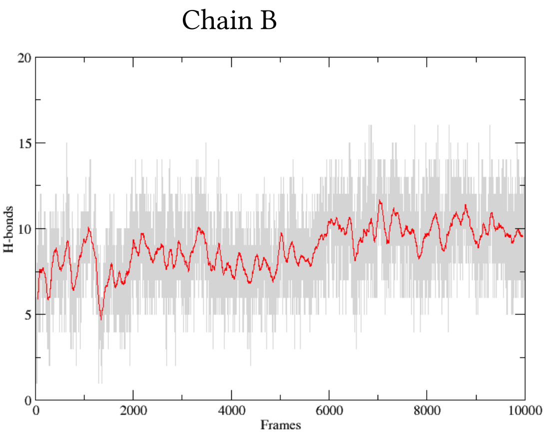


Figure 81: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain B (histone H4) during the 1$\mu$s MD trajectory.

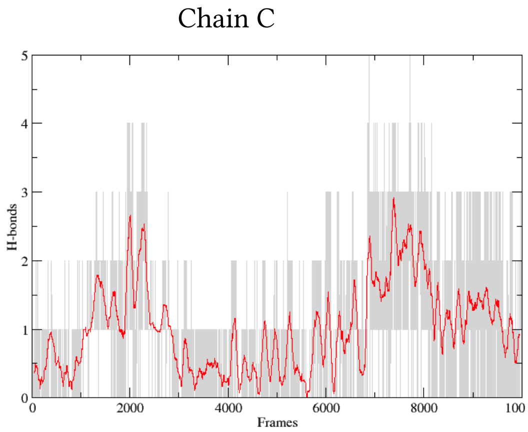


Figure 82: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain C (histone H2A) during the 1$\mu$s MD trajectory.

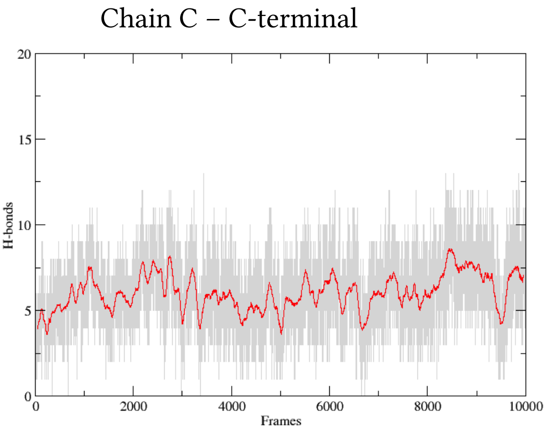


Figure 83: Hydrogen bonds formed between atoms of the C-terminal tail and atoms of the DNA in chain C (histone H2A) during the 1$\mu$s MD trajectory.

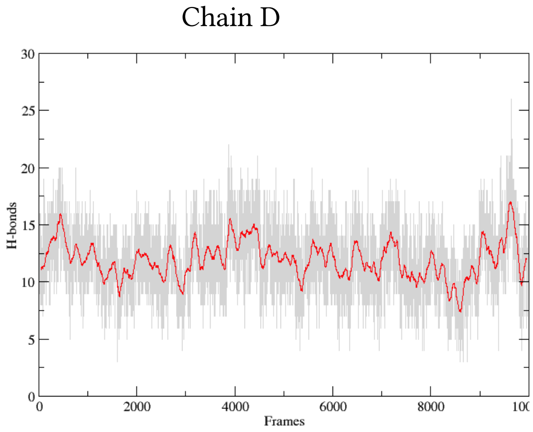


Figure 84: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain D (histone H2B) during the 1$\mu$s MD trajectory.

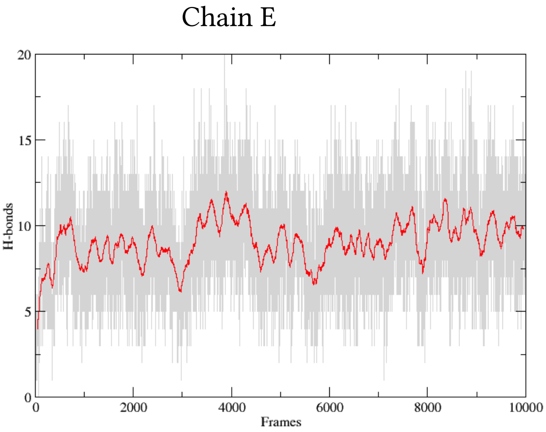


Figure 85: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain E (histone H3') during the 1$\mu$s MD trajectory.

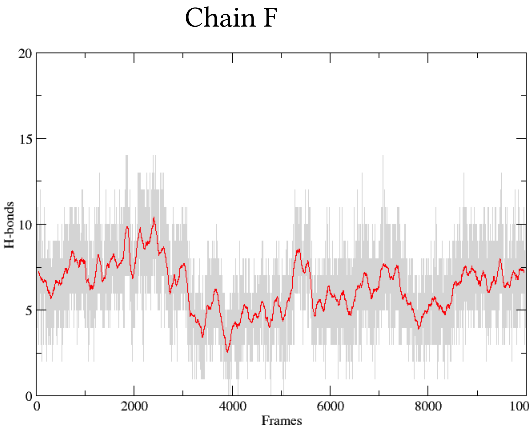


Figure 86: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain F (histone H4') during the 1$\mu$s MD trajectory.

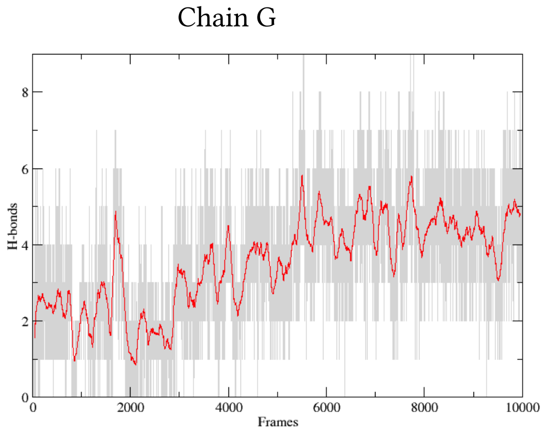


Figure 87: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain G (histone H2A') during the 1$\mu$s MD trajectory.

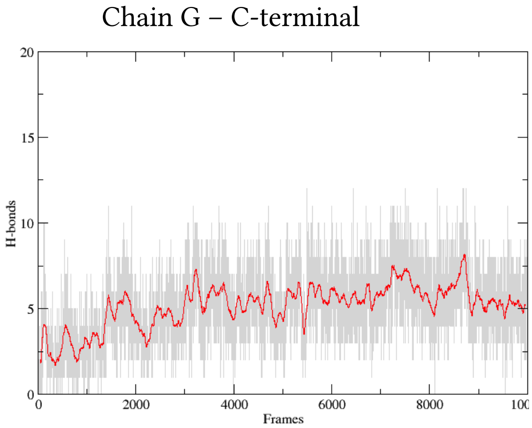


Figure 88: Hydrogen bonds formed between atoms of the C-terminal tail and atoms of the DNA in chain G (histone H2A') during the 1$\mu$s MD trajectory.

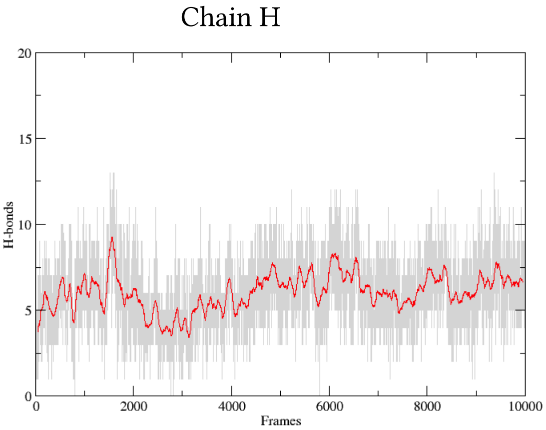


Figure 89: Hydrogen bonds formed between atoms of the N-terminal tail and atoms of the DNA in chain H (histone H2B') during the 1$\mu$s MD trajectory.

# Appendix B: Supplementary Information of Chapter 5

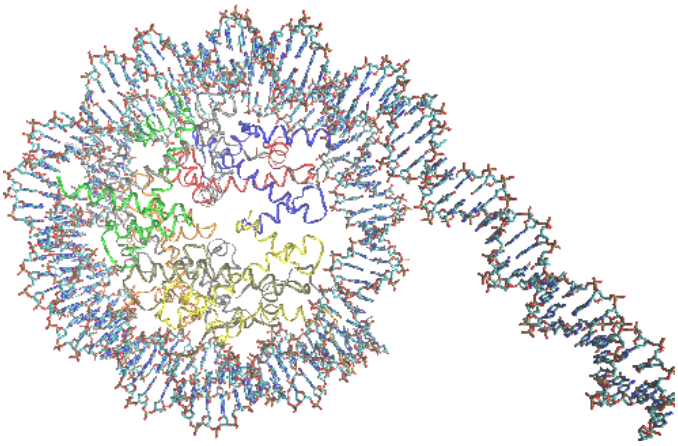


Figure 1: 3AFA nucleosome crystal structure with 25bps of linker DNA attached.

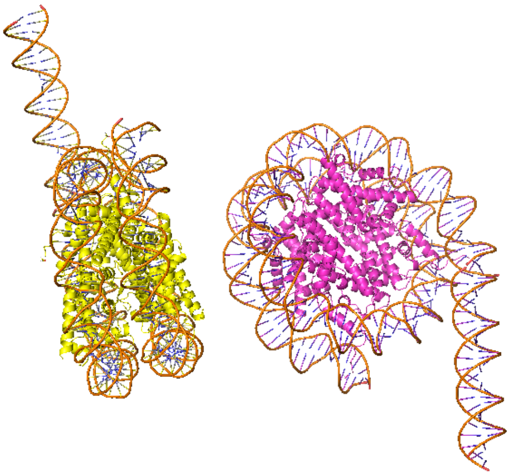


Figure 2: Example of structure composed of two 3AFA nucleosomes with linker DNA attached to each on which DelPhi calculations were performed. As in the case of the medoid2 structures, we explored a wide range of distances and relative orientations that represent interactions between nucleosomes in different parts of the chromatin fibre.

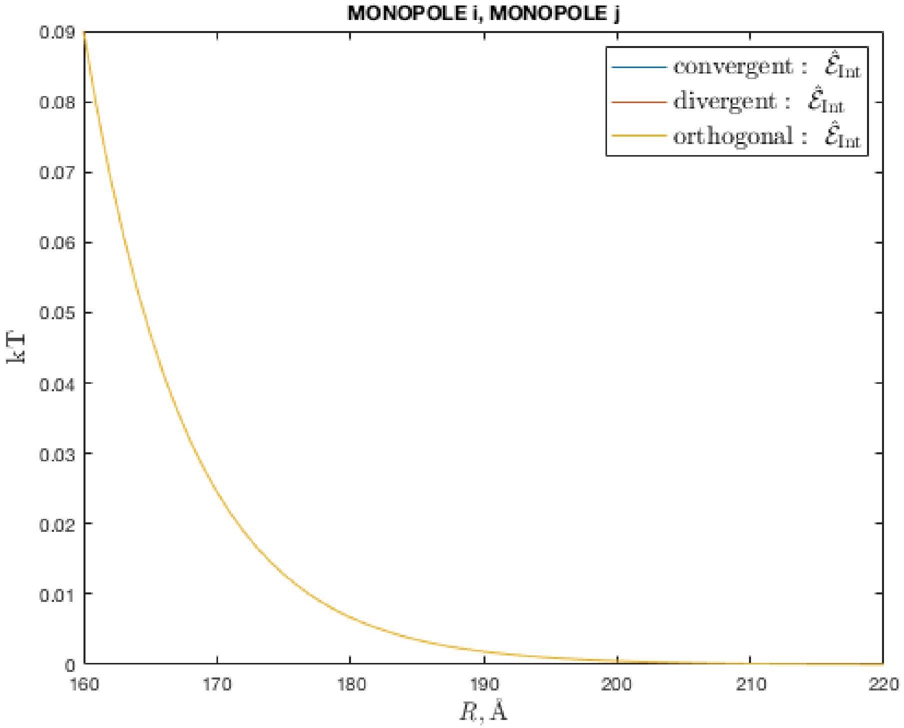


Figure 3: Electrostatic Interaction energy $\hat{\mathcal{E}}_{\mathrm{Int}}$ in the case of two interacting monopoles as defined in eq. (5.5.1). The value of the energy is given as a function of the distance between the centres of the spheres.

## Expressions for the Expansion Coefficients of Electrostatic Potentials

Henceforth $i = \overline{1,2}$ and $j = 3 - i$. Let us represent the potential expansion coefficients $G_{nm,i}$, $L_{nm,i}$, $0 \le m \le n$, and $H_{nm,i}$, $M_{nm,i}$, $1 \le m \le n$, by the infinite-size column-vectors $\mathbf{G}_{m,i}$, $\mathbf{H}_{m,i}$, $\mathbf{L}_{m,i}$, $\mathbf{M}_{m,i}$ defined as follows for arbitrary fixed $m$:

$$\mathbf{G}_{m,i} = \left\{ G_{nm,i} \right\}_{n\ge m}, \quad \mathbf{H}_{m,i} = \left\{ H_{nm,i} \right\}_{n\ge m}, \quad \mathbf{L}_{m,i} = \left\{ L_{nm,i} \right\}_{n\ge m}, \quad \mathbf{M}_{m,i} = \left\{ M_{nm,i} \right\}_{n\ge m}. \tag{1}$$

Let us also introduce the auxiliary matrices

$$\begin{aligned} \mathsf{A}_m(\tilde{a}_i, \varepsilon_i) &:= \text{diagonal}\left\{ \alpha_n(\tilde{a}_i, \varepsilon_i) \right\}_{n\ge m}, \qquad & \mathsf{B}_m(\tilde{a}_i, \varepsilon_i, \tilde{R}) &:= \left\{ \beta_{nml}(\tilde{a}_i, \varepsilon_i, \tilde{R}) \right\}_{n\ge m,\ l\ge m}, \\ \mathsf{K}_m(\tilde{a}_i) &:= \text{diagonal}\left\{ \frac{K_{n+1/2}(\tilde{a}_i)}{\tilde{a}_i^{n+1/2}} \right\}_{n\ge m}, & \mathbb{B}_m(\tilde{a}_i, \tilde{R}) &:= \left\{ \frac{b_{nml}(\tilde{a}_i, \tilde{R})}{\tilde{a}_i^{n}} \right\}_{n\ge m,\ l\ge m}, \end{aligned}$$

$$\mathsf{N}_{m;\,i,j} := \mathsf{A}_m(\tilde{a}_i, \varepsilon_i)^{-1}\mathsf{B}_m(\tilde{a}_i, \varepsilon_i, \tilde{R})\mathsf{A}_m(\tilde{a}_j, \varepsilon_j)^{-1}\mathsf{B}_m(\tilde{a}_j, \varepsilon_j, \tilde{R}),$$

where $\alpha_n(\tilde{a}_i, \varepsilon_i)$, $b_{nml}(\tilde{a}_i, \tilde{R})$ and $\beta_{nml}(\tilde{a}_i, \varepsilon_i, \tilde{R})$ are given by

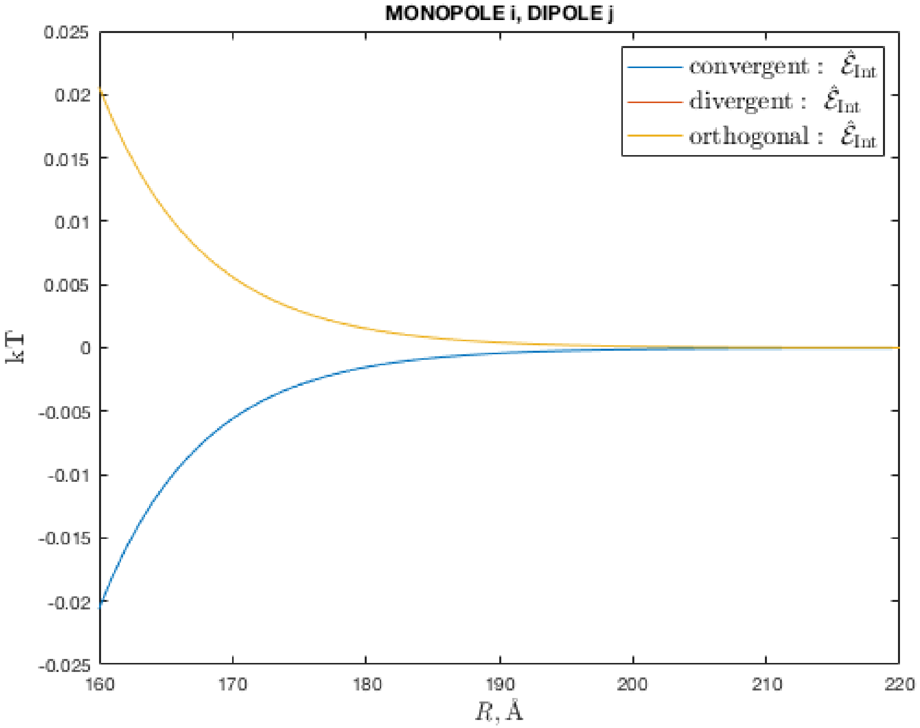


Figure 4: Electrostatic interaction energy $\hat{\mathcal{E}}_{\mathrm{Int}}$ in the Monopole-Dipole system case, as defined in eq. (5.5.1). The values of the energy are given as a function of the distance between the centres of the spheres for different relative dipole orientations: convergent (equivalent to divergent in this case), and orthogonal.

$$\begin{aligned}
b_{nml}(\tilde{a}_i,\tilde{R}) &:= (2l-1)!!\sum_{s=0}^{+\infty}\sum_{k=m}^{l}\frac{(-1)^{k+m}(l+m)!}{(k+m)!(l-k)!}\frac{\Omega_{l+s+1/2}(\tilde{a}_i,\tilde{R})}{\tilde{a}_i^{l-k}\tilde{R}^k}h_{kmls,n},\\
\alpha_n(\tilde{a}_i,\varepsilon_i) &:= \varepsilon_{\mathrm{m}}\frac{K_{n+3/2}(\tilde{a}_i)}{\tilde{a}_i^{1/2}} + n(\varepsilon_i-\varepsilon_{\mathrm{m}})\frac{K_{n+1/2}(\tilde{a}_i)}{\tilde{a}_i^{3/2}},\\
\beta_{nml}(\tilde{a}_i,\varepsilon_i,\tilde{R}) &:= \frac{\varepsilon_i n}{\tilde{a}_i}b_{nml}(\tilde{a}_i,\tilde{R}) - \varepsilon_{\mathrm{m}}\frac{\partial b_{nml}(\tilde{a}_i,\tilde{R})}{\partial\tilde{a}_i},\\
\Omega_{n+1/2}(\tilde{a}_i,\tilde{R}) &:= \sqrt{2\pi}\left(n+\frac{1}{2}\right)\frac{K_{n+1/2}(\tilde{R})I_{n+1/2}(\tilde{a}_i)}{\sqrt{\tilde{R}\tilde{a}_i}},\\
h_{kmls,n} &:= \frac{(2n+1)(n-m)!}{2(n+m)!}\int_{-1}^{1}C_s^{l+1/2}(x)P_n^m(x)P_k^m(x)dx,
\end{aligned}$$

with $C_s^{l+1/2}(\cdot)$ being the standard Gegenbauer ultraspherical polynomials.

## Charge – Charge System

In this case, denoting the corresponding potential expansion coefficients (1) as

$$\mathbf{G}_{m,i} := \mathbf{G}^{\mathrm{mon}}_{m,i}, \qquad \mathbf{H}_{m,i} := \mathbf{H}^{\mathrm{mon}}_{m,i}, \qquad \mathbf{L}_{m,i} := \mathbf{L}^{\mathrm{mon}}_{m,i}, \qquad \mathbf{M}_{m,i} := \mathbf{M}^{\mathrm{mon}}_{m,i}, \tag{2}$$

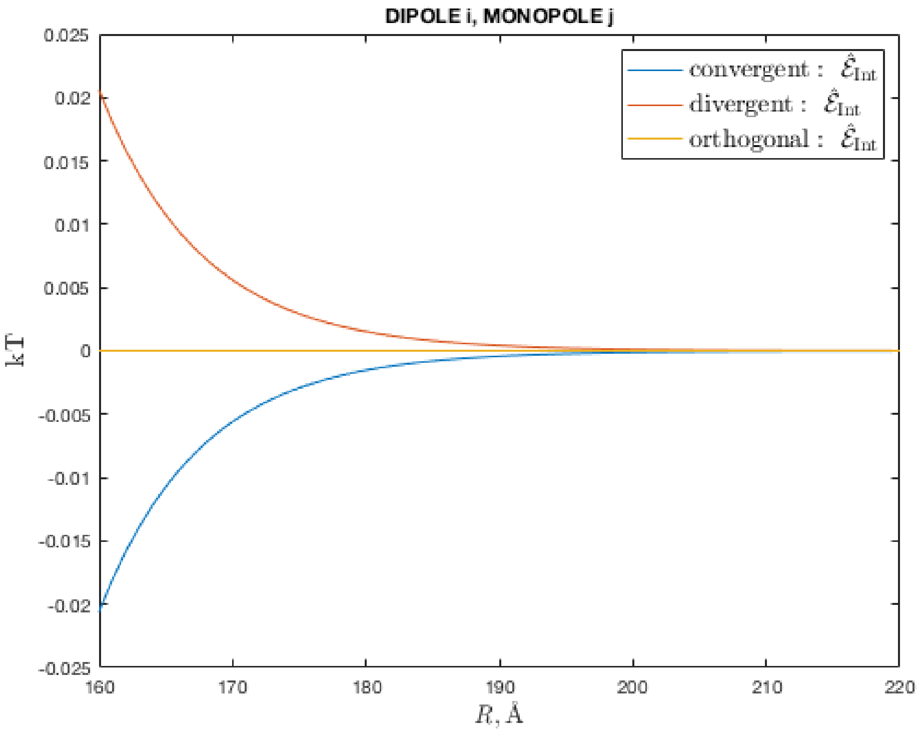


Figure 5: Electrostatic interaction energy $\hat{\mathcal{E}}_{\mathrm{Int}}$ in Dipole-Monopole system case, as defined in eq. (5.5.1). The values of the energy are given as a function of the distance between the centres of the spheres for different relative dipole orientations: convergent, divergent, and orthogonal.

one derives the following matrix power series:

$$\begin{aligned}
\mathbf{G}_{m,i} &= \sum_{s=0}^{+\infty} \left(\mathsf{N}_{m;\,i,j}\right)^{s} \mathfrak{F}_{m,i;G}, && m \geq 0,\\
\mathbf{H}_{m,i} &= \sum_{s=0}^{+\infty} \left(\mathsf{N}_{m;\,i,j}\right)^{s} \mathfrak{F}_{m,i;H}, && m \geq 1,\\
\mathbf{L}_{m,i} &= \mathfrak{C}_{m,i} + \sum_{s=0}^{+\infty}\Big(\mathsf{K}_m(\tilde{a}_i)\left(\mathsf{N}_{m;\,i,j}\right)^{s} \mathfrak{F}_{m,i;G} + \overset{\lceil}{\mathsf{B}}_m(\tilde{a}_i,\tilde{R})\left(\mathsf{N}_{m;\,j,i}\right)^{s} \mathfrak{F}_{m,j;G}\Big), && m \geq 0,\\
\mathbf{M}_{m,i} &= \mathfrak{D}_{m,i} + \sum_{s=0}^{+\infty}\Big(\mathsf{K}_m(\tilde{a}_i)\left(\mathsf{N}_{m;\,i,j}\right)^{s} \mathfrak{F}_{m,i;H} + \overset{\lceil}{\mathsf{B}}_m(\tilde{a}_i,\tilde{R})\left(\mathsf{N}_{m;\,j,i}\right)^{s} \mathfrak{F}_{m,j;H}\Big), && m \geq 1,
\end{aligned} \tag{3}$$

where the column-vectors $\mathfrak{F}_{m,i;G}$, $\mathfrak{F}_{m,i;H}$ are

$$\begin{aligned}
\mathfrak{F}_{m,i;G} &:= \mathsf{A}_m(\tilde{a}_i,\varepsilon_i)^{-1}\left(\mathbf{F}_{m,i;G} - \mathsf{B}_m(\tilde{a}_i,\varepsilon_i,\tilde{R})\mathsf{A}_m(\tilde{a}_j,\varepsilon_j)^{-1}\mathbf{F}_{m,j;G}\right),\\
\mathfrak{F}_{m,i;H} &:= \mathsf{A}_m(\tilde{a}_i,\varepsilon_i)^{-1}\left(\mathbf{F}_{m,i;H} - \mathsf{B}_m(\tilde{a}_i,\varepsilon_i,\tilde{R})\mathsf{A}_m(\tilde{a}_j,\varepsilon_j)^{-1}\mathbf{F}_{m,j;H}\right),
\end{aligned} \tag{4}$$

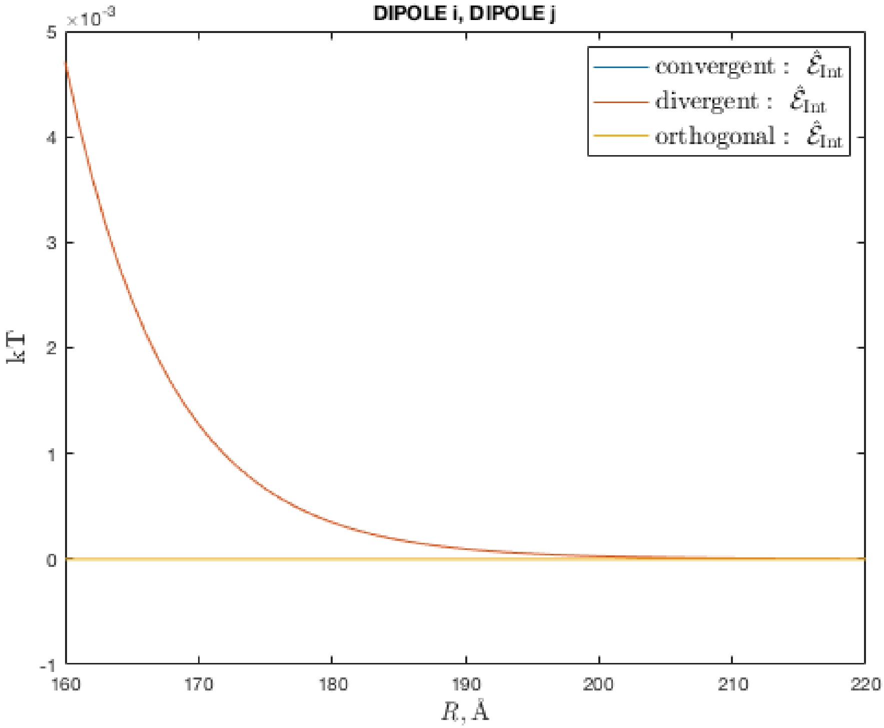


Figure 6: Electrostatic interaction energy $\hat{\mathcal{E}}_{\mathrm{Int}}$ in the Dipole-Dipole system case, as defined in eq. (5.5.1). The values of the energy are given as a function of the distance between the centres of the spheres for different relative dipole orientations: convergent (equivalent to divergent in this case), and orthogonal.

while, denoting a Kronecker delta by $\delta$, the column-vectors $\mathbf{F}_{m,i;G}$, $\mathbf{F}_{m,i;H}$, $\mathfrak{C}_{m,i}$, $\mathfrak{D}_{m,i}$ are, element-wisely,

$$\begin{aligned}
\mathbf{F}_{m,i;G} &= \mathbf{F}^{\mathrm{mon}}_{m,i;G} \;:= \left\{ \left( \frac{q_i\kappa}{4\pi\varepsilon_0\tilde{a}_i^2} - \frac{q_i\kappa e^{-\tilde{a}_i}}{4\pi\varepsilon_0}\left(\frac{1}{\tilde{a}_i}+\frac{1}{\tilde{a}_i^2}\right)\right)\delta_{n,0} - \frac{\sqrt{2}\kappa q_j}{4\pi^{3/2}\varepsilon_0\varepsilon_{\mathrm{m}}}\beta_{n00}(\tilde{a}_i,\varepsilon_i,\tilde{R})\delta_{m,0} \right\}_{n\geq m}, \\
\mathbf{F}_{m,i;H} &= \mathbf{F}^{\mathrm{mon}}_{m,i;H} \;:= \{0\}_{n\geq m}, \\
\mathfrak{C}_{m,i} &= \mathfrak{C}^{\mathrm{mon}}_{m,i} \;:= \left\{ \frac{1}{\tilde{a}_i^n}\left( \left( \frac{q_i\kappa e^{-\tilde{a}_i}}{4\pi\varepsilon_0\varepsilon_{\mathrm{m}}\tilde{a}_i} - \frac{q_i\kappa}{4\pi\varepsilon_0\varepsilon_i\tilde{a}_i}\right)\delta_{n,0} + b_{n00}(\tilde{a}_i,\tilde{R})\frac{\sqrt{2}\kappa q_j}{4\pi^{3/2}\varepsilon_0\varepsilon_{\mathrm{m}}}\delta_{m,0}\right)\right\}_{n\geq m}, \\
\mathfrak{D}_{m,i} &= \mathfrak{D}^{\mathrm{mon}}_{m,i} \;:= \{0\}_{n\geq m}\,.
\end{aligned} \tag{5}$$

In particular, it is easy to assert from relations (5) and (3), (4), that, in fact, all $G_{nm,i} = H_{nm,i} = L_{nm,i} = M_{nm,i} = 0$ for $n \geq m \geq 1$ which implicates the evident absence of any tesseral/sectoral harmonics in the azimuthally symmetric charge-charge system under consideration, but nonetheless we present these relations in a general form in order to facilitate further generalisations for dipole-dipole and charge-dipole systems considered below.

## Dipole – Dipole System

In this case, corresponding potential coefficient vectors

$$\mathbf{G}_{m,i} := \mathbf{G}_{m,i}^{\mathrm{dip}}, \qquad \mathbf{H}_{m,i} := \mathbf{H}_{m,i}^{\mathrm{dip}}, \qquad \mathbf{L}_{m,i} := \mathbf{L}_{m,i}^{\mathrm{dip}}, \qquad \mathbf{M}_{m,i} := \mathbf{M}_{m,i}^{\mathrm{dip}} \tag{6}$$

are still given by relations (3), (4), (5), but vectors $\mathbf{F}_{m,i;G}^{\mathrm{mon}}$, $\mathbf{F}_{m,i;H}^{\mathrm{mon}}$, $\mathfrak{C}_{m,i}^{\mathrm{mon}}$, $\mathfrak{D}_{m,i}^{\mathrm{mon}}$ in (5) should now be replaced by the vectors $\mathbf{F}_{m,i;G}^{\mathrm{dip}}$, $\mathbf{F}_{m,i;H}^{\mathrm{dip}}$, $\mathfrak{C}_{m,i}^{\mathrm{dip}}$, $\mathfrak{D}_{m,i}^{\mathrm{dip}}$, respectively, defined elementwisely as follows:

$$\begin{aligned}
\mathbf{F}_{m,i;G}^{\mathrm{dip}} &:= \left\{ e_{nm,i}^{\mathrm{dip}} - \frac{\varepsilon_i n}{\tilde{a}_i} c_{nm,i}^{\mathrm{dip}} + \varepsilon_{\mathrm{m}} \frac{\partial A_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R})}{\partial \tilde{a}_i} - \frac{\varepsilon_i n}{\tilde{a}_i} A_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R}) \right\}_{n \geq m}, \\
\mathbf{F}_{m,i;H}^{\mathrm{dip}} &:= \left\{ \left( \frac{\varepsilon_i n}{\tilde{a}_i} \hat{c}_i^{\mathrm{dip}} - \hat{e}_i^{\mathrm{dip}} \right) p_i \delta_{n,1} \sin \bar{\theta}_i \sin \bar{\varphi}_i + \varepsilon_{\mathrm{m}} \frac{\partial B_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R})}{\partial \tilde{a}_i} - \frac{\varepsilon_i n}{\tilde{a}_i} B_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R}) \right\}_{n \geq m}, \\
\mathfrak{C}_{m,i}^{\mathrm{dip}} &:= \left\{ \frac{c_{nm,i}^{\mathrm{dip}} + A_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R})}{\tilde{a}_i^n} \right\}_{n \geq m}, \\
\mathfrak{D}_{m,i}^{\mathrm{dip}} &:= \left\{ \frac{-\hat{c}_i^{\mathrm{dip}} p_i \delta_{n,1} \sin \bar{\theta}_i \sin \bar{\varphi}_i + B_{nm,j}^{\mathrm{dip}}(\tilde{a}_i, \tilde{R})}{\tilde{a}_i^n} \right\}_{n \geq m},
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
c_{nm,i}^{\mathrm{dip}} &:= \hat{c}_i^{\mathrm{dip}} p_i \delta_{n,1} \cdot \left\{ \begin{array}{ll} \cos \bar{\theta}_i, & m = 0 \\ -\sin \bar{\theta}_i \cos \bar{\varphi}_i, & m = 1 \end{array} \right\}, \quad & \hat{c}_i^{\mathrm{dip}} &:= \frac{\kappa^2 e^{-\tilde{a}_i}}{4\pi \varepsilon_0 \varepsilon_{\mathrm{m}}} \left( \frac{1}{\tilde{a}_i} + \frac{1}{\tilde{a}_i^2} \right) - \frac{\kappa^2}{4\pi \varepsilon_0 \varepsilon_i \tilde{a}_i^2}, \\
e_{nm,i}^{\mathrm{dip}} &:= \hat{e}_i^{\mathrm{dip}} p_i \delta_{n,1} \cdot \left\{ \begin{array}{ll} \cos \bar{\theta}_i, & m = 0 \\ -\sin \bar{\theta}_i \cos \bar{\varphi}_i, & m = 1 \end{array} \right\}, \quad & \hat{e}_i^{\mathrm{dip}} &:= \frac{\kappa^2}{2\pi \varepsilon_0 \tilde{a}_i^3} - \frac{\kappa^2 e^{-\tilde{a}_i}}{4\pi \varepsilon_0} \left( \frac{2}{\tilde{a}_i^3} + \frac{2}{\tilde{a}_i^2} + \frac{1}{\tilde{a}_i} \right),
\end{aligned} \tag{8}$$

and

$$\begin{aligned}
A_{n0,j}^{\mathrm{dip}}(\tilde{r}_i, \tilde{R}) &:= b_{n01}(\tilde{r}_i, \tilde{R}) \frac{\sqrt{2}\kappa^2}{4\pi \sqrt{\pi} \varepsilon_0 \varepsilon_{\mathrm{m}}} p_j \cos \bar{\theta}_j, \ n \geq 0, \\
A_{n1,j}^{\mathrm{dip}}(\tilde{r}_i, \tilde{R}) &:= -b_{n11}(\tilde{r}_i, \tilde{R}) \frac{\sqrt{2}\kappa^2}{4\pi \sqrt{\pi} \varepsilon_0 \varepsilon_{\mathrm{m}}} p_j \sin \bar{\theta}_j \cos \bar{\varphi}_j, \ n \geq 1, \\
B_{n1,j}^{\mathrm{dip}}(\tilde{r}_i, \tilde{R}) &:= -b_{n11}(\tilde{r}_i, \tilde{R}) \frac{\sqrt{2}\kappa^2}{4\pi \sqrt{\pi} \varepsilon_0 \varepsilon_{\mathrm{m}}} p_j \sin \bar{\theta}_j \sin \bar{\varphi}_j, \ n \geq 1, \\
A_{nm,j}^{\mathrm{dip}}(\tilde{r}_i, \tilde{R}) &:= 0 \quad \text{and} \quad B_{nm,j}^{\mathrm{dip}}(\tilde{r}_i, \tilde{R}) := 0 \quad \text{for} \quad \forall m \geq 2.
\end{aligned} \tag{9}$$

In particular, using the above definitions in (7), it is easy to assert that $G_{nm,i} = H_{nm,i} = L_{nm,i} = M_{nm,i} = 0$ when $n \geq m \geq 2$.

## Charge – Dipole System

Let us now consider the case of a mixed system consisting of a monopole and a dipole. Without loss of generality, we assume that the $i$-th particle contains the centred point charge (monopole) $q_i$, whereas the opposite $j$-th particle contains the centred point dipole with dipole moment $\mathbf{p}_j$. Placing a dipole of, say, dipole moment $p_i$ in the $i$-th sphere and a charge (say, $q_j$) in the opposite $j$-th sphere amounts to simply swapping the indices $i$ and $j$ in all the expressions (11) below.

As in the previously considered (charge-charge and dipole-dipole) cases, the expansion coefficients for potentials

$$\begin{aligned} \mathbf{G}_{m,i} &:= \mathbf{G}_{m,i}^{\text{mix,mon}}, & \mathbf{H}_{m,i} &:= \mathbf{H}_{m,i}^{\text{mix,mon}}, & \mathbf{L}_{m,i} &:= \mathbf{L}_{m,i}^{\text{mix,mon}}, & \mathbf{M}_{m,i} &:= \mathbf{M}_{m,i}^{\text{mix,mon}}, \\ \mathbf{G}_{m,j} &:= \mathbf{G}_{m,j}^{\text{mix,dip}}, & \mathbf{H}_{m,j} &:= \mathbf{H}_{m,j}^{\text{mix,dip}}, & \mathbf{L}_{m,j} &:= \mathbf{L}_{m,j}^{\text{mix,dip}}, & \mathbf{M}_{m,j} &:= \mathbf{M}_{m,j}^{\text{mix,dip}} \end{aligned} \tag{10}$$

are still given by (3), (4), (5), where $\mathbf{F}_{m,i;G}^{\text{mon}}$, $\mathbf{F}_{m,j;G}^{\text{mon}}$, $\mathfrak{C}_{nm,i}^{\text{mon}}$ and $\mathfrak{C}_{nm,j}^{\text{mon}}$ should now be replaced by $\mathbf{F}_{m,i;G}^{\text{mix,mon}}$, $\mathbf{F}_{m,j;G}^{\text{mix,dip}}$, $\mathfrak{C}_{nm,i}^{\text{mix,mon}}$ and $\mathfrak{C}_{nm,j}^{\text{mix,dip}}$, respectively, whereas $\mathbf{F}_{m,i;H}^{\text{mon}}$, $\mathbf{F}_{m,j;H}^{\text{mon}}$, $\mathfrak{D}_{nm,i}^{\text{mon}}$ and $\mathfrak{D}_{nm,j}^{\text{mon}}$ should be replaced by $\mathbf{F}_{m,i;H}^{\text{mix,mon}}$, $\mathbf{F}_{m,j;H}^{\text{mix,dip}}$, $\mathfrak{D}_{nm,i}^{\text{mix,mon}}$ and $\mathfrak{D}_{nm,j}^{\text{mix,dip}}$, respectively. These coefficients are defined element-wisely as follows:

$$\begin{aligned}
\mathbf{F}_{m,i;G}^{\text{mix,mon}} &:= \left\{ \left( \frac{q_i \kappa}{4\pi\varepsilon_0 \tilde{a}_i^2} - \frac{q_i \kappa e^{-\tilde{a}_i}}{4\pi\varepsilon_0} \left( \frac{1}{\tilde{a}_i} + \frac{1}{\tilde{a}_i^2} \right) \right) \delta_{n,0} + \varepsilon_{\mathrm{m}} \frac{\partial A_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R})}{\partial \tilde{a}_i} - \frac{\varepsilon_i n}{\tilde{a}_i} A_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R}) \right\}_{n \geq m}, \\
\mathbf{F}_{m,j;G}^{\text{mix,dip}} &:= \left\{ e_{nm,j}^{\text{dip}} - \frac{\varepsilon_j n}{\tilde{a}_j} c_{nm,j}^{\text{dip}} - q_i \frac{\sqrt{2}\kappa}{4\pi^{3/2}\varepsilon_0 \varepsilon_{\mathrm{m}}} \beta_{n00}(\tilde{a}_j, \varepsilon_j, \tilde{R}) \delta_{m,0} \right\}_{n \geq m}, \\
\mathbf{F}_{m,i;H}^{\text{mix,mon}} &:= \left\{ \varepsilon_{\mathrm{m}} \frac{\partial B_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R})}{\partial \tilde{a}_i} - \frac{\varepsilon_i n}{\tilde{a}_i} B_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R}) \right\}_{n \geq m}, \\
\mathbf{F}_{m,j;H}^{\text{mix,dip}} &:= \left\{ \left( \frac{\varepsilon_j n}{\tilde{a}_j} \hat{c}_j^{\text{dip}} - \hat{e}_j^{\text{dip}} \right) p_j \delta_{n,1} \sin \bar{\theta}_j \sin \bar{\varphi}_j \right\}_{n \geq m}, \\
\mathfrak{C}_{nm,i}^{\text{mix,mon}} &:= \left\{ \frac{1}{\tilde{a}_i^n} \left( \left( \frac{q_i \kappa e^{-\tilde{a}_i}}{4\pi\varepsilon_0 \varepsilon_{\mathrm{m}} \tilde{a}_i} - \frac{q_i \kappa}{4\pi\varepsilon_0 \varepsilon_i \tilde{a}_i} \right) \delta_{n,0} + A_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R}) \right) \right\}_{n \geq m}, \\
\mathfrak{C}_{nm,j}^{\text{mix,dip}} &:= \left\{ \frac{1}{\tilde{a}_j^n} \left( c_{nm,j}^{\text{dip}} + b_{n00}(\tilde{a}_j, \tilde{R}) \frac{\sqrt{2}\kappa q_i}{4\pi^{3/2}\varepsilon_0 \varepsilon_{\mathrm{m}}} \delta_{m,0} \right) \right\}_{n \geq m}, \\
\mathfrak{D}_{nm,i}^{\text{mix,mon}} &:= \left\{ \frac{1}{\tilde{a}_i^n} B_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R}) \right\}_{n \geq m}, \\
\mathfrak{D}_{nm,j}^{\text{mix,dip}} &:= \left\{ -\frac{1}{\tilde{a}_j^n} \hat{c}_j^{\text{dip}} p_j \delta_{n,1} \sin \bar{\theta}_j \sin \bar{\varphi}_j \right\}_{n \geq m},
\end{aligned} \tag{11}$$

where $c_{nm,j}^{\text{dip}}$, $\hat{c}_j^{\text{dip}}$, $e_{nm,j}^{\text{dip}}$, $\hat{e}_j^{\text{dip}}$ and $A_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R})$, $B_{nm,j}^{\text{dip}}(\tilde{a}_i, \tilde{R})$ are determined by (8) and (9), respectively.

## Charge-Dipole – Charge-Dipole System (General Case)

Finally, let us assume that both spheres $i = \overline{1,2}$ and $j = 3 - i$ carry point charges $q_i$, $q_j$, and point dipoles $\mathbf{p}_i$, $\mathbf{p}_j$, located at the spheres' centres. Then, for arbitrary $i \in \{1, 2\}$ one eventually gets the following superposition of (2), (6), (10) for the vectors $\mathbf{G}_{m,i} := \mathbf{G}_{m,i}^{\text{general}}$,

$\mathbf{H}_{m,i} := \mathbf{H}^{\text{general}}_{m,i}$, $\mathbf{L}_{m,i} := \mathbf{L}^{\text{general}}_{m,i}$, $\mathbf{M}_{m,i} := \mathbf{M}^{\text{general}}_{m,i}$:

$$\begin{aligned}
\mathbf{G}^{\text{general}}_{m,i} &= \frac{1}{2}\left(\mathbf{G}^{\text{mon}}_{m,i} + \mathbf{G}^{\text{mix,mon}}_{m,i} + \mathbf{G}^{\text{mix,dip}}_{m,i} + \mathbf{G}^{\text{dip}}_{m,i}\right),\\
\mathbf{H}^{\text{general}}_{m,i} &= \frac{1}{2}\left(\mathbf{H}^{\text{mon}}_{m,i} + \mathbf{H}^{\text{mix,mon}}_{m,i} + \mathbf{H}^{\text{mix,dip}}_{m,i} + \mathbf{H}^{\text{dip}}_{m,i}\right),\\
\mathbf{L}^{\text{general}}_{m,i} &= \frac{1}{2}\left(\mathbf{L}^{\text{mon}}_{m,i} + \mathbf{L}^{\text{mix,mon}}_{m,i} + \mathbf{L}^{\text{mix,dip}}_{m,i} + \mathbf{L}^{\text{dip}}_{m,i}\right),\\
\mathbf{M}^{\text{general}}_{m,i} &= \frac{1}{2}\left(\mathbf{M}^{\text{mon}}_{m,i} + \mathbf{M}^{\text{mix,mon}}_{m,i} + \mathbf{M}^{\text{mix,dip}}_{m,i} + \mathbf{M}^{\text{dip}}_{m,i}\right),
\end{aligned} \tag{12}$$

where from the electrostatic point of view the superscripts "mon", "mix,mon", "mix,dip", "dip" correspond to the interaction of $q_i$ with $q_j$, $q_i$ with $\mathbf{p}_j$, $\mathbf{p}_i$ with $q_j$, and $\mathbf{p}_i$ with $\mathbf{p}_j$, respectively, as described in the corresponding situations above.

In the case of parallel/antiparallel dipole moments $\mathbf{p}_i$ (where $i = \overline{1,2}$) one has $\sin\bar{\theta}_i = 0$, so that the system becomes azimuthally symmetric and only the coefficients with $m = 0$ contribute. In particular, using the exact analytical representations

$$K_{n+1/2}(x) = \sqrt{\frac{\pi}{2x}}\, e^{-x} \sum_{l=0}^{n} \frac{(n+l)!}{l!(n-l)!(2x)^l},$$

$$I_{n+1/2}(x) = \frac{1}{\sqrt{2\pi x}} \left( e^{x} \sum_{l=0}^{n} \frac{(-1)^l (n+l)!}{l!(n-l)!(2x)^l} + (-1)^{n+1} e^{-x} \sum_{l=0}^{n} \frac{(n+l)!}{l!(n-l)!(2x)^l} \right),$$

and asymptotic relations

$$\frac{K_{n+1/2}(x)}{\sqrt{x}} \sim \sqrt{\frac{\pi}{2}}\frac{e^{-x}}{x} \quad (\text{as } x \longrightarrow \infty)$$

$$\frac{I_{n+1/2}(x)}{\sqrt{x}} \sim \sqrt{\frac{2}{\pi}}\frac{x^n}{(2n+1)!!} \quad \frac{K_{n+1/2}(x)}{\sqrt{x}} \sim \sqrt{\frac{\pi}{2}}\frac{(2n-1)!!}{x^{n+1}} \quad (\text{as } x \longrightarrow 0)$$

for modified Bessel functions $K_{n+1/2}(\cdot)$ and $I_{n+1/2}(\cdot)$ of semi-integer order, after extensive algebraic manipulations one gets the following coefficients of the external potential $\tilde{\Phi}_{\text{out},i}$ resulting from (12):

$$\begin{aligned}
G^{\text{general}}_{00,i} &= \frac{q_i \kappa \sqrt{2}}{4\pi^{3/2}\varepsilon_0\varepsilon_{\text{m}}} \left( \frac{e^{\tilde{a}_i}}{1+\tilde{a}_i} - 1 \right) + \frac{\sqrt{2} e^{\tilde{a}_j - \tilde{R}}}{8\pi^{3/2}\varepsilon_0 \tilde{R}} \left( \frac{q_j \kappa}{(1+\tilde{a}_j)\varepsilon_{\text{m}}} + \frac{3\kappa^2 p_j \cos\bar{\theta}_j}{(\varepsilon_j + 2\varepsilon_{\text{m}})(1+\tilde{a}_j) + \varepsilon_{\text{m}}\tilde{a}_j^2} \left(1 + \frac{1}{\tilde{R}}\right)\right)\\
&\quad \times \left( \frac{e^{2\tilde{a}_i}(\tilde{a}_i - 1)}{\tilde{a}_i + 1} + 1 \right) + O(\omega),\\
G^{\text{general}}_{10,i} &= \frac{\sqrt{2} p_i \kappa^2 \cos\bar{\theta}_i}{4\pi^{3/2}\varepsilon_0} \left( \frac{3e^{\tilde{a}_i}}{(\varepsilon_i + 2\varepsilon_{\text{m}})(1+\tilde{a}_i) + \varepsilon_{\text{m}}\tilde{a}_i^2} - \frac{1}{\varepsilon_{\text{m}}} \right) - \frac{3\sqrt{2} e^{\tilde{a}_j - \tilde{R}}}{8\pi^{3/2}\varepsilon_0 \tilde{R}} \left( e^{2\tilde{a}_i} \frac{(\varepsilon_i + 2\varepsilon_{\text{m}})(\tilde{a}_i - 1) - \varepsilon_{\text{m}}\tilde{a}_i^2}{(\varepsilon_i + 2\varepsilon_{\text{m}})(1+\tilde{a}_i) + \varepsilon_{\text{m}}\tilde{a}_i^2} + 1 \right)\\
&\quad \times \left( \frac{q_j \kappa}{(1+\tilde{a}_j)\varepsilon_{\text{m}}} \left(1 + \frac{1}{\tilde{R}}\right) + \frac{3\kappa^2 p_j \cos\bar{\theta}_j}{(\varepsilon_j + 2\varepsilon_{\text{m}})(1+\tilde{a}_j) + \varepsilon_{\text{m}}\tilde{a}_j^2} \left(1 + \frac{2}{\tilde{R}} + \frac{2}{\tilde{R}^2}\right)\right) + O(\omega),
\end{aligned} \tag{13}$$

where the dimensionless doubly-screened factor $\omega := \frac{a_i a_j e^{-2\kappa R}}{R^2}$. Coefficients (13) were used to construct the empirical model for predicting the external potential outside the "nonlinearity corona". Furthermore, numerical simulations show that increasing the number of terms beyond $G^{\text{general}}_{00,i}$ and $G^{\text{general}}_{10,i}$, i.e., using $G^{\text{general}}_{20,i}$ and higher-order series terms in eq.(5.12) , has very little effect on the final prediction result.

## Nonlinear PBE and Charge Renormalisation

### Nonlinear PBE for a Single Sphere

Let us consider a single sphere of radius $a$ and dielectric constant $\varepsilon$, immersed into a polarisable ionic medium described by a dielectric constant $\varepsilon_{\mathrm{m}}$ and a Debye length $\lambda := \kappa^{-1}$. It is assumed that the sphere contains a point charge $q$ located in its centre. The corresponding normalised (i.e. dimensionless by dividing by $kT/e$) electrostatic potential $\varphi$ **inside** the sphere is then given by

$$\varphi = \frac{e}{kT}\frac{q}{4\pi\varepsilon_0\varepsilon r} = \frac{e^2}{4\pi\varepsilon_0 kT\text{Å}}\frac{q_e}{\varepsilon r_{\text{Å}}} = \frac{q_e}{\tilde{\varepsilon} r_{\text{Å}}}, \tag{14}$$

where we normalised the charge $q$ with respect to the elementary charge $e$ and the radial position $r$ to portions of Å, so that $q_e := q/e$, $r_{\text{Å}} := r/\text{Å}$, and $\tilde{\varepsilon} := \varepsilon\tilde{\varepsilon}_0$, where the dimensionless parameter $\tilde{\varepsilon}_0 := 4\pi\varepsilon_0 kT\text{Å}/e^2$ ($\tilde{\varepsilon}_0^{-1} \approx 560.74$ at the temperature of 25° Celsius).

The corresponding spherically symmetric self-consistent normalised potential $\varphi(r_{\text{Å}})$ **outside** the sphere (i.e., $r_{\text{Å}} > a_{\text{Å}} := a/\text{Å}$) is governed by the nonlinear PBE $\Delta_{\text{Å}}\varphi(r_{\text{Å}}) - \lambda_{\text{Å}}^{-2}\sinh\varphi(r_{\text{Å}}) = 0$. We apply the following boundary condition at the spherical surface:

$$\varepsilon\ \varphi'\big|_{r=a-0} = \varepsilon_{\mathrm{m}}\ \varphi'\big|_{r=a+0}\,. \tag{15}$$

Employing the spherical symmetry in the notation of the Laplace operator and substituting (14) into (15), one gets the following nonlinear boundary-value problem on the infinite interval $r_{\text{Å}} \geq a_{\text{Å}}$:

$$\begin{cases} \dfrac{d^2\varphi(r_{\text{Å}})}{dr_{\text{Å}}^2} + \dfrac{2}{r_{\text{Å}}}\dfrac{d\varphi(r_{\text{Å}})}{dr_{\text{Å}}} - \dfrac{\sinh\varphi(r_{\text{Å}})}{\lambda_{\text{Å}}^2} = 0, & r_{\text{Å}} > a_{\text{Å}}, \\ \left.\dfrac{d\varphi(r_{\text{Å}})}{dr_{\text{Å}}}\right|_{r_{\text{Å}}=a_{\text{Å}}} = -\dfrac{q_e}{\tilde{\varepsilon}_0\varepsilon_{\mathrm{m}} a_{\text{Å}}^2}, & \varphi(r_{\text{Å}})\big|_{r_{\text{Å}}\to+\infty} = 0. \end{cases} \tag{16}$$

When evaluating all distances as multiples of $\lambda$ (thus, the new dimensionless radial coordinate is $r_\lambda := r/\lambda$) one can also recast (16) to the following form:

$$\begin{cases} \dfrac{d^2\varphi(r_\lambda)}{dr_\lambda^2} + \dfrac{2}{r_\lambda}\dfrac{d\varphi(r_\lambda)}{dr_\lambda} - \sinh\varphi(r_\lambda) = 0, & r_\lambda > a_\lambda, \\ \left.\dfrac{d\varphi(r_\lambda)}{dr_\lambda}\right|_{r_\lambda=a_\lambda} = -\dfrac{q_e}{\tilde{\varepsilon}_0\varepsilon_{\mathrm{m}}\lambda_{\text{Å}} a_\lambda^2}, & \varphi(r_\lambda)\big|_{r_\lambda\to+\infty} = 0. \end{cases} \tag{17}$$

To numerically solve (17) or its equivalent (16) one must first transform the infinite interval $r_\lambda \geq a_\lambda$ into a finite one, say, $[0,1]$. This can be done using the mappings $t = t(r_\lambda) = 1 - e^{-(r_\lambda - a_\lambda)}$ or $t = t(r_\lambda) = 2\arctan(r_\lambda - a_\lambda)/\pi$ to a new auxiliary variable $t \in [0,1]$.

In particular, using the latter mapping and expressing derivatives through the chain rule (e.g., for $\frac{d\varphi}{dr_\lambda}$ one has $\frac{d\varphi}{dr_\lambda} = \frac{d\varphi}{dt} \cdot \frac{dt}{dr_\lambda} = \frac{d\varphi}{dt} \cdot \frac{2}{\pi} \frac{1}{1+(r_\lambda - a_\lambda)^2} = \frac{d\varphi}{dt} \cdot \frac{2}{\pi} \frac{1}{1+\tan^2(\pi t/2)} = \frac{d\varphi}{dt} \cdot \frac{2}{\pi} \cos^2 \frac{\pi t}{2}$, etc.), the problem (17) then reduces to the following one:

$$\begin{cases} \left(\frac{4}{\pi^2} \cos^4 \frac{\pi t}{2}\right) \frac{d^2\varphi(t)}{dt} + \left(\frac{2}{a_\lambda + \tan(\pi t/2)} - \sin \pi t\right) \left(\frac{2}{\pi} \cos^2 \frac{\pi t}{2}\right) \frac{d\varphi(t)}{dt} - \sinh \varphi(t) = 0, \quad t \in (0,1), \\ \left.\frac{d\varphi(t)}{dt}\right|_{t=0} = -\frac{\pi}{2} \frac{q_e}{\tilde{\varepsilon}_0 \varepsilon_{\mathrm{m}} \lambda_{\mathrm{\AA}} a_\lambda^2}, \qquad \varphi(t)|_{t=1} = 0. \end{cases} \tag{18}$$

Now the boundary-value problem (18) can be handled numerically. The numerical calculations performed in `Maple 2019` are presented below; the built-in midpoint solvers `middefer` and `midrich` were used (note that (18) has a singular point at $t = 1$, thus the conventional trapezoidal solvers are impracticable). The absolute integration error level was $10^{-7}$.

## The Nonlinearity Regime for a Single Sphere

Let us consider the distance $b > a$ such that $|\varphi(b)| = 10^{-1}$, which we will use as a criterion for identifying the transition to the linear regime. Note that the monotonic decrease of the potential magnitude $|\varphi|$ as $r \longrightarrow +\infty$, see [1], which indicates that, after such a "nonlinearity threshold" $r = b$, the potential magnitude will always be decaying and its module will be less than $10^{-1}$. Indeed, let us point out that $|\sinh x - x| < 1.67 \cdot 10^{-4}$ and $|\sinh x - x| < 1.67 \cdot 10^{-3} |\sinh x|$ for arbitrary $|x| \leq 10^{-1}$, thus one has the approximation $\sinh \varphi \approx \varphi$ with the relative error less than 0.17% as $|\varphi| \leq 10^{-1}$. This permits the use of the linearised PBE

$$\frac{d^2\varphi(r_\lambda)}{dr_\lambda^2} + \frac{2}{r_\lambda} \frac{d\varphi(r_\lambda)}{dr_\lambda} - \varphi(r_\lambda) = 0 \tag{19}$$

in the "linearity" domain $r_\lambda \geq b_\lambda > a_\lambda$ instead of its nonlinear counterpart (17).

The charge density in the solvent is $\rho_{\mathrm{solvent}} \mathrm{\AA}^3/e = -\tilde{\varepsilon}_0 \varepsilon_{\mathrm{m}} \sinh(\varphi(\mathbf{r}_{\mathrm{\AA}}))/4\pi\lambda_{\mathrm{\AA}}^2$. Then, the normalized total charge (of counterions) $Q_{NL}$ in the nonlinearity layer is given by

$$\frac{Q_{NL}}{e} = \int\limits_{\text{nonlinearity layer}} \frac{\rho_{\mathrm{solvent}} \mathrm{\AA}^3}{e} d\mathbf{r}_{\mathrm{\AA}} = -\tilde{\varepsilon}_0 \varepsilon_{\mathrm{m}} \lambda_{\mathrm{\AA}} \int\limits_{a_\lambda}^{b_\lambda} r_\lambda^2 \sinh(\varphi(r_\lambda)) dr_\lambda, \tag{20}$$

where $\varphi(r_\lambda)$ is the solution to the nonlinear PBE boundary-value problem (17).

## Renormalization of Charge and Radius

It therefore follows from the arguments presented in the previous Subsection that far from the sphere the potential determined by the nonlinear PBE (17) can be accurately approximated by the potential coming from the linearised PBE (19). It is easy to assert

---

[1] V.I. Vishnyakov, G.S. Dragan, V.M. Evtuhov, Nonlinear Poisson-Boltzmann equation in spherical symmetry, *Physical Review E*, **76**, 036402 (2007).

that the general solution of (19) vanishing at infinity is given by

$$\varphi(r_\lambda) = C\frac{e^{-r_\lambda}}{r_\lambda}$$

with the general constant $C$ to be determined from the remaining boundary condition; particularly, using (15) one immediately arrives at

$$\varphi(r_\lambda) = \frac{q_e e^{a_\lambda}}{\tilde{\varepsilon}_0 \varepsilon_{\mathrm{m}}(1 + a_\lambda)\lambda_{\text{Å}}}\frac{e^{-r_\lambda}}{r_\lambda}. \tag{21}$$

Let us renormalise the parameters $q_e$ and $a_\lambda$ in (21) by replacing:

- $q_e$ with $q_e + Q_{NL}/e$ (i.e, the bare charge $q_e$ plus the counterions' charge $Q_{NL}/e$ of the nonlinearity layer);
- the original normalized spherical radius $a_\lambda$ with the normalized threshold distance $b_\lambda$ (where it is assumed the nonlinearity regime ends – see Subsection 6).

The potential (21) normalised in such a way is to be used in the linearity domain $r \geq b$ instead of the solution to the original nonlinear PBE (see (16), (17)). Figure 10 shows the reliability and robustness of such a renormalisation even for extremely highly charged small spheres.

## Numerical Examples

We present numerical assessments of the quantities $b$ and $Q_{NL}$. Taking $\lambda = 8.071$ Å, $\varepsilon_{\mathrm{m}} = 80$, $\varepsilon = 2$, a numerical solution of the problem (18) for radius $a = 60$ Å, gives us Figs. 7, 8 and 9, showing: the value of the potential $\varphi$ at the spherical surface $r = a$; show the difference $b - a$, i.e. the width of the non-linearity layer; the charge $Q_{NL}$ defined in (20). For $q_e = -144$ we get $\varphi(a) = -1.788$, $b - a = 20.473$ Å, and $Q_{NL} = 131.401e$.

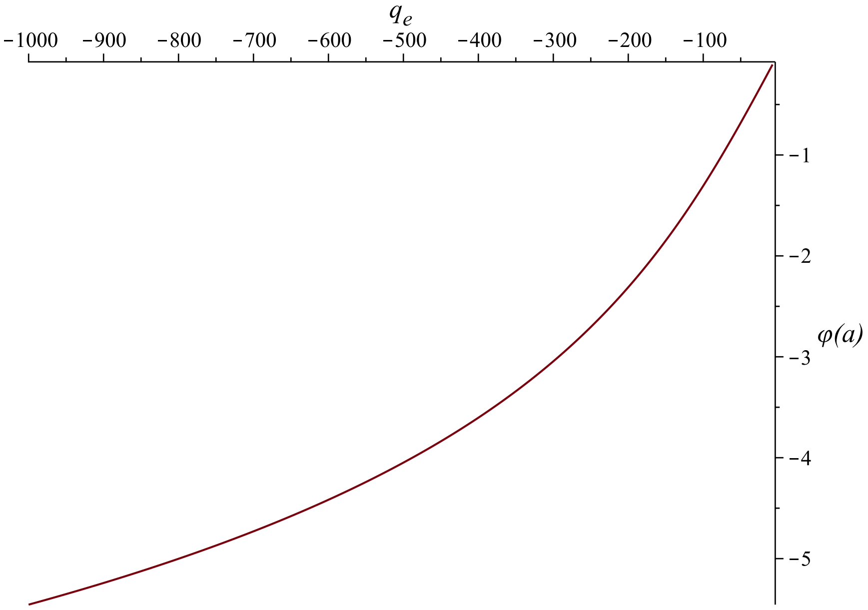


Figure 7: The value of $\varphi(a)$ as a function of $q_e$; $a$ = 60 Å.

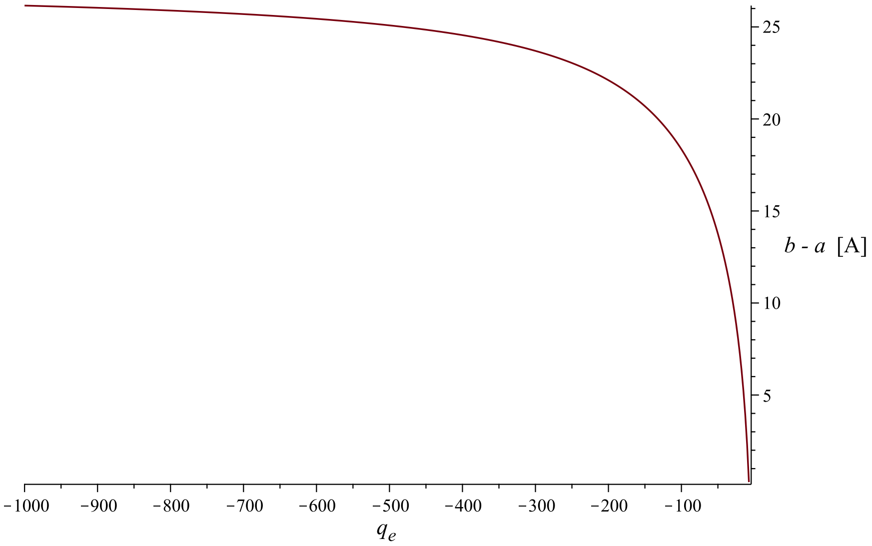


Figure 8: The nonlinear layer thickness $b - a$ as a function of $q_e$; $a$ = 60 Å.

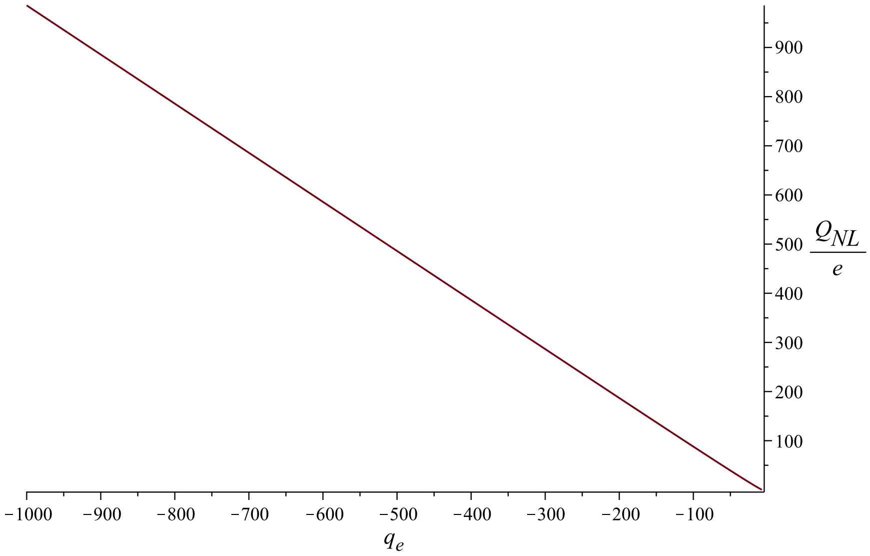


Figure 9: Nonlinear layer charge $Q_{NL}$ as a function of $q_e$; $a$ = 60 Å.

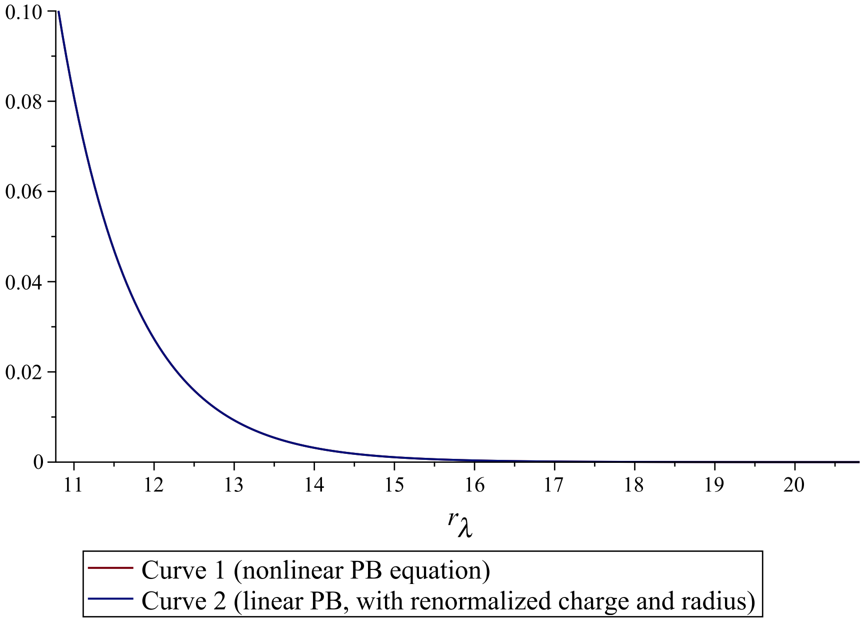


Figure 10: Curves 1 and 2 denote the solution to the original nonlinear problem (17) and the solution (21) with renormalized charge and radius, respectively; $q_e = 10^5$, $a$ = 60 Å, $b_\lambda \leq r_\lambda \leq b_\lambda + 10$. The maximum difference between two curves is $3.75 \cdot 10^{-5}$.