Unexpected mechanical properties of nucleic acids

Aleksander V. Drozdetski

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Physics

Alexey V. Onufriev, Chair
Michel J. Pleimling, Co-Chair
Leo Piilonen
Uwe C. Täuber

April 27, 2016
Blacksburg, Virginia

Keywords: Nucleic Acids, Persistence Length, Molecular Dynamics

Unexpected mechanical properties of nucleic acids

Aleksander V. Drozdetski

(ABSTRACT)

Mechanical deformations of nucleic acids (NA) play a very important role in many biological life processes. The bending persistence length of DNA is of specific interest, because so much eukaryotic DNA that stores genetic information is tightly packed inside cell nuclei, even though DNA is considered to be a relatively stiff biopolymer. However, recent experiments suggest that DNA may be more flexible than its persistence length ($\sim$ 150 bp or $\sim$ 47 nm) suggests, especially for fragments shorter than 100 bp. It is important to reconcile these two seemingly competing pictures of DNA bending by providing a model that can explain the novel results without discrediting old experiments and the widely-accepted worm-like chain model. Another factor that influences both molecular geometry as well as mechanical properties is the ionic atmosphere surrounding the NA. It is known that multivalent ions with charge of $+3e$ and higher can condense DNA into aggregates at high enough concentration. However, most conventional models cannot explain why RNA and DNA condense at different concentrations. Furthermore, our recent simulation results suggest that even though DNA persistence length decreases with multivalent ion concentration due to increasing electrostatic screening, RNA actually becomes stiffer due to a structural transition from the internal binding of the counterions.

Contents

List of Figures

List of Tables

Publications

Journal

1. **A. Drozdetski**, I. S. Tolokh, A. V. Onufriev - Mechanical properties of NA in the presence of multivalent ions (in preparation)

2. **A. Drozdetski**, A. Mukhopadhyay, A. V. Onufriev - Stongly bent double-stranded DNA : reconciling theory and experiment (in preparation)

3. **A. Drozdetski**, I. S. Tolokh, L. Pollack, N. Baker, A. V. Onufriev - Opposing effects of multivalent ions on flexibility of DNA and RNA (under review)

4. S. A. Pabit, A. M. Katz, I. S. Tolokh, **A. Drozdetski**, N. Baker, A. V. Onufriev, L. Pollack - Understanding Nucleic Acid Structural Changes by Comparing Wide-Angle X-ray Scattering (WAXS) Experiments to Molecular Dynamics Simulations, *The Journal of Chemical Physics* (2016)

5. I. S. Tolokh, **A. V. Drozdetski**, L. Pollack, N. A. Baker, A. V. Onufriev - Multi-shell model of ion-induced nucleic acid condensation, *The Journal of Chemical Physics* (2016)

6. I. S. Tolokh, S. A. Pabit, A. M. Katz, Y. Chen, **A. Drozdetski**, N. Baker, L. Pollack, A. V. Onufriev - Why double-stranded RNA resists condensation, *Nucleic Acids Res.* 42(16):10823-31 (2014)

7. R. Anandakrishnan, **A. Drozdetski**, R. C. Walker, A. V. Onufriev - Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations, *Biophysical journal* 108.5: 1153-1164 (2015).

The thesis is based on the publications 2,3,5 and 6. Other publications are unrelated to this thesis.

Research Presentations

1. Condensed Matter Seminar, Physics Department, Virginia Tech - October 26, 2015

2. Molecular Biophysics Symposium, Virginia Bioinformatics Institute, Virginia Tech (poster)

3. Biophysical Society 59th Annual Meeting, Baltimore, MD, USA - February 7-11, 2015 (poster)

4. Biophysical Society 58th Annual Meeting, San Francisco, CA USA - February 15-19, 2014 (poster)

5. Biophysical Society 57th Annual Meeting: Philadelphia, PA - February 2-6, 2013 (platform talk)

6. Biophysical Society Pennsylvania Networking Event, Physics and Chemistry of Biological Systems (poster)

Chapter 1

Introduction

Nucleic acids (NA) are arguably the most important polymer in the world, because they are responsible for the genetic information that holds the key to all living organisms. Ribonucleic acids (RNA) were first to evolve into molecules heredity, but they are somewhat unstable, especially compared to its ultimate successor, deoxyribonucleic acid (DNA). While RNA still plays a huge role in cellular processes, particularly transcription and translation, the genetic information of almost all living organisms today is stored using DNA. Many molecular mechanisms important to life involve processes of mechanical deformation that affect DNA functions such as the compact packing of DNA into chromatin gene expression, and formation of DNA-protein complexes. While the famous double-helix structure of DNA has been known since Watson and Crick published their theoretical paper on its geometry in 1953[269], there have been many studies refining the structure and investigating its mechanical properties such as stretching, twisting, and bending rigidity. While all mechanical properties provide important insights into the dynamics of NA, bending deformations play a significantly larger role in most DNA processes compared to stretching and twisting. In fact, roughly 80% of all eukaryotic genomic DNA is packed tightly inside cell nuclei [209], which requires a significant amount of bending. But in order to understand how DNA bends, it is crucial to have a clear picture of the underlying mechanisms that are responsible for influencing DNA flexibility. There are many known factors that affect the bending flexibility of DNA, such as its structural conformation, ionic strength conditions, and even genetic composition (*i.e.* sequence). Furthermore, recent experimental evidence suggests that DNA may be more flexible than predicted by the classical picture[54, 55, 274, 262], which has stirred up quite a bit of controversy in the field. While many argue that the widely-accepted "harmonic" bending is a valid model for elastic DNA deformation [70, 172, 266, 176], there has been more and more evidence, both experimental and from simulations, of the breakdown of the classic model in the so-called *strong* bending regime.

NA condensation mediated by multivalent counterions is another biologically important phenomenon. It has been studied extensively [155, 20, 167, 33, 86, 272, 27, 28], and many models

exist to describe the process [189, 36, 244, 178, 72, 125, 126, 5, 250, 251, 127, 118], and yet there are many unanswered questions regarding the mechanism on the atomic level. Since NA are negatively charged molecules, they inherently repel other NA, and attract positively charged counterions. When bound to the NA molecules, these counterions effectively neutralize the phosphates by "screening" the electrostatic repulsion. If the degree of neutralization is high enough, the NA molecules can actually condense into aggregates if the charge of the counterions is $+3e$ or higher[272, 27, 28]. This phenomenon has been studied extensively [92, 93, 276, 97, 34, 206, 157, 211, 212, 29, 146, 128, 231, 182, 170, 127, 110, 151, 119], yet there is no single accurate model of NA condensation and there are many unexplained effects. A better understanding of this process can help shed light on many biological processes involving NA, such as packing and gene expression.

1.1 Background Information

1.1.1 NA structure

NA are biopolymers typically consisting of nucleotides; even though there are many examples of single-stranded NA (such as messenger RNA used in gene expression), this thesis focuses specifically on double-stranded NA. Each nucleotide consists of a nitrogenous base (purine or pyrimidine), a 5-carbon sugar ring (ribose in RNA and deoxyribose in DNA), and a phosphate group (see Fig.1.1). Within each strand, the nucleotides are joined by phosphodiester bonds, and the two strands are held together by hydrogen bonds between the bases, with two complimentary nucleotides forming a base-pair (bp). In DNA, Adenine (A) is paired with Thymine (T), and Guanine (G) is paired with Cytosine (C). For RNA, the complimentary base to Adenine is Uracil (U), rather than Thymine. These are usually referred to as Watson-Crick pairs. While other pairings occur in nature such as in the case of DNA triplex, Watson-Crick pairing are by far the most common configuration of base-pairs found in naturally occurring double-stranded NA.

Due to its highly-charged nature (linear charge density of $-2e$ per bp), that is a result of the phosphates in the backbone, double-helical NA are unstable in vacuum due to electrostatic repulsion between the complimentary strands. In solution, it is stabilized by the presence of positively-charged, neutralizing counterions, typically monovalent ions such as Na^+ (from salt such as NaCl). The self-repulsion of phosphates also contributes to the molecule's apparent stiffness. In fact, there have been several attempts to quantify the electrostatic contribution to the bending persistence length of DNA [204, 235, 165]

Most DNA is considered to be in what is known as canonical B-form conformation, which is roughly 10 bp/helical turn (about 36° inter-bp twist) with a helical rise of ~ 34 Å and a diameter of ~ 20Å. While this is considered to be the most common form of DNA, another biologically relevant conformation of DNA is the A-form, which has a shorter helical rise

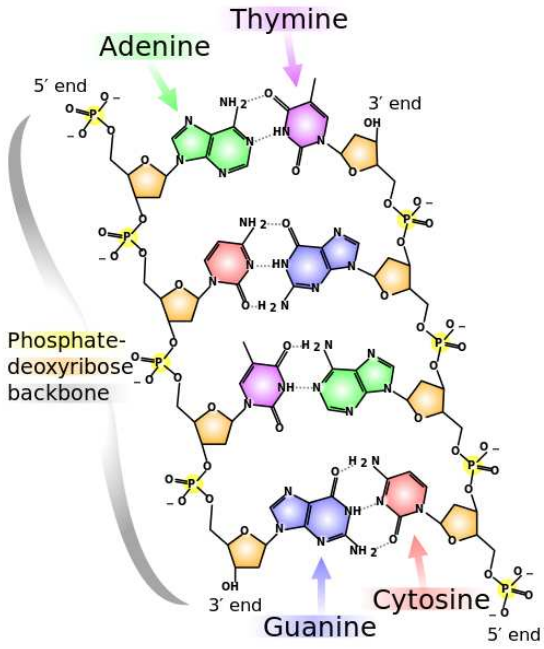

Figure 1.1: A schematic representation of the DNA structure.

(∼ 30 Å) than B-form, but is slightly wider (∼ 23 Å), and has a smaller inter-bp twist (∼ 33°). The characteristic features of A-form DNA are tilted base-pairs and a small opening along the helical axis when viewed from above. It is worth noting that while the two forms are discussed here in the context of DNA, they are also applicable to double-stranded RNA.

The main difference between DNA and RNA is the hydroxyl group bonded to the 2′ C atom in RNA, which is absent in DNA. While this seems like a minor difference that doesn't change some of the other important qualities of RNA compared to DNA (like charge density), this structural difference becomes crucial in determining possible conformations of double-stranded RNA (dsRNA). Notably, the presence of the extra oxygen restricts the possible conformations of the backbone, forcing RNA to have an A-form double-helix with the B-form being highly unfavorable energetically. Furthermore, the presence of the hydroxyl group results in reduced stability of RNA due to its susceptibility to hydrolysis. While most RNA *in vivo* is actually single-stranded (ssRNA), recent findings indicate that dsRNA plays a more important role in living organisms than previously thought[255, 98, 38].

1.1.2 Polymer bending and the worm-like chain

The behaviour of a flexible polymer with respect to bending is governed by its bending persistence length L_p, which is a measure of the length-scale at which the tangents to the polymer become uncorrelated:

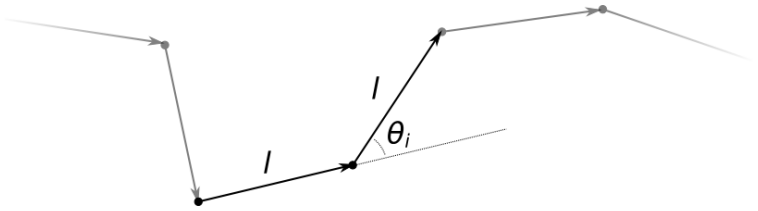

Figure 1.2: A schematic representation of a part of a long polymer, showing a bending site with angle θ_i between two successive segments of length l.

$$\langle \cos\theta \rangle_s = e^{-s/L_p} \tag{1.1}$$

here θ is the angle between two tangents separated by a distance s along the polymer. In other words, segments of the polymer shorter than L_p can be considered more or less as rigid rods, unable to bend significantly from thermal fluctuations.

The current most widely-accepted model of DNA bending is the worm-like chain (WLC) model, which treats the polymer as a chain of discreet segments of length $l \ll L_p$ and assumes the fluctuations of individual bend angles between successive segments θ_i to be independent. A real polymer can also stretch and twist, but the simplest form of WLC decouples these modes and considers only the bending degree of freedom with the total energy defined as:

$$E^{WLC}_{chain}(\theta_1, \theta_2, \ldots, \theta_N) = \sum_{i=1}^{N} E_{bend}(\theta_i) \tag{1.2}$$

where θ_i is the angle between two consecutive segments (see Fig. 1.2), or a bending site, and $E_{bend}(\theta)$ is the effective bending energy function. If we expand $E_{bend}(\theta_i)$ around a global minimum θ_0 ($\partial E_{bend}(\theta_i = \theta_0)/\partial\theta_i = 0$), the lowest order remaining term is necessarily quadratic:

$$E^{hWLC}_{chain}(\theta_1, \theta_2, \ldots, \theta_N) = \sum_{i=1}^{N} \left(\frac{1}{2} a(\theta_i - \theta_0)^2 + \mathcal{O}(\theta_i^3) \right) \tag{1.3}$$

where a is a constant. Furthermore, without loss of generality we can assume no intrinsic curvature of the polymer ($\theta_0 = 0$), *i.e.* the ground state is a straight line. Then, for small enough θ_i, just the quadratic term should provide an accurate enough approximation:

$$E^{hWLC}_{chain}(\theta_1, \theta_2, \ldots, \theta_N) = \sum_{i=1}^{N} \frac{1}{2} k_B T \frac{L_p}{l} (\theta_i)^2 \tag{1.4}$$

where k_B is the Boltzmann constant and T is the temperature. This is called the *harmonic* WLC (or hWLC for short), which is essentially the bending equivalent of Hooke's law. In the continuum macroscopic limit, it corresponds to Landau's model of a thin elastic rod. The hWLC model has been quite successful at describing the behaviour of DNA, and many classical experiments [19, 59, 59, 265, 232, 99, 100, 254, 70] have found it to be in perfect agreement with measurements. The harmonic model of DNA bending has led to the development of a set of detailed, sequence-dependent energy functions that describe DNA deformation at the level of base-pair steps[56]. Within the range of their applicability, these models are extremely useful in practice as they can predict relative free energies of deformation of realistic DNA sequences.

However, given the nature of the approximation, it is evident that for large enough values of θ_i it will no longer be accurate, as the contribution from the higher order terms can no longer be neglected. These terms can potentially lead to non-trivial effects: for example, it is the an-harmonic terms in atomic potential that lead to the existence of thermal expansion phenomenon in solids[11]. In the context of DNA, this begs the question: is this threshold outside of biologically relevant bending, or does the hWLC model break down well before the strong bending regime? While this question remains unanswered, some recent experiments suggest that the latter might be the case[55, 274, 262]. It is important that the deviations were seen only on relatively short length scales: central limiting theorem necessitates that the polymer behavior approaches hWLC on long length scales[274, 273]. It was noted[54] that the "classical" experiments were typically performed on long length scales and therefore may have missed these new effects.

The picture becomes even more complicated when ionic strength of the solvent is taken into account. Since NA are negatively-charged polymers, the presence of positively charged counterions can affect the bending rigidity of the molecule due to Debye-Hückel electrostatic screening of the phosphates. Conventional models suggest that the bending flexibility should increase with ionic concentration, and this has been found to agree with experimental evidence for monovalent counterions. One might expect that multivalent counterions produce the same result, but there have not been enough studies done to confirm this hypothesis, especially for RNA. While there have been many studies on DNA bending persistence length and how its affected by the structure of the molecule and its surrounding environment, dsRNA has only recently become of interest to researchers. Specifically, it has been found to play a key role in RNA interference[255], and may have potential medicinal applications [98, 38].

1.1.3 Molecular Dynamics

While it is unlikely that any single approach, experimental or theoretical, will be able to provide a definitive resolution of the issues outlined above, atomic level models and molecular dynamics (MD) simulations can complement experiment and provide unique in-

sights. Modern force fields[57, 160, 81, 47] have been shown to accurately reproduce a variety of structural and dynamical properties of small fragments of canonical and non-canonical nucleic acids in water[49, 237, 48, 177, 243, 228, 50, 89, 25, 75, 76, 183, 22], their interaction with solvent[23, 77, 224, 162], counterions[177, 101, 77, 12, 208], and proteins[31, 24, 14, 253, 135]. Importantly, forcefields such as AMBER[43] can reproduce high-level quantum mechanical calculations for hydrogen bonding and back stacking interactions[107, 4, 195, 241, 240, 242]. Moreover, these forcefields can now provide reasonable descriptions of non-trivial conformational changes such as the $A \rightarrow B$ transition in duplex and triplex DNA[49, 237, 48, 177, 243, 228, 51].

1.1.4 Implicit Solvent and the Generalized Born approximation

The traditional all atom explicit solvent molecular dynamics can yield a very detailed picture of complex interactions between biomolecules and the surrounding solvent, but it is too demanding computationally[281] to accomplish all of the tasks set in this thesis. Recently, novel approaches that greatly enhance conformational sampling and provide simple means of estimating free energy of conformational states have become available for nucleic acids[53, 121, 281]. These methods have been employed to model DNA free in solution[282, 267, 261, 238, 213], binding between proteins and nucleic acids[282, 14, 115, 62, 6, 53], as well as for energetic analysis of conformational changes such as the $A \rightarrow B$ transition[260].

The implicit solvation methodology[58, 108, 21, 161, 88, 220, 156] represents water by a uniform continuous medium that has the properties of water, but without individual water molecules. The model is relatively new, but has already shown considerable promise[247, 104, 105, 221, 200, 71, 114, 87, 58, 18, 144, 78], especially in molecular dynamics applications[69, 61, 239, 40, 259, 268, 188, 84, 234, 185, 143].

The total energy of a solvated molecule includes solvation effects implicitly, via $E_{\text{tot}} = E_{\text{vac}} + \Delta G^{\text{solv}}$, where E_{vac} is molecule's energy in vacuum (gas-phase), and ΔG^{solv} is the free energy of transferring the molecule from vacuum into solvent, *i.e.* solvation free energy. The vacuum energy is computed in the same manner as E_{tot} in the explicit solvent approach, but without water or counterions. To compute the solvation contribution, ΔG^{solv}, it is decomposed into the electrostatic and non-electrostatic parts:

$$\Delta G^{\text{solv}} = \Delta G^{\text{solv}}_{\text{el}} + \Delta G^{\text{solv}}_{\text{nonpolar}}, \tag{1.5}$$

The ΔG_{el} part also includes the screening effects of ionic atmosphere. In practice, ΔG_{el} can be computed either at the level of Poisson-Boltzmann (PB) equation[108], or its analytical approximations such as the generalized Born (GB) model[247, 18]. The hydrophobic effects are implicitly included in $\Delta G_{\text{nonpolar}}$, typically through a surface tension (SA) term.

The continuum, or implicit solvent, model has several critical advantages over the explicit

water representation:

1. Considerably lower computational costs;
2. improved sampling (mainly due to the absence of viscosity);
3. ability to easily estimate relative energies of molecular structures, even very large ones. Note that E_{tot} defined above implicitly includes free energy of solvent rearrangement, including the contribution from the ionic atmosphere.
4. Absence of potential artifacts of periodic boundary conditions. This may be especially important for highly charged systems where large conformational changes are expected.

The rate of conformational sampling in explicit solvent is expected to be enhanced by 20 to, perhaps, even 100 fold for the DNA[260, 278]. Thus, a 10 ns simulation effectively corresponds to hundreds of nanoseconds or more of "real" time. The implicit solvent approach can therefore be considered as a technique that greatly enhances conformational sampling while preserving the thermodynamic properties of the system.

The particular implicit solvent method used in my research was the GB approximation, which represents the electrostatic term of the solvation contribution ΔG^{solv} as:

$$\Delta G_{el} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp -r_{ij}^2/(4R_i R_j)}} \tag{1.6}$$

where r_{ij} is the spatial separation between two charges q_i and q_j, $R_{i/j}$ are the effective Born radii that quantify the degree of how buried the atoms are within the molecular cavity, and $\epsilon_{in}/\epsilon_{out}$ are the dielectric constants inside and outside the molecule respectively.

1.1.5 Further advancements in computational methods

Another approach to speeding up simulation rate to obtain longer time-scales is better, optimized hardware. One way to do this is to build a machine specifically designed to carry out MD simulations, such as Anton[226]. These powerful supercomputers are capable of accessing millisecond time-scales with explicit solvent[227]. However, not only is this expensive, it is somewhat impractical due to its narrow applicability and limited accessibility to researchers.

The alternative is to use available hardware that can be adapted (via special software) for faster calculations used in MD simulations. Specifically, parallel computing is ideal for calculating atomic energies at a much faster rate than conventional methods. Supercomputers with multiple CPUs per node show up to tenfold increase in simulation speeds, and can

be used by many researchers who have access to local or remote clusters. However, these supercomputers are frequently overloaded with jobs for many other research projects and sometimes require formal requests for use, limited to a certain number of computational hours. This may be undesirable for MD simulations, which can sometimes take months even on these powerful machines.

A more accessible option to researchers is graphical processing unit (GPU) accelerated computation. GPUs are specifically designed hardware for parallel computation, originally created for graphics-intensive applications such as ray-tracing. Recent advances in computational methods allow these modules to be used for Molecular Dynamics, such as in conjunction with the AMBER platform. While powerful GPU cards can be considered somewhat expensive, the total cost is usually significantly lower than computational time on a supercomputer, and can be acquired and maintained by researchers without the need to share the resources with other research groups. Most of the MD simulations done for this thesis were performed on NVIDIA GPUs using the CUDA framework, either on our group's own machines or using the HokieSpeed supercomputer.

1.2 Overview

The main goal of my research was to use computational tools and methods to investigate the non-trivial aspects of mechanical properties of nucleic acids and nucleic acid condensation, and to propose theoretical models that are in agreement with both experimental data and predictions from simulations.

In Chapter 2, we have outlined a general framework of polymer bending and applied it to the case of DNA cyclization, or the formation of loops from linear fragments. The CHC model shows profound agreement with recent experimental data that suggests the probability of cyclization of fragments smaller than 100 bp is magnitudes higher than predicted by classical models. We then investigate the origin of non-convexity of DNA bending energy that leads to CHC-like behavior of the molecule. Finally, we investigate the relationship between strength of confinement of DNA-protein complexes and the bimodal bending observed in unconfined, constrained polymers with CHC bending energy.

In Chapter 3 we present a multi-shell model of NA condensation that explains the difference in condensation propensity between DNA and RNA in the presence of multivalent counterion Cobalt(III) Hexammine, ($Co(NH_3)_6^{3+}$) or simply referred to as CoHex. First we focused on the MD simulations that helped understand the underlying mechanism and showed that the number of ions bound *externally* to the duplexes correlated really well with measured NA condensation propensity. Then we developed a novel, quantitative multi-shell model of NA condensation, by taking into account the observations from MD simulations. The model successfully explains aggregation of short NA molecules and makes predictions for longer molecules.

In Chapter 4 we investigated a novel effect observed in MD simulations of dsRNA with multivalent ions. Contrary to the expectation, CoHex counterions cause RNA persistence to increase, rather than decrease, by binding inside the electrostatically favorable major groove and effectively "locking up" the double-helix. We have found this effect to be robust to ion type and concentration, sequence details, and force-field parameters, and developed a simple qualitative model that explains this unexpected phenomenon.

Works unrelated to this thesis

Apart from the work discussed in the following chapters, I also have collaborated on several additional projects which are not a part of this thesis.

My very first project involved implicit solvent MD simulations of asymmetric phosphate neutralization of short (21 bp) DNA molecules. There have been both experimental[248] and computational[134, 102] studies that showed electrostatically-induced bending of DNA when one side of the duplex was partially neutralized by replacing several negatively-charged phosphate groups with neutral methylphosphonates. I attempted to reproduce this effect using implicit solvent MD, which allows for much faster simulation speed and thus can provide longer simulation times necessary for equilibration of the system to obtain more accurate statistics. While I was able to confirm there was in fact comparable bending to the previous studies, in order to obtain a more complete picture, different combinations of methylphosphonate isomers had to be simulated, so the work was never published. Despite this fact, this work gave us confidence in using implicit solvent MD simulations for large systems that are either computationally expensive or even impossible to create due to the exceptionally large size of solvent box and number of water molecules required.

I have worked with Ramu Anandakrishnan on investigating the computational speed-up (or slowdown) of implicit solvent simulations of small, large and mixed conformational changes with respect to explicit solvent MD. I helped conduct some MD simulations for the project, performed parts of the analysis, and participated in the preparation of the manuscript.

We have collaborated with an experimental group at Cornell University and a theoretical group at Pacific Northwest National Laboratory on using wide-angle x-ray scattering (WAXS) measurements in combination with MD simulations to understand the structural changes that occur when CoHex counterions are added to the solution. On its own, WAXS data can be quite difficult to interpret, but when used in conjunction with MD, we can gain insight into the structural transitions that occur when counterions bind to the NA duplexes. Furthermore, this shows that WAXS measurements can be used to refine and improve MD force-fields. My part in the project was helping conduct necessary MD simulations and participation in the preparation of the manuscript.

I am also currently collaborating with Arman Fathizadeh from University of Illinois at Chicago on investigating the potential relationship between bending persistence length es-

timates and nucleosome binding affinity. We have simulated several long DNA molecules with various sequences in the hope that we can gain insight into any possible link between bending persistence length of DNA and its nucleosome binding affinity. Specifically, we are looking at known nucleosomal sequences (such as the one in the well-known X-ray structure of nucleosome core particle NCP147, PDB ID: 1KX5), as well as sequences considered to be unfavorable for nucleosome binding (such as the TGGA repeat).

Chapter 2

Strongly bent double-stranded DNA: reconciling theory and experiment

2.1 Introduction

Polymer deformation is ubiquitous, and studying the elastic properties of these macromolecules is crucial for understanding their dynamics. Biopolymers are abundant in nature and play crucial roles in many biological processes[96]. Given their relevance to many key cellular processes[85, 39, 180], it is very important to understand not only their structure, but physical properties as well. [113, 263, 91, 120] While there are many different kinds of biopolymers, DNA stands out as a case of its own. Understanding DNA flexibility is crucial for mechanistic grasp of vital cellular functions such as packaging of DNA compactly into chromatin, formation of protein/DNA complexes and protein regulation[85, 196]. An all-important example of DNA deformation, relevant to a variety of biological processes that depend on the elastic properties, is DNA looping, which occurs in many prokaryotic [94] and eukariotic [155, 209] systems. There are many regulatory proteins that can loop DNA in various conformations, most notably the nucleosome, which is involved in DNA packing within cell nuclei. Not only is DNA looping crucial in regulation of many biological processes involving DNA[222], but the majority of eukareotic genomic DNA (75-80%) is packed tightly into nucleosomes and subsequently chromatin fibers [209]. The nucleosome consists of roughly 150 bp of DNA looped almost twice around a histone core, which is significant bending on such a small scale. However, despite decades of intense experimental and theoretical effort, the story of how this arguably most important polymer behaves under deformation is far from complete, with controversies and new developments abound[262, 266, 176, 219].

The bending flexibility of a polymer is conventionally quantified in terms of its persistence length, L_p, a length scale below which the polymer behaves more or less like a rigid rod. Specifically, L_p is defined as length of DNA segment over which the time-averaged orientation

of the polymer becomes uncorrelated; for fragments smaller than L_p, the thermal fluctuation alone is not enough to induce significant ($\sim$ 1 rad [258]) bending. This definition of L_p can be exploited to qualitatively separate the two bending regimes: if no considerable bending occurs on length scales shorter than L_p, the polymer can be deemed *weakly* bent; otherwise the bending is assumed to be *strong*. A variety of experimental techniques [19, 59, 59, 265, 232, 99, 100, 254, 70], reveal that DNA is unusually stiff, with L_p roughly equal to 500 Å($\approx$ 150 bp) for mixed sequence B-form DNA. Thus, the fundamental level of DNA packaging – the nucleosome – contains strongly bent double-stranded DNA.

Response of DNA to mechanical stress has been studied extensively [82, 233, 168, 169, 19, 29, 85, 139, 54, 274, 15, 26, 32, 55, 68, 70, 103, 138, 173, 225, 248, 199, 196], which led to a consensus in modeling the weak bending regime, while the strong bending regime is highly controversial, as we show below. Arguably the most widely used simplified model of DNA bending is the worm-like chain (WLC) model. In the original WLC[136, 173, 70], the polymer is modeled as a continuous, isotropic elastic rod with its deformation energy being a quadratic function of the deformation angle. A more common variant of this model is the discrete version, where the bending energy of the polymer consisting of N segments of length l is:

$$E = \sum_{i}^{N-1} \frac{1}{2} k_B T \frac{L_p}{l} \theta_i^2 \tag{2.1}$$

where θ_i is the angle between two consecutive segments (see inset of Fig. 2.1). Various theoretical models of DNA bending, including sequence-dependence [56, 140, 67, 83], were consequently developed that assumed harmonic angular deformation energy of DNA. There is little doubt that the Hookean, "elastic rod" models accurately describe many polymers in the weak bending regime [96], including double-stranded DNA[70, 173, 175]. Indeed, lowest order term of a Taylor series expansion of any well-behaved function around its local minimum is quadratic, which means that for small deviations from equilibrium, the response function can be considered harmonic. However, by the same logic it should be expected that beyond a certain threshold the bending energy may no longer be approximated by the quadratic term alone, and that the strong bending regime contains regions of the energy function where the behavior of the polymer differs from the weak bending regime.

Within the last few decades, the prevailing view of DNA as a Hookean polymer was challenged when experiments were able to investigate the flexibility of DNA on scales smaller than several L_p. Previously, only very large fragments (hundreds to thousands of base-pairs) were investigated[19, 59], and thus only the weak bending regime was probed. Small DNA fragments ($\approx$ 100bp) were found to have a higher probability of cyclization (spontaneous formation of loops) than that predicted by the WLC theory[55]. This discovery sparked considerable controversy, which still remains unresolved. What is particularly puzzling is that strongly bent DNA appears less rigid than the DNA in the Hookean regime. Some of the follow-up experimental and theoretical work supports the validity of WLC even on small scales [70, 172, 266, 176], while others show that smaller fragments of DNA are indeed much more flexible[274, 262] than previously thought, and that bending of short double-helical

DNA fragments can not be described by a harmonic model[274].

In this chapter we propose, and verify against available experiment, a unified theoretical description of polymer bending that treats the weak and strong bending regimes on the same footing, guided by a simple physical principle.

2.2 The proposed unified framework of polymer bending

Consider a polymer chain made of $N \gg 1$ inextensible, identical monomer subunits with "effective" (implicitly includes long-range interactions between subunits) bending deformation energy $E(\theta_i)$ for each bending site, where θ_i is the angle between two successive segments (see inset of Fig. 2.1). The total energy of the polymer is $E_{chain} = \sum_i^N E(\theta_i)$, and without the loss of generality we assume no intrinsic bends, *i.e.* $E(0) = 0$. In order to induce an average non-zero bend in the chain, the polymer must be constrained, and the problem of finding the equilibrium polymer conformation is reduced to minimizing E_{chain}, subject to the specific constraint of the problem. Here we assume that entropic effects can be neglected at length scales of interest (not much larger than the polymer persistence length) – an assumption that we explicitly confirm below by numerical experiments.

In classical thermodynamics, for a system to be stable against a macroscopic fluctuation in energy, the entropy of the system as a function of energy ($S(E)$) must be concave (non-convex). Any chord connecting two points on a graph of $S(E)$ must lie below the curve itself in order to satisfy the second law of thermodynamics (maximum S). Conversely, the inverse function $E(S)$ must be convex. If, however, $E(S)$ is not convex over some region, the system phase-separates once this region is reached, with the properties of the two phases corresponding to the end points of the convex hull of the non-convex region. The actual, physical average energy of the system follows the convex hull, which makes the energy manifestly convex. This very general reasoning is applicable to polymer stretching[219] and macroscopic mechanical deformations[13, 37]. Here we use the analogy to describe polymer response to bending.

Consider a uniformly bent polymer – constant curvature along the chain. A closed circular loop would be an example. By construction, such a polymer consists of identically bent subunits with the bending angle equal to the average deformation angle, $\bar{\theta} = N^{-1} \sum_i^N \theta_i$, and its total energy $E_{chain} = NE(\bar{\theta})$. Here $E(\theta)$ is the *effective* bending energy of the polymer, which takes into account all the interactions, short- and long- range, between its monomers.

Next we consider a realistic situation where the polymer bending is enforced by just one constraint: that the total sum of the monomer bend angles remains constant, $\alpha = \sum_i^N \theta_i = const.$. A circular loop, $\sum_i^N \theta_i = 2\pi$ would be an example. Mathematically, the problem

of finding the minimum energy conformation of the polymer can be represented as energy minimization under a specific constraint:

$$NE(\bar{\theta}) = \min_{\sum_{i=1}^{N} \theta_i = N\bar{\theta} = \alpha} \{\sum_{i}^{N} E(\theta_i)\} \tag{2.2}$$

Using Lagrange multipliers, Eq. 2.2 can be reduced to $\min\{E(\theta_1)+\cdots+E(\theta_N)-\lambda(\theta_1+\cdots+\theta_N-\alpha)\}$. Differentiating with respect to θ_i gives a set of equations $\partial_{\theta_i} E(\theta_i) - \lambda = 0$ (for all i) which leads to a set of equalities $\partial_{\theta_1} E(\theta_1) = \partial_{\theta_2} E(\theta_2) = \cdots = \partial_{\theta_N} E(\theta_N)$. For a convex functional form of $E(\theta)$, $\partial_\theta E(\theta)$ monotonically increases with θ, and therefore the equalities are satisfied only if $\theta_1 = \theta_2 = \cdots = \theta_N$. However, for a non-convex function we can have two stable conformational states characterized by θ_a and θ_b bend angles respectively, such that $\partial_{\theta_i} E(\theta_a) = \partial_{\theta_i} E(\theta_b)$ for some $\theta_a < \theta_b$. Therefore, in the energy minimized state of a polymer with a non-convex effective bending energy, all subunits are bent to one of the two stable conformations: in one, the units are bent stronger than in the other. Let $0 < p < 1$ represent the fraction of all the bending sites that are in the state θ_a. Since the remaining sites are in the state θ_b, their fraction of the total number of sites N is given by $1-p$. Then the total bending angle in terms of θ_a and θ_b is given by $Np\theta_a + N(1-p)\theta_b = N\bar{\theta} = \alpha$, and the minimum bending energy per monomer in the non-convex region is $E(\bar{\theta}) = pE(\theta_a) + (1-p)E(\theta_b)$. Rewriting $p = (\bar{\theta} - \theta_b)/(\theta_a - \theta_b)$, we have

$$E(\bar{\theta}) = \frac{\bar{\theta} - \theta_b}{\theta_a - \theta_b}(E(\theta_a) - E(\theta_b)) + E(\theta_b) \tag{2.3}$$

which proves that in the non-convex region, the polymer energy is a linear function of the average deformation $\bar{\theta}$ (see Fig. 2.1).

2.3 Bending of a circular loop, weak and strong

While many different types of constraints can be physically realized, one of the most important ones is the closed loop constraint, which is also used in DNA cyclization experiments[233, 70, 55] critical in the study of the strong bending regime. Consider the case of a single closed loop $\alpha = \sum_i^N \theta_i = 2\pi$. While in general nothing restricts the loop from bending in 3D, unless there is a lot of torsional energy in the closed polymer, the loop can be assumed to be confined to a 2D plane to simplify the mathematics of the following model. As we show using coarse-grained simulations of such closed loops (with a defined bending energy), the energy minimized states exhibit minimal out of plane bending, so this assumption should not significantly affect the outcome of our theoretical predictions (see Fig. 2.2).

To arrive at a theory that can account for both the weak and strong bending regimes simultaneously, we use the form of Eq. 2.3 for the strong bending regime, while retaining WLC for the weak bending. In the proposed Convex Hull Chain(CHC) model, the average per unit

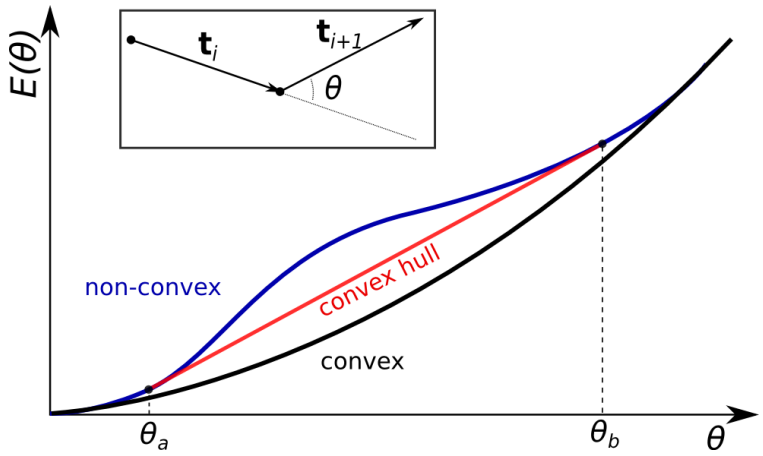

Figure 2.1: Two different forms for a bending energy profile of a homopolymer. Shown is the (effective) bending energy per site $E(\theta)$. If the profile is purely convex down (black curve), the minimal energy conformations of the polymer is uniform bending (all sites are identically bent). If the function has a non-convex region (blue curve), non-uniform bending is more energetically favorable. In this case the total energy of the system follows the convex hull of the energy curve (red line).

(e.g. per base-pair) bending energy is described by a single piece-wise polynomial: quadratic WLC (Eq. 2.1) for $\bar{\theta} < \theta_a$, and a linear function (convex hull of $E(\theta)$) for $\theta_a < \bar{\theta} < \theta_b$:

$$E(\bar{\theta}) = \begin{cases} \frac{1}{2} k_B T L_p \bar{\theta}^2 & \text{if } \bar{\theta} \leq \theta_a \\ \\ k_B T L_p \theta_a (\bar{\theta} - \frac{1}{2}\theta_a) & \text{if } \theta_a < \bar{\theta} < \theta_b \end{cases} \tag{2.4}$$

where L_p, the persistence length, is expressed in number of bending sites (*e.g.* number of base pairs for DNA loops); k_B is the Boltzmann constant and T is the temperature. Here we are not interested in the very strong bending regime $\bar{\theta} > \theta_b$, since for DNA this would correspond to loops smaller than 10bp, which are physically impossible due to steric constraints and are much smaller than anything observed in cyclization studies[262, 149]. From Eq. 2.4, the total bending energy of a closed loop of total length L (in bp) is given by $E_{loop} = E(\bar{\theta})L$. Since $\bar{\theta} = \frac{2\pi}{L}$, the bending energy of the loop is:

$$E_{loop}(L) = \begin{cases} 2\pi^2 k_B T \frac{L_p}{L} & \text{if } L > \frac{2\pi}{\theta_a} \\ \\ k_B T L_p \theta_a \left(2\pi - \frac{1}{2} L \theta_a\right) & \text{if } \frac{2\pi}{\theta_b} < L < \frac{2\pi}{\theta_a} \end{cases} \tag{2.5}$$

Note that, where defined, the new function depends on just one new parameter: θ_a – left boundary of the non-convex domain. In fact, the energy beyond θ_b has no effect on the result of two-state bending (except for defining one of the stable states and the upper bound of

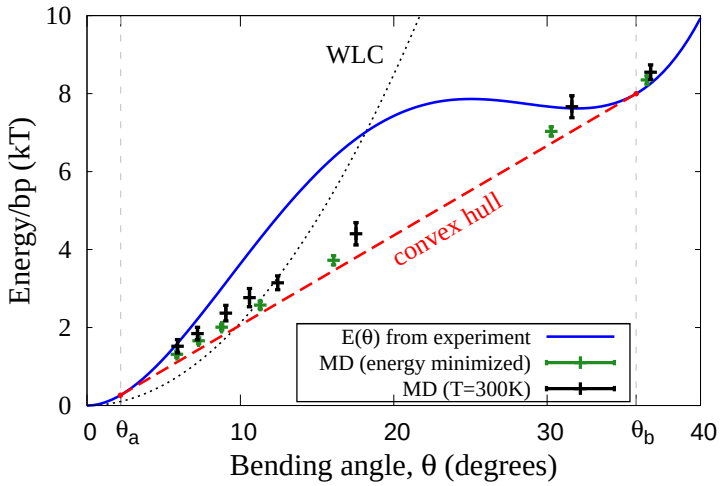

Figure 2.2: DNA bending energy $E(\theta)$ (per bp) extracted from the probability distribution[70] of DNA bends that naturally occur in protein-DNA complexes (blue line), and the average energy of unrestrained DNA closed loops simulated via coarse-grained MD with the same $E(\theta)$ (crosses). Green symbols: energy minimized (simulated annealing) loops. Black symbols: loops simulated at T=300K. In both cases, the average loop energy as a function of average bend angle $\bar{\theta} = \theta$ follows the convex hull of $E(\theta)$. The small deviation of the T=300K points from the convex hull are due to the entropic contributions neglected within the CHC model based on energy minimization. However, the almost perfect agreement of the energy minimized structures with the convex hull suggests that there is minimal out of plane bending, which justifies our assumption for the theory of closed loop formation.

the non-convex region). In an actual cyclization experiment, the loop can have a "teardrop" geometry so that $\sum_{i=1}^{N} \theta_i = \alpha = 2\pi - \phi$, $\phi \geq 0$. However, a non-zero ϕ will not change the form of Eq. 2.5, affecting only the over-all multiplicative pre-factor in the cyclization probability discussed in the next section.

2.4 Application to double-stranded DNA

The preceding discussion was completely general: non-uniform bending and the linear bending regime can be a feature of any polymer. However, since DNA is arguably the most important polymer, and it exhibits looping in many different biological systems, we will focus on double-stranded DNA for the rest of the study. An effective bending energy (per bp) calculated from a statistical analysis of experimental PDB structures of DNA-protein complexes [70] is shown in Fig. 2.2. This effective bending energy function has a non-convex region, and thus a convex hull, the end points of which are $\theta_a = 2.2°$ and $\theta_b = 35.8°$

Numerical simulations (see Section 2.12) demonstrate that a polymer with this effective

bending energy between monomers exhibits all of the features discussed above. For large loop sizes the bending angles are small (weak bending) – the system samples the convex (harmonic) region of the energy function, Fig. 2.2. However, as the loop size decreases, the average angle per bending site $\bar{\theta}$ increases, and eventually crosses the θ_a threshold. Once this happens, the total energy of the system will increase linearly with $\bar{\theta}$ (until $\bar{\theta} > \theta_b$, *e.g.* Eq. 2.5), due to the fact that it is energetically more favorable for the bending sites to be a linear combination of θ_a and θ_b states, as discussed in the framework section. Coarse-grained simulations at T=300K confirm that the average energy per bp follows the convex hull in this region (see Fig. 2.2).

2.5 Comparison with cyclization experiments

Most experimental cyclization results are expressed in terms of the Jacobson-Stockmayer *j-factor* for DNA fragments of different lengths. [55, 70] *j-factor* of a polymer is used to estimate the probability that a linear polymer forms a closed loop by joining its cohesive ends [233, 112]. For an unconstrained DNA [275, 229, 7] it is typically modelled using the following equation,

$$J(L) \simeq \frac{k}{L_p^3}\left(\frac{L_p}{L}\right)^5 \exp\left(-\frac{E_{loop}}{k_B T} + \frac{L}{4L_p}\right) \tag{2.6}$$

where $\frac{k}{L_p^3}\left(\frac{L_p}{L}\right)^5 \exp\left(\frac{L}{4L_p}\right)$ accounts for the entropic contribution, averaged over possible looping geometries, and $\exp\left(-\frac{E_{loop}}{k_B T}\right)$ is the energy penalty of bending the DNA fragment to form the loop. We note that the pre-factor k depends on the loop closing geometry and can be expected to remain invariant over a relatively short range of loop lengths L, within the same experiment. To make a direct connection with cyclization experiments for non-integer numbers of helical repeats, we modulate (see Section 2.11) with $cos(2\pi L/10)$ the torsionally independent loop energy from Eq. 2.6 (with $E_{loop}(L)$ from Eq. 2.5 for CHC and $E_{loop}(L) = 2\pi^2 k_B T \frac{L_p}{L}$ for WLC), which is a standard way [229] to account for the periodic variation of the *j-factor* due to the torsional component of the energy [229]. For more details on the calculation of the *j-factor*, please see Section 2.11. We compare the proposed model and WLC with the most recent experiment [262] in Fig. 2.3.

As seen from Fig. 2.3, the effective potential modeled with CHC leads to an excellent quantitative agreement with the cyclization experiments, while the j-factors predicted by conventional WLC are off by several orders of magnitude in the strong bending regime (WLC is known to work well in the weak bending regime where it coincides with CHC by construction). Counterintuitively, the predicted envelope function of *j-factor* (*i.e. j-factor* without the 10 bp periodic component, Eq. 2.6 for CHC) has a minimum near 45 bp and begins to increase for even smaller loops, whereas for WLC *j-factor* decreases sharply for small loops. The over-all variation of the *j-factor* as a function of the loop length for the two

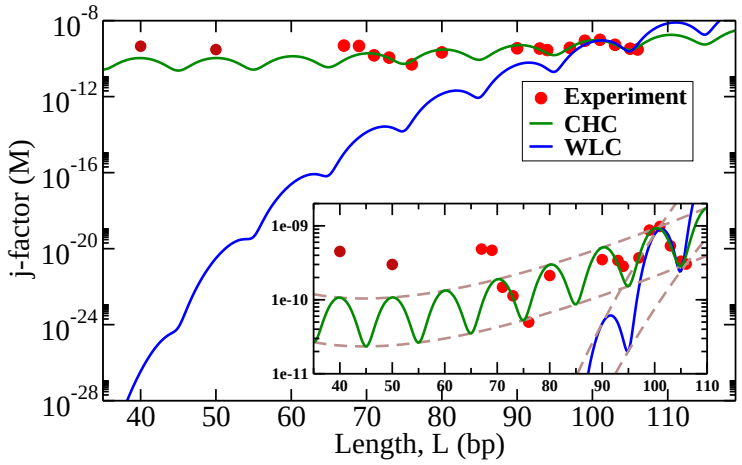

Figure 2.3: DNA cyclization *j-factors* computed using the proposed model (green line) and WLC (blue line) are compared with recent experiment [262] (red dots, $L > 60$ bp). Experimental values of persistence length, $L_p = 150$ bp and $\theta_a = 2.2°$ (Fig. 2.2) were used; the value of the prefactor k in Eq. 2.6 was obtained independently for each model as best fit against two experimental data points for fragment length L = 101 and 106 bp. The envelopes of the *j-factor* (brown dashed lines) for CHC and WLC are shown in the inset. Predicted envelope for CHC *j-factor* has a minimum near 45 bp. The experimental data points $L = 50$ and $L = 40$ bp were shared by Taekjip Ha in private communication to assess model performance after the model was constructed.

models is governed by the interplay between entropic and energetic costs $E_{loop}(L)$ of forming the loop. For small loops, the entropic penalty of forming the loop decreases with the loop size L; however, $E_{loop}(L) \to \infty$ for small L within WLC, which leads to a steep decrease in the cyclization probability. In contrast, CHC loop energy, Eq. 2.5, approaches a constant for $L \to 0$, which explains why the *j-factor* increases for small enough L. The very different qualitative behavior of WLC and CHC *j-factors* for small loops can be used as a definitive experimental test of the models.

2.6 Origin of non-covexity of DNA bending energy

Within the proposed CHC framework, the central role is played by convexity properties of the effective bending energy between individual monomers. For the DNA, that energy derived from statistical analysis of experimental structures of protein-DNA complexes (Fig. 2.2), has a clear non-convex region, which is responsible for the "softer", linear bending mode of short DNA loops.

In order to investigate the physical origin of the non-convexity of the DNA effective bending

energy, we extracted the average bending energy per bp from all-atom Molecular Dynamic (MD) simulations of DNA mini-circles. The total bending energy profile obtained from these simulations, along with the breakdown into components of different physical origin, are shown in Fig. 2.4; one can clearly see a prominent non-convex region, in qualitative agreement with the experiment, Fig. 2.2. The key parameter θ_a is $\sim 1.5^\circ$ for the MD simulations, which compares well to the experimental value of 2.2°, Fig. 2.2. Some discrepancy is likely due to sequence effects[142] or force-field issues[46, 215]. Importantly, Fig. 2.4 also suggests that the primary contribution to the non-convexity of the bending energy function can be mostly attributed to two non-bonded contributions: Van-Der Waals (VDW) and electrostatic. For small bending angles, the total energy closely follows the electrostatic term (which can be approximated by a quadratic function in this regime), while contributions form the VDW and internal (bond, angle and dihedral terms) degrees of freedom are nearly opposite, and cancel each other out. However, once the bending reaches the transition angle θ_a, the VDW energy decreases at a rate faster than the increase of the other terms combined, which results in a non-convex region of the total $E(\theta)$, Fig. 2.4. It is this sharp decrease in the VDW contribution that gives rise to the existence of a non-convex region in the DNA bending energy. Further analysis, (inset in Fig. 2.4), reveals that the VDW interactions between DNA backbone atoms (backbone-backbone) are peculiar and critical to the counterintuitive sharp decrease. While the base-stacking (basepair-basepair) interaction energy increases somewhat with the bending angle, the backbone-backbone term decreases quickly, similarly to the behavior of total VDW energy over the same angle range. The key role of the backbone-backbone VDW term suggests that it is the overall structure of DNA, rather than sequence details, that is responsible for the existence of the convex hull in the polymer's effective bending profile. Variation in the DNA sequence may alter the range of bending angles over which the convex hull exists.

To illustrate, at the atomic level, the origin of the non-convex behavior of the backbone-backbone VDW energy, consider pairs of oxygen (O) atoms in the backbone. We choose these atoms because they have one of the lowest VDW energy minima out of all atom pairs in the DNA, and have most atom pairs within the effective short range of the interaction. The O-O VDW energy as a function of pairwise distance r_{ij} has a potential well at $r_{min} \sim 3.3\mathring{A}$, Fig. 2.5. As the double helix bends, the geometry of the backbone deforms in a way that more oxygen atoms pairs fall into this well ($\sim 3.1 - 3.5$Å): beyond θ_a the accumulation of the attractive contributions begins to sharply lower the total interaction energy compared to the native, unbent state where all the oxygen atoms are significantly further apart (>4.0Å) than the well minimum. As the helix bends further, these O-O pairs eventually pass the LJ minimum, and begin to climb onto the repulsive wall, which increases the total VDW interaction energy.

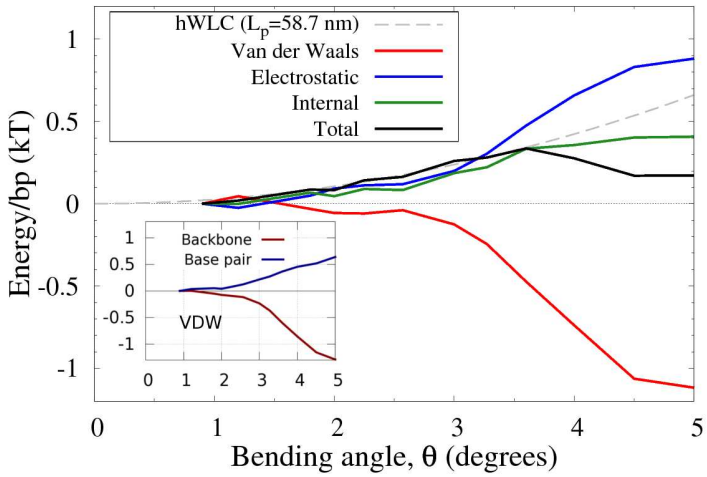

Figure 2.4: The bending part of the DNA potential energy ΔE and its components (per bp) as a function of the bending angle θ, from all-atom MD simulations of DNA circles of variable lengths (50-400 bp). For reference, an WLC fit for the small angle data ($\approx 1 - 3.5°$) (grey dashed line) yields the persistence length of 58.2 nm ($\approx$ 172 bp), quite close to the experimental value of $\approx$ 50 nm ($\approx$ 150 bp) . The main contribution to the non-convexity of the total potential energy comes from the Van der Waals (VDW) interactions. Further examination suggests that it is the backbone-backbone interaction in particular that contribute the most to non-convexity due to a sharp decrease in energy for $3° < \theta < 4°$, as shown in the inset.

2.7 Beyond closed loops: the "nucleosome"

The proposed framework is based on one main assumption: despite constraints, the polymer chain is still free to explore sufficient conformational space to search for minimum energy. So far, we focused on DNA loops because of direct connection to key cyclization experiments; the single constraint $\sum_i^N \theta_i = \alpha = 2\pi$ is minimally restrictive. However, other realistic scenarios of DNA bending, notably in protein-DNA complexes, may involve different types of constraints that can confine the polymer strongly enough to potentially violate the main assumption to various degrees. To investiage to which extent our main conclusion – deformation energy of strongly bent DNA follows the convex hull of $E(\theta)$ – may still hold in a realistic protein-DNA complex, we explore a system that mimics what is arguably the most important DNA-protein complex: the nucleosome, Fig. 2.6. A negatively charged polymer ("DNA", monomer charge q_s) is effectively restricted to bending in a plane (details in Methods), and allowed to wrap around a charged disk (the "nuclesome core"). We vary the total (central) charge Q of the disk to modulate the electrostatic attraction of the polymer to the disk, and, hence, the degree of the polymer confinement. In the limit of very strong confinement ($|Q/q_s| \to \infty$), the polymer is forced to be confined to a circular, uniformly bent path and has very few degrees of freedom left to explore in this regime, solid red line in Fig. 2.6.

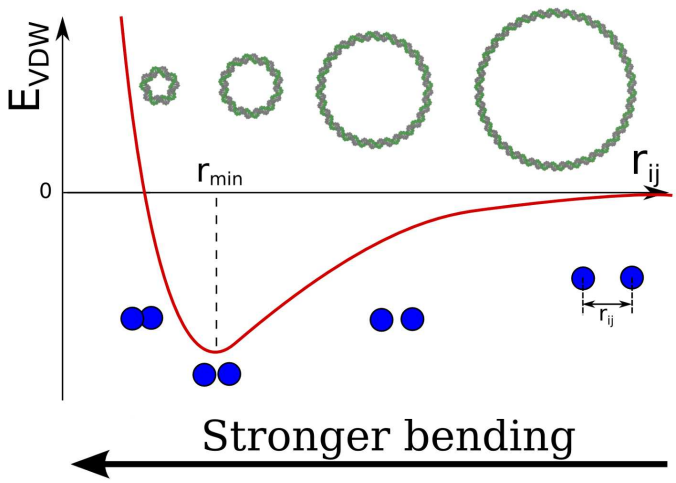

Figure 2.5: The increase of the bending angle of the DNA duplex causes a decrease in the average distance between atoms pairs that contribute significantly to the total VDW energy; the atoms move deeper into the LJ potential well. The shape of the well is conducive of a sharp decrease in the total VDW energy upon small changes in r_{ij} caused by DNA bending. Once the average distance passes the well minimum, the VDW energy begins to increase again.

Since all the monomers are identically bent in this case, the average bending energy simply follows the given functional form of $E(\theta)$ (red dashed line in Fig. 2.6), and CHC clearly does not apply. As we decrease the confinement strength, the polymer is allowed to explore more degrees of freedom, and can assume non-uniform bending conformations while lowering its total bending energy. The effective bending energy per monomer begins to approach the convex hull (solid green and blue lines), making CHC more applicable. At the lower end of the confinement strength (solid purple line), the Coulomb energy of the chain is smaller than the total bending energy of the polymer, but still strong enough to keep it loosely bound to the disk against thermal fluctuations. In this case the polymer is allowed to relax close to the minimal energy state conformation, while the sum of the bent angles is still constrained. This is the limiting case described by our CHC model, and the resulting energy per monomer follows the convex hull fairly closely (offset by entropic effects at T=300K, see Fig. 2.2).

We argue that it is this low confinement regime, where CHC is relevant, that corresponds to the real nucleosome. To this end, we use a coarse-grained 147-bp DNA fragment placed next to a cylinder with relative dimensions of the actual histone core[79] (see Section 2.9.2); as the core charge Q is increased, the whole fragment starts to wrap around the cylinder once the confinement strength is $|Q/q_s| \geq 90$. At this value of the DNA confinement, the energy cost of pulling away a fragment of ~ 20 bp in our model is $\approx 10 k_B T$, which is in the same order of magnitude as $\approx 6 k_B T$ estimated from experiment [85] for the fragment of the same length in the actual nucleosome, which is reasonably close.

The nucleosome represents one of the strongest confined DNA-protein complexes that still exhibits CHC-like behavior, so complexes with similar or weaker confinement strength will also have a CHC bending regime provided they are under certain constraints. Moreover, even for higher degrees of confinement ($|Q/q_s| \simeq 200$, blue lines in Fig. 2.6) still approximate the convex hull, and so CHC is probably applicable, at least qualitatively.

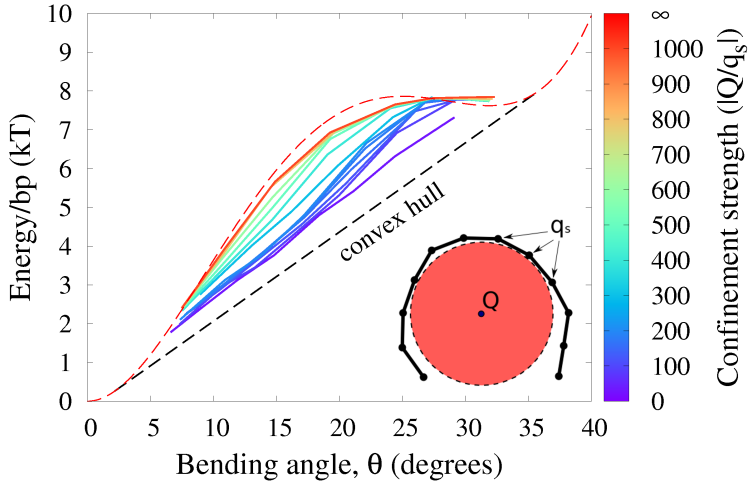

Figure 2.6: Polymer bending in a "protein-DNA complex" model with variable strength of polymer confinement, see Section 2.9.2. Shown is the average energy per bead against the average bending angle θ, at different confinement strengths (governed by the charge ratio $|Q/q_s|$, of the polymer charge to the opposite confining charge). The intrinsic bending of the polymer is described by $E(\theta)$ from Fig. 2.2; the system follows its convex hull under weak confinement, while approaching $E(\theta)$ for strong confinement.

The same considerations can be applied to any effective bending energy that has a distinct non-convex region; the qualitative results are the same if a recently proposed kinkable WLC (kWLC) potential [266] is used instead of $E(\theta)$ from Fig. 2.2

2.8 Conclusion

It is well-known that DNA behaves like a (harmonic) WLC polymer in the regime of weak bending, but if fragments of DNA of around one persistence length (and smaller) have significant angular deformation from the straight (or equilibrium) conformation, the bending can no longer be accurately described with the classical model. Here we have proposed a novel, general framework for bending of polymers, which is based on the consideration of convex properties of the effective bending energy between successive monomers. Within the framework, the convex hull of such an energy function defines region of "phase transition" from weakly bent regime where the bending energy is harmonic, and the strongly bent regime

in which the system's energy is a linear function of the average bending angle.

The new theory agrees with recent experimental cyclization data from Vafabaksh and Ha [262], but it does not have to stop at DNA, and can be applied to other polymers as well. Future advances in material and polymer science can potentially allow polymers to be designed to a specific bending energy profile to allow for a transition to the two-state phase at predetermined bending thresholds.

2.9 Computational Methods

2.9.1 Molecular Dynamics simulation of closed DNA loops

DNA circles of various sizes (50-400 bp), sequence poly(dA).poly(dT) and helical repeat of 10 bp, were generated using AMBER Tools software NAB. For all of the MD simulations, implicit solvent Generalized Born (AMBER option igb=1) approximation was used to treat solvation effects, including 0.145M of monovalent salt. All circles were initially minimized for 1000 steps with P atoms restrained harmonically with a force constant of 1.0 kcal/mol/Å^2. Each system was then heated to 300K and equilibrated for 100 ps with the same restraints as for minimization. Finally, the DNA circles were subject to 1 ns of molecular dynamics with P atoms restrained harmonically with a force constant of 0.1 kcal/mol/Å^2. The energies were saved every 20 fs, and averaged over the whole trajectory. In Fig. 2.4, ΔE was computed as the difference between per bp energies of the given circle and the largest circle simulated, which is virtually unbent.

2.9.2 Coarse-grained simulations of confined fragments

ESPResSo software [148, 9] was used to create and simulate a coarse-grained 20 bead fragment of "DNA" bound to a spherical charged "protein". The polymer fragment was set up with the angular bending energy $E(\theta)$ per bead taken from experimental data[70] (see Fig. 2.2). Each bead carries unit charge q_s =-1; however non-bonded interactions between the beads were turned off. The bead charges interact only with a positive charge Q surrounded by a spherical, impenetrable constraint of radius R. In addition, two impenetrable walls were placed above and below the charge Q to minimize out-of-plane bending of the "DNA". The confining charge Q was varied from 10-1000 effectively increasing confinement strength (defined here as $|Q/q_s|$). The constraint radius R was also varied to sample various total bending angles.

For the nucleosome model, the system described above was modified to mimic the confinement of DNA around realistic histone core. The "DNA" fragment size was increased to 147 bp, and the non-bonded interactions between monomers were turned on for additional

realism[79]. The fragment was confined around a cylinder of fixed diameter ∼100Å, and the walls were placed ∼50Å apart (approximate dimensions of the nucleosome complex[79]).

2.10 Convex vs. non-convex bending energy function

An infinitesimal perturbation $\Delta\theta$ that reduces the angle at one bending site means that another bending site must increase its angle by $\Delta\theta$ in order to satisfy the constraint $\alpha = N\bar{\theta}$. This perturbation changes the total energy by $E(\bar{\theta}+\Delta\theta) + E(\bar{\theta}-\Delta\theta) - 2E(\bar{\theta})$. If $E(\theta)$ is a convex function (e.g. the black curve in Fig. 2.7), the perturbed system will have a higher energy than the initial uniformly bent case. This is because, by definition, a convex curve always lies below its chords, so that $2E(\bar{\theta}) < E(\bar{\theta}+\Delta\theta) + E(\bar{\theta}-\Delta\theta)$. Thus, the system would gain energy due to the perturbation, and therefore the minimum energy conformation for a polymer with a convex effective bending energy will remain that of a uniformly bent chain. However, if the function $E(\theta)$ has a non-convex region, e.g. the blue curve in Fig. 2.7, then for some $\bar{\theta}$ in that region $2E(\bar{\theta}) > E(\bar{\theta}+\Delta\theta) + E(\bar{\theta}-\Delta\theta)$, which means it is possible to lower the energy of the polymer by bending the sites to two different bending angles, with the new value of the chain energy falling on the line connecting the two new bending states ($\bar{\theta}+\Delta\theta$ and $\bar{\theta}-\Delta\theta$) on the energy curve (see dashed red line in Fig. 2.7). In fact, as long as the line that connects the two perturbed states on the energy curve is below the curve itself, the energy can still be lowered until its value is on that line. The energy can not be lowered further once the system reaches the convex hull (solid red line in Fig. 2.7), a line tangent to the non-convex curve at two points, such that between these two points the value of the function is greater than that of the convex hull.

2.11 Fitting *j-factor* with experiments

To derive the expression of the *j-factor*, we will assume the following functional form adopted from Ref [229]. Assuming the same entropic contribution for a WLC polymer, the expression for the *j-factor* can be written as:

$$J(L) \simeq \frac{k}{L_p^3}\left(\frac{L_p}{L}\right)^5 \exp\left(-\frac{E_{loop}}{k_B T} + \frac{L}{4L_p}\right) \tag{2.7}$$

where L_p is the persistence length (150 base pairs), L is the loop length and $k_B T$ is the Boltzmann factor. k is an unknown constant that is specific to the bending energy model; the value of k in this work will be deduced from experimental data of the *j-factor*. The bending energy, E_{loop}, for a given DNA loop within the CHC model is given by (see Section

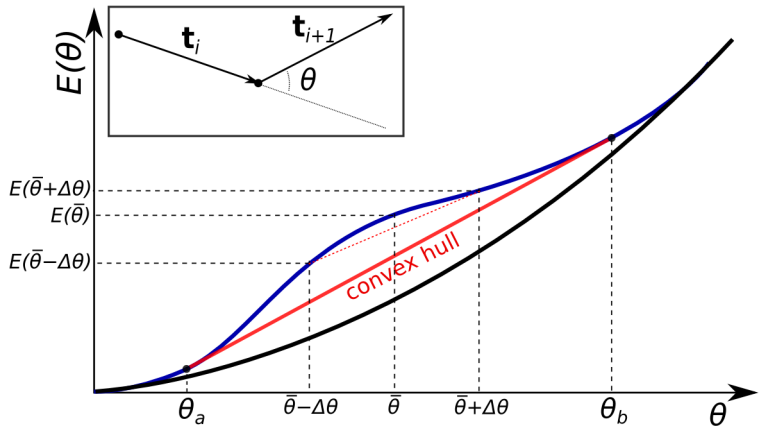

Figure 2.7: Two different forms for a bending energy profile of a homopolymer. Shown is the (effective) bending energy per site $E(\theta)$. If the profile is purely convex down (black curve), the minimal energy conformations of the polymer is uniform bending (all sites are identically bent). If the function has a non-convex region (blue curve), non-uniform bending is more energetically favorable. In this case the total energy of the system follows the convex hull of the energy curve (red line).

2.3 for details),

$$E_{loop}(L) = \begin{cases} 2\pi^2 k_B T \frac{L_p}{L} & \text{if } L > \frac{2\pi}{\theta_a} \\ k_B T L_p \theta_a \left(2\pi - \frac{1}{2} L \theta_a\right) & \text{if } \frac{2\pi}{\theta_b} < L < \frac{2\pi}{\theta_a} \end{cases} \tag{2.8}$$

This new loop energy function depends on just one parameter: θ_a – left boundary of the non-convex domain. Note that, the energy beyond θ_b has no effect on the result of two-state bending (except for defining one of the stable states and the upper bound of the non-convex region).

Using the value of $\theta_a = 2.2°$ extracted from experimental structures [70], for loop length larger than $\frac{2\pi}{\theta_a} \simeq 164$ bp the loop bending energy is given by the WLC energy function, the CHC function is used otherwise. The value of k is deduced from the two experimental *j-factor* data points for loop lengths – $L = 101$, and 106 bp. To do this we use the Eq. 2.7 to define the asymptotes. The torsional dependence of the *j-factor* is represented by a cosine function, similar to that of the original work [229], to oscillate between the upper and lower asymptotes periodically in an interval of 10 bp:

$$J(L) = \frac{1}{2}\left[J^{top}\left(1 + \cos\left(\frac{2\pi L}{10}\right)\right) + J^{bot}\left(1 - \cos\left(\frac{2\pi L}{10}\right)\right)\right] \tag{2.9}$$

where the asymptotic curves are obtained directly using the functional form of the *j-factor* defined in Eq. 2.7,

$$\begin{aligned}\frac{J^{top}(L)}{J(101)} &= \left(\frac{101}{L}\right)^5 e^{(L-101)\left\{\frac{1}{4L_p}+\frac{L_p\theta_a^2}{2}\right\}} \\ \frac{J^{bot}(L)}{J(106)} &= \left(\frac{106}{L}\right)^5 e^{(L-106)\left\{\frac{1}{4L_p}+\frac{L_p\theta_a^2}{2}\right\}}\end{aligned} \tag{2.10}$$

For the WLC region we use the same constant k found above but replace the expression for E_{loop} as defined in Eq. 2.8. The full expression of the *j-factor* spanning both regions is given by a similar expression as Eq. 2.9 where the expression for J^{top} and J^{bot} is replaced by:

$$\begin{aligned}J^{top}(L) = J(101) \times \left(\frac{101}{L}\right)^5 \Big[\mathbf{\Theta}\left(\frac{2\pi}{\theta_a}-L\right) e^{(L-101)\left\{\frac{1}{4L_p}+\frac{L_p\theta_a^2}{2}\right\}} \\ + \mathbf{\Theta}\left(L-\frac{2\pi}{\theta_a}\right) e^{\left\{-\frac{2\pi^2 L_p}{L}+L_p\theta_a\left(2\pi-\frac{101\theta_a}{2}\right)+\frac{L-101}{4L_p}\right\}}\Big]\end{aligned}$$

$$\begin{aligned}J^{bot}(L) = J(106) \times \left(\frac{106}{L}\right)^5 \Big[\mathbf{\Theta}\left(\frac{2\pi}{\theta_a}-L\right) e^{(L-106)\left\{\frac{1}{4L_p}+\frac{L_p\theta_a^2}{2}\right\}} \\ + \mathbf{\Theta}\left(L-\frac{2\pi}{\theta_a}\right) e^{\left\{-\frac{2\pi^2 L_p}{L}+L_p\theta_a\left(2\pi-\frac{106\theta_a}{2}\right)+\frac{L-106}{4L_p}\right\}}\Big]\end{aligned} \tag{2.11}$$

where $\mathbf{\Theta}(x)$ is the standard Heaviside theta function, such that for $x > 0$, $\mathbf{\Theta}(x) = 1$ and $\mathbf{\Theta}(x) = 0$ otherwise. The resulting curves obtained from Eqn.2.9 and 2.11 are shown in Fig. 3 in the Main text.

To compare directly with the case of pure WLC, i.e. when the loop energy is given by WLC in both the tight bending and the weak bending region, the expressions for J^{top}, and J^{bot} is given by:

$$\begin{aligned}J^{top}(L) &= J(101) \times \left(\frac{101}{L}\right)^5 e^{\left\{\frac{L-101}{4L_p}+2\pi^2 L_p\left(\frac{1}{101}-\frac{1}{L}\right)\right\}} \\ J^{bot}(L) &= J(106) \times \left(\frac{106}{L}\right)^5 e^{\left\{\frac{L-106}{4L_p}+2\pi^2 L_p\left(\frac{1}{106}-\frac{1}{L}\right)\right\}}\end{aligned} \tag{2.12}$$

Table 2.1: *j-factor* ratios, J(L)/J(L^{ref}), predicted using CHC and WLC models compared with experiment[262].

L(bp)	L^{ref}(bp)	Experiment	CHC	WLC
40	50	1.50	9.93×10^{-1}	1.12×10^{-6}
71	101	1.51×10^{-1}	2.01×10^{-1}	3.08×10^{-6}
80	101	2.17×10^{-1}	3.28×10^{-1}	6.10×10^{-4}
90	101	3.56×10^{-1}	5.59×10^{-1}	3.73×10^{-2}

2.11.1 Minima of the asymptotes and ratio comparison

The minima of the expression for the two models can be derived directly from the expression of *j-factor*. For CHC, the minima is given by:

$$L = \frac{5}{\frac{1}{4L_p} + \frac{L_p\theta_a^2}{2}} \tag{2.13}$$

Using the value of $L_p = 150$ bp and $\theta_a = 2.2°$, the minima of *j-factor* is found at $L \simeq 45$ bp. No such minima exists in the WLC case, as the J-ratio approaches 0 as the loop size decreases.

Using this equation and the function $\Delta G_{loop}(L)$, a formula for the ratio of *j-factors* $J(L)/J(L^{ref})$ can be used to predict the difference between *j-factors* for different sized fragments and compare it with known experimental results, using the value $\theta_a = 2.2°$ from the effective DNA bending energy extracted from experimental PDB structures[70]. We also extracted the value for θ_a from the experimental data and equation 2.7 and persistence length $L_p = 500$ Å(150 bp). The results are shown in Table 2.1.

2.12 Coarse-grain simulations of loops with non-convex bending potential

In order to understand how the non-convexity of the bending potential affects the angular distribution of different sized loops, we performed coarse-grain simulations using esspressoMD ([148]). Loops of different sizes (6 - 600 bp) were created with only effective bending energy defined by the function from[70]. The bonds were made practically inextensible (very large coefficient of the quadratic bond energy) to adhere to the model outlined in the main test. The loops were then simulated at $T = 300K$ long enough to create 100,000 statistically significant snapshots that were then analyzed by finding the angular probability distribution of each loop size, which is shown in Figure 2.12.

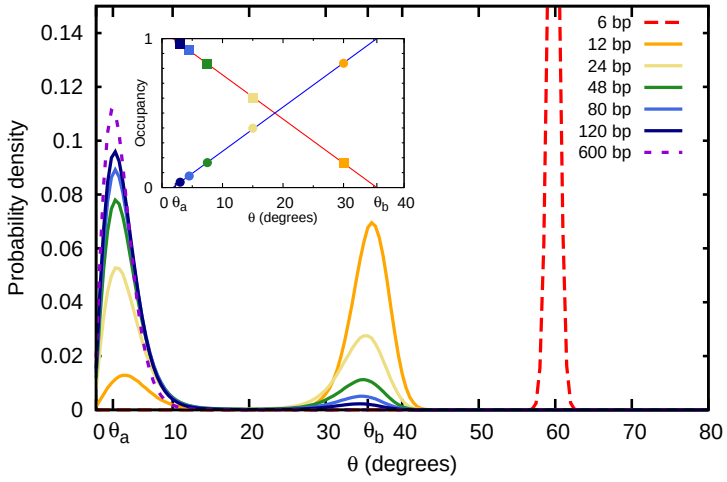

Figure 2.8: Angular probability distributions resulting from a non-convex potential[70]. As the loop size increases, the average bending angle per bp decreases. When the average angle falls into the convex hull range, the angular distribution becomes binomial with peaks at θ_a and θ_b. The fractional occupancy of both of these states is shown in the inset as a function of the average angle $\bar{\theta}$.

Chapter 3

Multi-shell model of Nucleic Acid condensation

3.1 Introduction

Condensation of highly charged DNA and RNA molecules by cationic agents is biologically important for processes such as packaging of genetic material inside living cells and viruses [155, 20, 167, 33, 86], compactization and delivery of small interfering RNA molecules for gene silencing [2] and gene therapy [166]. In aqueous solution, DNA condensation requires cations with charges of $+3e$ or higher [272, 27, 28]; e.g., trivalent cobalt(III) hexammine (CoHex), trivalent spermidine or tetravalent spermine can generally condense DNA at room temperatures, while divalent cations cannot.

Several decades of experimental and theoretical studies of DNA condensation have resulted in a general physical picture wherein the major contribution to the effective attraction is due to electrostatic interactions [92, 93, 276, 97, 34, 206, 157, 211, 212, 29, 146, 128, 231, 182, 170, 127, 110, 151, 119]. Theoretical models have previously been developed to clarify the details of nucleic acid (NA) interactions, starting from the models of interacting uniformly charged cylinders immersed in an implicit ionic bath [189, 36, 244, 178], to nucleic acid models with more realistic helical geometries of molecular charge distributions [72, 125, 126, 5, 250, 251, 127, 118]. Counterion electrostatic treatments range from simple mean-field descriptions to strong coupling models [211, 181, 145, 178, 179, 118, 119] with the more sophisticated strong coupling models able to reproduce NA-NA attraction. It has been shown that mean-field description of the counterion atmosphere always leads to repulsion between the oppositely charged cylinders [97, 119] – correlations between counterions is a necessary condition for the attraction [189, 145, 65, 179]. All-atom explicit solvent molecular dynamics (MD) simulations of short DNA fragments demonstrated the existence of short-range attractive forces between B-form DNA duplexes at sufficient degree of duplex neutralization by multivalent ions [153,

60, 277].

A commonly accepted view has emerged: the effective attraction is mainly due to electrostatic contributions. However, the general picture as well as the atomistic mechanism of NA condensation, including the role of hydration forces [206, 64] or multivalent counterion correlations [211, 151], are still incomplete and cannot be fully explained.

Despite progress in understanding NA-NA interactions, an atomic-level mechanism of multivalent ion-induced nucleic acid condensation has not yet been fully established and some recent experimental data are difficult to rationalize within the accepted models. For example, the condensation propensity of double-stranded (ds) RNA in the presence of trivalent CoHex ion was recently found to be much smaller than for the equivalent sequence of dsDNA [147, 257]. In other words, dsRNA helices resist CoHex-induced condensation under conditions where the DNA duplexes readily condense. The unexpected findings are still very recent, with only a limited number of theoretical models attempting to rationalize it so far. For example, a recent extension [129] of an earlier model [128] of DNA condensation suggests that the striking difference between RNA and DNA condensation stems from the differences in intrinsic helical parameters of the duplexes. However, it has recently been shown experimentally that significant differences in condensation propensity can arise in some NA duplexes without significant differences in their helical parameters [257]. Nevertheless, the model [129] emphasizes critical role of counterion distributions in NA condensation; some of the distributions of bound trivalent counterions assumed by the model are consistent with the latest all-atom molecular dynamics (MD) simulations described in Section 3.2.4, but some are very different, pointing to the critical need to take these atomistic details into account for quantitative and even qualitative agreement with recent condensation experiments.

Because dsDNA is typically found in B-type helical forms, while dsRNA stays in A-like forms, it was suggested that counterion distributions around the differing helical forms may explain the difference in condensation behavior of DNA and RNA [147]. However, atomically detailed CoHex distributions are hard to measure in solution samples. In principle, all-atom explicit solvent molecular dynamics (MD) simulations [132, 133, 153, 60, 45, 123, 124, 279] can provide the necessary details of NA interaction with multivalent ions, including ion distributions around NA, as well as preferred ion binding sites and their occupancies. To ensure equilibration of multivalent ion distributions, especially around RNA, such simulations will have to go well beyond the time scale of tens of nanoseconds – the longest reported atomistic MD simulation to date of small DNA fragments interacting with tri- or tetravalent ions in solution [60].

At this point, only atomistic simulations that treat the solvent (water and ions) explicitly can provide several key details of counterion binding and distributions around nucleic acids, which are otherwise near impossible to obtain experimentally. Among these critical details is thermodynamic characterization of CoHex binding to different regions of NA, which, as we shall see, is necessary for a quantitative description of condensation. However, at this point experiment can only provide average, aggregate binding affinity to NA, but this is

not the key quantity of interest for condensation. Atomistic simulations can bridge the gap with experiment in that respect and provide estimates of the counterion distributions and affinities to various loci in the NA duplexes. In what follows we show that such estimates become valuable in building a quantitative picture of NA condensation – picture that can be verified experimentally.

To uncover the mechanism of RNA resistance to condensation and to determine key factors responsible for different NA condensation propensities, we studied CoHex-induced condensation and CoHex binding properties of four short NA helical constructs using experiments and experimentally guided atomistic MD simulations on appropriate time-scales.

3.2 Materials and Methods

3.2.1 Materials

Single-stranded NAs were purchased from IDT. Four double-stranded NA constructs were made by annealing in STE buffer (50 mM NaCl, 10 mM TRIS, 1 mM EDTA, pH 7.4) at 94°C for 2 min. The 25 bp duplex DNA, RNA and DNA:RNA hybrid all used the same mixed sequence (GCA TCT GGG CTA TAA AAG GGC GTC G, with T replaced by U in RNA strands) employed in our previous studies [147, 8, 193]. For the homopolymeric 25 bp poly(dA):poly(dT) DNA duplex, a UV melt was taken to verify monophasic annealing. CoHex chloride powder was purchased from Sigma-Aldrich and dissolved in water. CoHex concentrations were verified by measuring the absorption at 473 nm using extinction coefficient 56.2 $M^{-1}cm^{-1}$ as reported by Deng and Bloomfield[63]. Duplexes were dialyzed in pH 7.0 Na-MOPS buffer containing NaCl.

3.2.2 CoHex-induced NA condensation as monitored by UV absorption

The NA constructs were dialyzed in 20 mM NaCl and 1 mM pH 7.0 Na-MOPS buffer and separated in 40 μM 100 μL aliquots. Each aliquot was spiked with 5 μL of concentrated CoHex and incubated at 4°C for 2 hours. The aliquots were centrifuged at 10,000 RPM for 10 min and the supernatant was separated for UV-VIS absorption spectroscopy measurements. The absorption spectra were recorded from 220 to 600 nm using a Cary 50 spectrophotometer. The optical density at 260 nm (OD_{260}) was monitored to determine the amount of NA left in the solution. The recorded value was normalized using the OD_{260} of the sample with no added CoHex and reported in Figure 3.1 as the fraction of soluble sample in the supernatant. The error bars shown were determined by propagation of errors due to baseline differences at OD_{600} and presence of residual CoHex at OD_{473}.

3.2.3 Circular dichroism (CD)

CD measurements were used to verify the helical form of the double helices and examine the effect of added 1 mM CoHex. The NA concentrations were 20 μM. Samples (20 μL) were kept at room temperature in a solution of 100 mM NaCl and 1 mM pH 7.0 Na-MOPS buffer. Since NA condensation is modulated by the amount of monovalent ions present in solution [272, 130], we chose higher NaCl concentration to prevent the spontaneous condensation observed in Figure 3.1. These conditions complement the monovalent ion concentrations used for X-ray scattering experiments reported in [147]. Measurements were made on a BioLogic MOS 450 configured in CD mode. 5 scans per spectrum, each from 200-320 nm with 5s/nm steps, were used for averaging.

3.2.4 All-atom molecular dynamics (MD) simulations

All MD simulations of the 25 bp DNA, RNA and hybrid duplexes were carried out using AMBER12 [42] and ff99bsc0 force field [47, 194]. The 25 bp B′-form structure of poly(dA·dT) duplex was constructed from the helical parameters of 1PLY (PDB) X-ray structure [44] using 3D-DART server [264] and 3DNA software [152]. P-O3' bonds were optimized by energy minimization keeping atoms P,OP1,OP2 and N* restrained (100 kcal/mol/Å^2 force constant). The 25 bp mixed sequence DNA and RNA duplexes were constructed in canonical B- and A-form, respectively, using Nucleic Acid Builder (NAB) [158]. The 25 bp mixed sequence DNA:RNA hybrid was constructed with the help of MD simulations. Two structures corresponding to canonical B- and A-forms were initially built using NAB-based "make-na" server (http://structure.usc.edu/make-na/) by J. Stroud. Each structure was solvated with 16838 TIP3P water molecules, 72 Na^+ and 24 Cl^- ions, and equilibrated, unrestrained, for 200 ns at 300 K. Both hybrid structures converged to essentially the same A-like structure. The final structure from the A-form trajectory was used for simulations of the DNA:RNA hybrid with CoHex. To approximate the experimental NA condensation conditions of a very low Na^+ concentration and to avoid uncertainties associated with Na^+ force field parameters [280], all of the following MD simulations, which involved CoHex, were carried out without any mobile ions other than CoHex. This approximation is justified because the experimental Na^+ concentration is far lower than the concentrations necessary to suppress CoHex induced DNA condensation at mM levels of CoHex [272] used here, and because a negligible effect of Na^+ on the bound CoHex ions was observed at Na^+ concentrations below 40 mM [34]. For the simulations, each of four NA duplexes was solvated with 16880 TIP3P water molecules and 16 CoHex ions (neutralizing amount). After initial water minimization (2000 steps), the systems were equilibrated for 0.5 ns in canonical NVT ensemble and 0.5 ns in isothermal-isobaric (NPT) ensemble using 1 fs time step and achieving 1 atm pressure and 300 K temperature. The latter was maintained using Langevin dynamics with the collision frequency of 1 ps^{-1}. Periodic boundary conditions and the particle mesh Ewald (PME) method were used. During the minimization and equilibration, all NA atoms were

harmonically restrained with 100 kcal/mol/Å^2 force constant. Then, using NVT ensemble and 2 fs time step, 320 ns production trajectories were generated for each system. For DNA atoms, the restrain force constant was reduced to 50 kcal/mol/Å^2 in the production stage. The RNA and DNA:RNA hybrid duplexes were simulated unrestrained. The need for ∼300 ns trajectories was dictated by the long time scale fluctuations (50-100ns) in the number of bound CoHex ions in the unrestrained systems (see Section 3.7). CoHex distributions are presented as the average numbers of CoHex ions in consecutive thin (0.25 Å) cylindrical layers around duplexes. For these calculations, the first 40 ns of each trajectory were not taken into account. Robustness of the CoHex distributions to the choice of water model was confirmed, see Section 3.8.

3.2.5 Continuum electrostatics calculations

Electrostatic potentials and fields around the 25 bp mixed sequence B-DNA and A-RNA duplexes were calculated in continuum solvent approximation via the Poisson-Boltzmann (PB) equation with MEAD [17] 2.2.9 PB solver. Representative configurations of the duplexes with 15 bound CoHex ions were used. Grid spacing of 0.3 Å, dielectric constants of 4 (solute) and 80 (water), and solvent probe radius of 3 Å (CoHex radius) were employed. The potentials and fields were visualized using GEM software [90].

3.3 Why RNA resists condensation

3.3.1 Experimental measurement of duplex condensation propensity

The propensity for condensation, which is accompanied by precipitation of NA duplexes, was probed by measuring the fraction of duplexes remaining in solution after the addition of CoHex. We examined 25 base-pair mixed-sequence duplexes of RNA and DNA, a mixed sequence DNA:RNA hybrid, as well as a DNA homopolymer made of poly(dA):poly(dT). As in Li *et al.* [147], this fraction was determined by measuring the change in UV absorption of the supernatant. Results, shown in Figure 3.1, reveal marked differences in the CoHex induced condensation of the four different NA constructs.

A striking difference in condensation is measured in the RNA-containing structures (dsRNA and DNA:RNA hybrid) relative to the two DNA structures. Homopolymeric DNA begins to condense at the lowest added CoHex levels, followed by mixed-sequence DNA. The RNA-containing duplexes both require more added CoHex to precipitate. The DNA duplexes are suggested to be B-like, while the latter two are A-like [210, 214, 3, 73, 184]. These results support our initial suggestion [147] that NA condensation propensities are related to NA duplex geometry. However, Figure 3.1 shows that this simplistic interpretation is

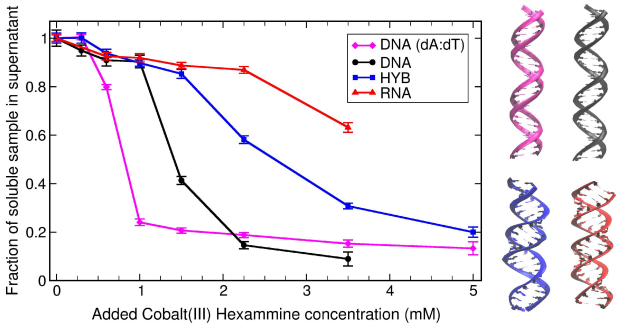

Figure 3.1: Fraction of unprecipitated short 25 bp nucleic acid duplexes (DNA homopolymer, mixed sequence DNA, RNA, and DNA:RNA hybrid) calculated from UV absorption as a function of CoHex concentration in solution (starting duplex concentration is 40 μM). The condensation propensity of each duplex is characterized by the CoHex concentration at the midpoint of duplex condensation – the higher the value the lower the propensity. The right panel illustrates the over-all suggested structures of these duplexes as either B-form or A-like form helices. The color matches the lines in the panel on the left.

not completely correct. Substantial differences in condensation are measured within each structural family. Of the two A-like structures, pure RNA requires more added CoHex to precipitate than does the DNA:RNA hybrid with the same sequence. Differences in B-type helices are detected as well: less added CoHex is required to condense homopolymeric poly(dA):poly(dT) than mixed sequence DNA.

3.3.2 NA duplexes that condense differently can have the same helical structure

To gain insight into the potential connection between duplex structure and condensation, we used Circular Dichroism (CD) spectroscopy to characterize each of the four NA duplexes. To account for changes induced by CoHex-NA binding, spectra were measured for each sample in the absence and presence of CoHex ions. CD provides unique experimental signatures of A- and B-form NA helices. In general, B-DNA has CD peaks at 220 and 275 nm and a dip at 245 nm [30], while A-form RNA has a peak at 260 nm and a dip at 210 nm [210].

Figure 3.2 shows the CD spectra of our four NA duplexes. Not surprisingly, both DNA helices display B-like features. The spectrum of the homopolymeric poly(dA):poly(dT) duplex (Figure 3.2A) strongly resembles a B′-form [3] of DNA, with peaks at 217, 258 and 282nm, and a dip at 247nm [95]. The mixed sequence DNA spectrum (Figure 3.2B) is typical of canonical B-form [137]. No notable spectral shifts in the CD spectra of both DNA duplexes are observed after the addition of CoHex suggesting the helical forms of our 25 bp DNA

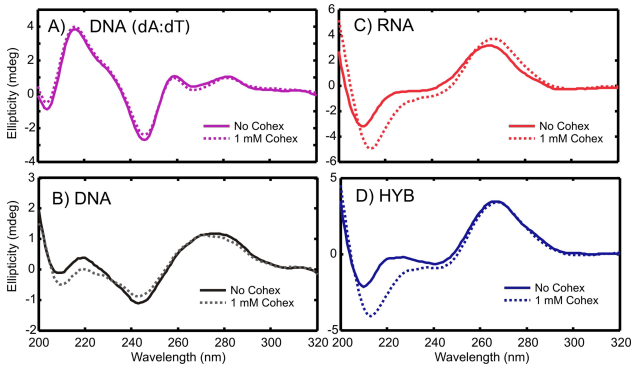

Figure 3.2: CD spectra of the nucleic acid helices with and without CoHex

structures are unaffected by the addition of CoHex. The homopolymer remains in B′-form and the mixed sequence DNA in the canonical B-form.

The spectra of the RNA and DNA:RNA hybrid duplexes without CoHex (Figure 3.2C,D) differ slightly from each other, but closely resemble A-form RNA [210]. Upon addition of CoHex, both spectra slightly change. The differences between the CD spectra with and without CoHex are significant and above the experimental noise. With CoHex, the RNA and the hybrid become virtually identical. Therefore, since in the presence of CoHex both RNA containing duplexes are in practically the same A-like form, the observed differences in CoHex induced condensation between DNA:RNA hybrid and RNA shown in Figure 3.1 can not be explained solely based on differences in the helical structure parameters, as suggested in Ref. [129].

3.3.3 Experiment guides atomistic modeling

The experimental CD spectra (Figure 3.2) indicate a substantial difference in the response of RNA and DNA structures to the addition of CoHex ions. The RNA and hybrid duplexes change their helical geometries while retaining the major features of A-form structure in the presence of CoHex. On the other hand, the structures of the DNA duplexes remain mostly unchanged. Thus, in our atomistic simulations we assume rigid B′- and B-form DNA structures, as enforced by applying positional restraints to all of the DNA atoms. This assumption of rigidity minimizes possible bias due to the use of imperfect modern force-fields [159]. In contrast, to reflect the measured CoHex induced changes in both RNA and the hybrid structures, these molecules are simulated without any positional restraints. The MD simulations were performed to provide sufficient equilibration and sampling of bound configurations of 16 (neutralizing amount) CoHex ions in the simulation box, see Section 3.2.4. After addition of CoHex, the simulations report that both the hybrid and the RNA structures relax to similar A-like form helices consistent with the CD experimental data.

3.3.4 NA structure strongly determines differences in CoHex binding and distribution

Sub-microsecond all-atom MD simulations (see Section 3.2.4) reveal a detailed picture of CoHex distributions around the four 25 bp duplexes used in the condensation experiments, see Figure 3.3. In all four systems, CoHex ions bind preferentially to the phosphate groups whose centers are about 10 Å from the helix axis in all structures. However, the structural differences between the B- and A-form of NA, including differences in the width of the major grooves and in the relative orientation of the phosphate oxygens, lead to striking differences in CoHex binding, Figure 3.3 and Section 3.5. We identify two dominant spatial modes of CoHex binding: an *internal* mode for ions bound inside the major groove, 7 to 12 Å from the helical axis, and an *external* mode for ions bound at the external surface of the phosphate groups exposed to the bulk, 12 to 16 Å from the helical axis. As illustrated in the inset of Figure 3.3, the vast majority of CoHex ions in the *external* mode are outside the DNA major groove. The ions are mobile and mostly found "hovering over" the phosphate backbone and the minor groove, see Section 3.5 for details. The cylindrical shells in the CoHex distributions, which correspond to these modes, are color-coded green and light blue in Figure 3.3. The upper boundary of the *external* binding shell roughly corresponds to the largest distance (from the axis) at which CoHex can still make direct contacts with the phosphate groups. Most CoHex ions are found within these two shells. A small fraction of ions are found deeply buried inside the DNA major groove: about two CoHex ions are bound to the nucleotide bases at Guanine-phosphate-Cytosine (GpC) steps in DNA [35] and GpG steps in RNA and hybrid [122], see Section 3.5. The existence of several CoHex binding modes in the CoHex-DNA interaction is in agreement with the heterogeneity in the magnetic environments sampled by CoHex (^{59}Co NMR) adjacent to DNA [34, 35].

For the RNA and hybrid, the simulations find most of the bound CoHex ions reside in the *internal* shell. These ions are buried within the major grooves with distribution peaks at 10 Å from the helical axis (see Figure 3.3 and Section 3.5). In contrast, CoHex binds to both DNA structures by preferring the *external* binding mode, with the distribution peaks around 14 Å.

The major *external–internal* dichotomy in the preference for binding of CoHex counterions to B- and A-form NA is consistent with the major differences in the distribution of electrostatic potential around the NA surface[52] accessible to large CoHex ions, Figure 3.4. In the B-form geometry (DNA), the strength of the potential in the *internal* shell (major groove) is similar to that in the *external* shell. However, the *external* binding mode is energetically preferred: CoHex ions bound in the *external* shell are, on average, further apart and more exposed to solvent compared to internally bound CoHex, which are closer to the helical axis and therefore experience stronger ion-ion repulsion. In contrast, in the A-form geometry found in RNA and the hybrid structure, the negative electrostatic potential in the *internal* shell (major groove) is at least 10 kcal/mol/$|e|$ stronger than in the *external* shell, Figure 3.4. The resulting increase in the relative affinity of the *internal* shell to CoHex is sufficient

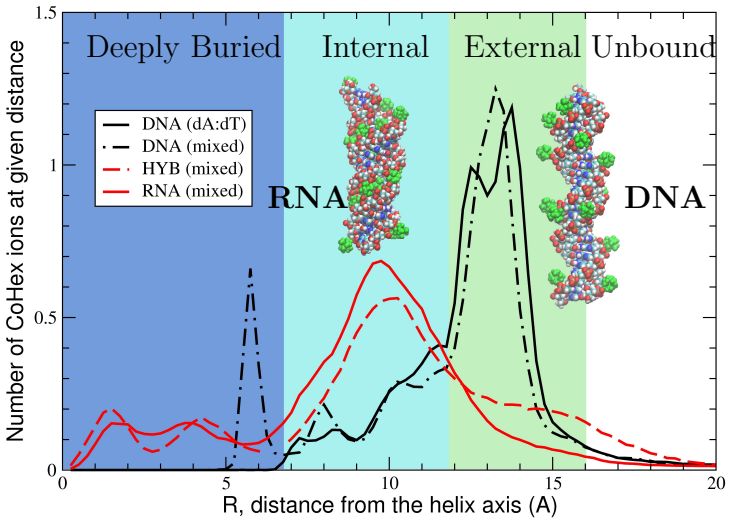

Figure 3.3: CoHex distributions and ion binding modes (shells) around four types of NA duplexes. DNA duplexes exhibit mostly external CoHex binding (12-16 Å). For RNA and DNA:RNA hybrid, most neutralizing CoHex ions are bound internally (7-12 Å). RNA (left) and DNA (right) structures with bound CoHex ions (green). Shown are the average numbers of CoHex ions in thin (0.25 Å) cylindrical layers around duplexes, at the given distance R from the helical axis. Insets show representative snapshots of RNA (left) and DNA (right) structures with bound CoHex ions (green). In DNA, 80 to 100% of bound ions are localized at the surface of the phosphate backbone.

to overwhelm the additional ion-ion repulsion, which explains why most of the CoHex ion bind internally in A-form NA, Figure 3.3.

3.3.5 The connection between the counterion distribution around NA and condensation

As CoHex counterions bind to free NA, the resulting charge neutralization patterns for A- and B-form duplexes are expected to be very different, consistent with the differences in the ion binding patterns. Since it is ultimately attraction between opposite charges that, under right conditions, leads to CoHex induced NA condensation, we expect the differences in bound counterion distributions seen in Figure 3.3 to be directly relevant to the observed differences in the condensation behavior. Table 3.1 summarizes the results of our simulations and the experimental condensation propensities, Figure 3.1, and offers insight into the connection between counterion distribution and condensation. For each of the four duplexes, we list the number of CoHex ions in the *external*, *internal*, and deeply buried binding shells; the degree of duplex neutralization afforded by the bound ions; and the relative condensation

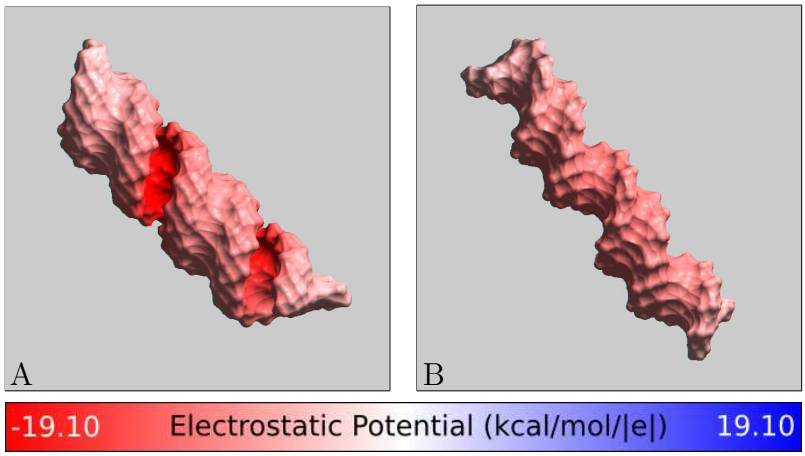

Figure 3.4: Electrostatic potential at the "CoHex-accessible" surface of (**A**) A-form RNA and (**B**) B-form DNA structures, without bound CoHex ions. Shown is the electrostatic potential computed 3 Å away (CoHex radius) from the molecular surface. B-DNA minor groove is sterically inaccessible to large CoHex ions.

propensities determined from the experimental data shown in Figure 3.1).

Table 3.1: Relative measured condensation propensities from Figure 3.1 and average numbers of bound CoHex ions in each binding shell shown in Figure 3.3. 100% neutralization would correspond to 16 bound CoHex counterions

	DNA (dA:dT)	DNA	HYB	RNA
Condensation propensity	**highest**	**high**	**low**	**lowest**
External shell ions	**9.8**	**8.6**	**3.4**	**2.0**
Internal shell ions	4.6	4.2	8.0	9.4
Deeply buried ions	0.1	1.6	2.7	3.3
Duplex neutralization	91%	90%	88%	92%

The number of ions in the *external* shell clearly emerges as the key "order parameter" for these systems: it correlates well with the propensity of the four duplexes to condense. In other words, the greater the number of ions in the *external* shell of the duplex, the more readily the duplex condenses.

To explain the unique role that the distribution of bound CoHex ions relative to the helical axis plays in determining the major differences in condensation propensity between A- and B-form duplexes, we compare the charge neutralization patterns of mixed sequence B-form DNA and A-form RNA helices, Figure 3.5. To characterize these patterns we calculate the electric field at the external surface of NA-CoHex complexes (the connection between distribution of electric field strength in space and the system's electrostatic energy ($\propto \epsilon\,(\nabla\phi)^2$)

[111] motivates this type of analysis). The different preferable CoHex binding modes for B- and A-form duplexes create distinctive paterns of the net electric field at the duplex surface.

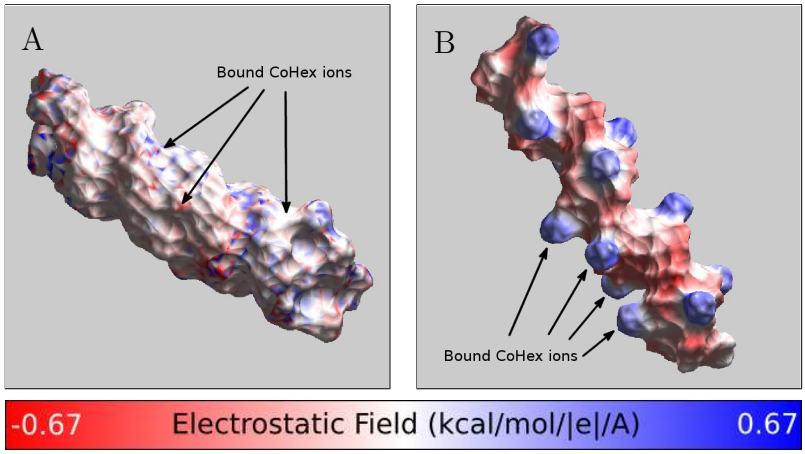

Figure 3.5: Charge neutralization patterns of NA duplexes by bound CoHex ions, assessed by the strength of the electric field near the NA-CoHex complex surface. **A**: A-form mixed sequence RNA with CoHex counterions, which bind mostly in the major groove. **B**: B-form mixed sequence DNA with CoHex ions, which are bound mostly externally. The specific snapshots are chosen to illustrate the *internal* (A) and *external* (B) binding modes from Figure 3.3 and reflect the actual average binding preferences; each snapshot has 15 bound (near neutralizing) CoHex ions, and is taken from the corresponding 320 ns-long all-atom MD simulation described in Section 3.2. See Section 3.5 for a detailed visual characterization of CoHex ion distributions around these structures. The field is computed 3 Å away from the NA-CoHex complex molecular surface.

At ∼90% neutralization of the duplex charge, which is a precursor for NA-NA association [34, 197], there is a substantial difference in the field patterns around B-DNA and A-RNA, Figure 3.5. The strong, localized field of the mobile and correlated 3+ counterions bound mainly at the surface of the phosphate backbone of the B-form duplex is not compensated locally by the more uniform background field of the duplex, thus resulting in a strong alternating pattern of electrostatic field along the surface of the phosphate backbones of the B-DNA, Figure 3.5B. This mobile alternating pattern makes it possible, by an appropriate mutual arrangement [127] of the duplexes in the NA-NA complex, to decrease the electrostatic energy substantially and thus create the NA-NA attraction necessary for the condensation [211]. In contrast, in the A-form geometry, found in RNA and the hybrid structure, CoHex ions bind mostly internally, in the major groove, where the potential is initially much stronger than at the outside surface. The strong, highly correlated CoHex binding leads to a more uniform suppression of the electric field and its smaller spatial variation around the RNA-CoHex complex (Figure 3.5A). The net result for the A-form structures are nearly uniformly charged duplexes: bringing together uniformly charged cylinders cannot lower the

electrostatic energy of the system, which would be necessary to generate an effective attraction. Implicit in the above reasoning is the assumption that no drastic re-distribution of CoHex ions bound to individual duplexes occurs when the duplexes approach each other at distances relevant for condensation. This approximation is borne out by single molecule DNA condensation experiments at mM CoHex concentrations [256], and by our simulations, see Section 3.6.

We can now rationalize the more subtle structure-condensation relationships within B-family helices (mixed sequence vs. homopolymeric DNA) and A-like form helices (RNA vs. DNA:RNA hybrid), Table 3.1. Changing numbers of *external* shell ions can explain differences in condensation of the homopolymeric B′-DNA and the canonical B-form mixed sequence DNA. To trace the origin of the change, we performed another MD simulation of the homopolymer with the structure restrained in the canonical B-form conformation rather than the B′-form observed via CD and used in the simulations. This additional simulation shows the same 91% duplex neutralization by all bound CoHex ions with approximately the same number of ions bound in the *external* shell (9.4 as compared to 9.8 ions in B′-form). Similarly, no CoHex ions are found closer than 7 Å to the helical axis, compared to 1.6 ions bound to GpC steps in the mixed sequence DNA duplex.

Thus, it is not the minor structural difference between B′ and canonical B-form that accounts for the observed differences in condensation. Instead, specific sequence details; e.g., the presence of GpC steps, are responsible for the smaller number of bound CoHex ions at the external DNA surface. The GpC steps, which serve as additional binding sites in mixed sequence DNA [35], bind about two CoHex ions (see Table 3.1) reducing CoHex binding to the phosphate groups in mostly the *external* shell which is key to condensation.

To address the differences between the pure RNA and the DNA:RNA hybrid A-like form structures, we analyzed the duplex configurations from the MD trajectories and found very similar distributions of the phosphorus-phosphorus (P-P) distances across the major groove, but a wider spread of the phosphate group oxygen orientations in the DNA strand of the DNA:RNA hybrid compared to the RNA strands (not shown). The less restrictive sugar-phosphate backbone of the DNA strand [210] allows orientations of the phosphate groups where unbridging oxygens are more directed toward the outside of the duplex. We suggest that such orientations are less favorable for CoHex binding inside the major groove of the hybrid, and promote the binding in the *external* shell, increasing the number of CoHex ions at the hybrid duplex external surface.

Thus, two A-like structures, having practically the same helical geometry but different distributions of the phosphate group orientations, can have different bound counterion distributions with respect to the helical axis, resulting in different condensation propensities.

3.4 Bridging theory and experiment

Our analysis suggests that the spatial distribution of ions around a nucleic acid double-helix has a dramatic effect on its ability to interact with neighboring helices. Specifically, the distribution of bound CoHex ions relative to the helical axis is the key "order parameter": at near neutralizing conditions necessary for condensation it is the number of externally bound ions, rather than net bound, that predicts the experimentally observed condensation propensities of the short NA duplexes. We have rationalized the connection between location of the bound ions and condensation propensity by general electrostatic arguments that explain why DNA condenses more readily than RNA.

The following argument further supports the connection between the counterion distribution around NA and condensation induced by multivalent counterions. Consider two parallel NA duplexes, with externally bound CoHex ions, approaching each other. An ion bound to duplex 1 experiences attractive net force from the nearly neutralized opposite duplex 2; this force becomes substantial when the ion on duplex 1 finds itself in one of the binding shells of duplex 2, *i.e.* when the shells overlap. This increase of the ion attraction in the shell overlapping region implies both nonspecific binding of the ions to the NA surface and a correlation between the bound ions.

The CoHex binding shells, Figure 3.3, for adjacent parallel duplexes can overlap in two different patterns, as depicted in Figure 3.6: *external–external* and *external–internal*. In the former, the axial separation is about 28 Å, while the overlap of *external* and *internal* shells results in smaller interaxial distances, 24-26 Å; the specific pattern is determined by the axial separation. Due to steric restrictions, overlap of the *internal–internal* shells is insignificant and can be ignored.

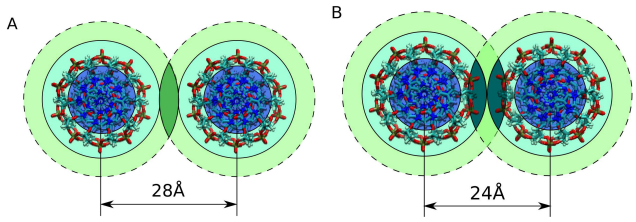

Figure 3.6: Schematic of (A) the *external–external* and (B) *external–internal* CoHex shell overlaps at different interaxial distances. The shell colors correspond to Figure 3.3. The overlapping shell regions are indicated by a darker color to guide the eye to the differences between (A) and (B).

Remarkably, the experimentally reported interaxial distances between CoHex condensed DNA molecules are 27-28 Å [206, 202]. At these distances, the only significantly overlapping shells are the *external* ones, providing additional support for our proposal linking externally

bound ions to condensation.

Finally we comment on experimental observation that RNA eventually begins to condense at higher CoHex concentrations [147]. While the precise spatial arrangement of RNA duplexes in CoHex-induced aggregates is unknown, it is difficult to imagine the absence of close side-by-side helix packing in such aggregates. If RNA packing is similar to that in DNA aggregates, we predict that RNA condensation should occur at interhelical distances smaller than those found in DNA aggregates. This is because at 27-28 Å interaxial distance, only the nearly empty RNA *external* shells overlap, while at shorter distances the attractive contribution can originate from counterions in the populated *external–internal* shell overlap, Figure 3.6B. Such a decrease in the duplex-duplex separation is likely accompanied by both the entropy loss and appreciable increase in the residual electrostatic repulsion. Note that end-to-end stacking of the RNA duplexes in the presence of CoHex was observed in the X-ray scattering experiments [147]. This end-to-end RNA duplex interaction may result from favorable base-stacking interactions between charge compensated molecules [193] and may potentially be promoted by the solvent exposed bound CoHex ions at the ends of the RNA duplexes seen in our simulations, see Section 3.5.

3.5 Details of CoHex ion distribution around DNA and RNA duplexes

In addition to the radial distributions of the bound CoHex ions around NA molecules, shown in Figure 3 in Section 3.3, we present here the spatial distributions of CoHex around the DNA and RNA duplexes.

In the DNA, Figure 3.7, most of CoHex ions bound to phosphate groups are in the external ion binding shell (12 – 16 Å from the helical axis) "hovering over" the phosphate backbone and the minor groove. Only a small fraction ($\sim 1/3$) of the bound ions are in the major groove with almost all of them being in the internal ion binding shell (7 – 12 Å from the helical axis). CoHex density is substantially reduced near the terminal phosphate groups of each DNA strand.

In the RNA, Figure 3.8, CoHex ions bind almost exclusively inside the major groove. Only a tiny fraction of ions bind to the surface of the RNA minor groove or to the external surface of the sugar-phosphate backbone. Note the exposed bound CoHex ions at both ends of the duplex where the major grooves end. These ions may participate in end-to-end attraction of the RNA duplexes that may result in creation of longer filaments.

Sequence specificity has some effect on the distribution of the bound CoHex ions. For example, the mutual configuration of Guanine base oxygens in the GpC steps leads to a

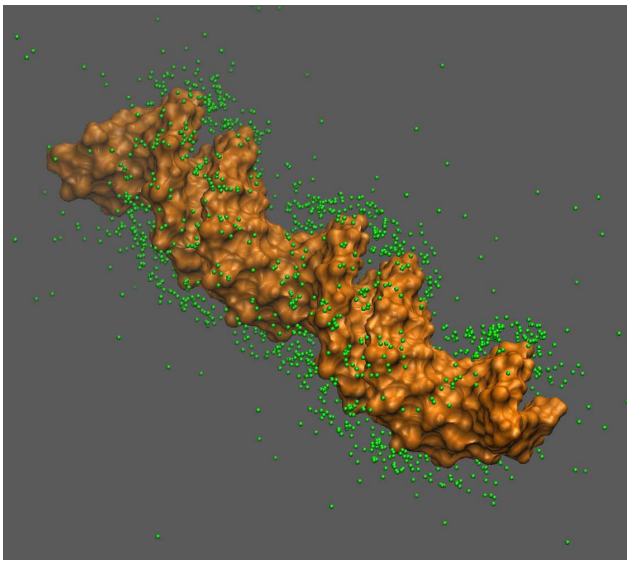

Figure 3.7: Spatial distribution of CoHex ions around homopolymeric 25 bp poly(dA):poly(dT) DNA. Green dots represent positions of CoHex ions per nanosecond sampled over a 100-ns region of the MD trajectory.

relatively high, localized CoHex affinity in the major groove, Figure 3.9.

3.6 Insignificant re-distribution of bound CoHex ions between the bound ion shells around NA upon formation of a pair of interacting duplexes

The discussion in Section 3.3 assumed that no substantial re-distribution of bound CoHex between external and internal ion shells around a single NA molecule occurs when the two duplexes approach each other. Here we verify that an approaching RNA duplex does not "pull" a substantial number of CoHex ions from the internal shell of the other duplex.

For RNA, this verification is presented in Figure 3.10, where we have compared the sum of independent simulated CoHex distributions around single RNA duplex with the CoHex distribution around a pair of such duplexes. Without loss of generality, the RNA molecules

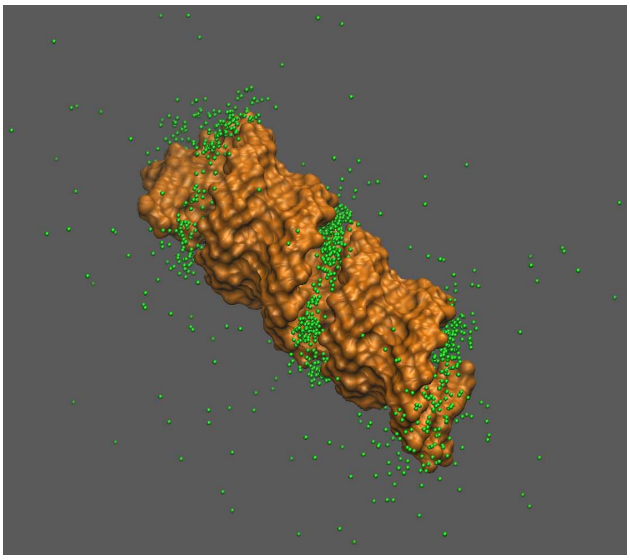

Figure 3.8: Spatial distribution of CoHex ions around 25 bp mixed sequence RNA. Green dots represent positions of CoHex ions per nanosecond sampled over a 100-ns region of the MD trajectory. Most of CoHex ions are bound inside the major groove of RNA within 12 Å from the helix axis.

were restrained to canonical A-form. In the case of the pair of duplexes, the molecules were kept at a 26 Å distance between their helical axes. This separation allows the tails of CoHex distribution formed by the externally bound CoHex ions to overlap. As seen from Figure 3.10, the presence of the second RNA duplex has an insignificant effect on the distribution of strongly bound CoHex within 12 Å from the axis of the first duplex. In fact, the small differences between two curves in Figure 3.10 average out almost completely upon integration that yields the total numbers of bound ions: 13.5 ions in both cases. The difference in the ion densities beyond 12 Å is due to a small re-distribution of CoHex in the external shells of two interacting duplexes.

Figure 3.11 presents evidence for the absence of substantial re-distribution of bound CoHex between external and internal ion shells around DNA duplexes when they approach each other. The sum of independent simulated CoHex distributions around a single DNA (poly(dA):poly(dT)) duplex is compared with the CoHex distribution around a pair of such duplexes. As in the case of RNA, two DNA molecules were kept at a 26 Å distance between their helical axes. The numbers of CoHex ions in the internal shell of a single duplex differs from the corresponding number in the case of a pair of interacting duplexes by less than 0.2 ions. A somewhat greater difference, 1.1 ions, is found between the external shells of the duplexes, in the case of the interacting duplex pair compared to the case of non-interacting

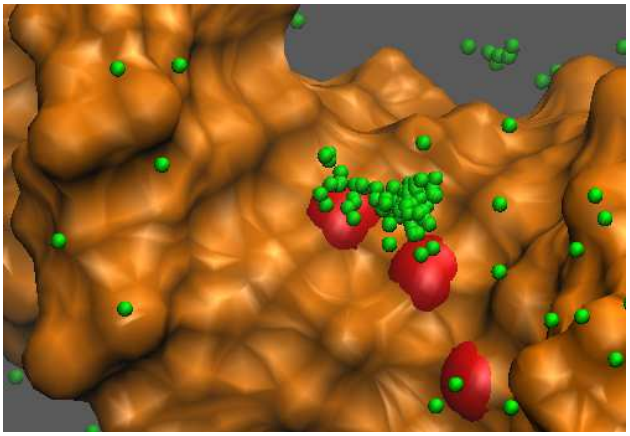

Figure 3.9: Deeply buried CoHex ion bound to a GpC step in the mixed sequence DNA. Zoomed-in view of the major groove. Green dots represent positions of the ion sampled every 1 ns over 100 ns part of the MD trajectory. Red spheres represent Guanine base oxygens.

DNA duplexes. Since the number of all bound CoHex ions around the DNA pair differs insignificantly from the corresponding number of ions around the two non-interacting DNA duplexes, the difference of 1.1 ions is suggested to be due to a re-distribution of the bound CoHex ions only within the external shells of the interacting DNA duplexes. We suggest that such a re-distribution will be substantially reduced in the case of hexagonally packed DNA aggregates when the six external shell overlapping regions are formed around each DNA duplex.

3.7 Hundreds of nanoseconds of all-atom MD are needed to adequately represent the fluctuating CoHex atmosphere around RNA duplex

The number of bound ions is expected to fluctuate significantly around NA duplexes. These fluctuations affect the accuracy of the calculated averages of the number of bound CoHex ions presented in Section 3.3. As seen from Figure 3.12, a ∼300 ns averaging window is sufficient to obtain well converged averages. In calculations of CoHex distributions and average numbers of bound CoHex ions in different ion binding shells around the NA duplexes we have ignored the first 40 ns of each MD trajectory to allow for equilibration of the ionic atmosphere.

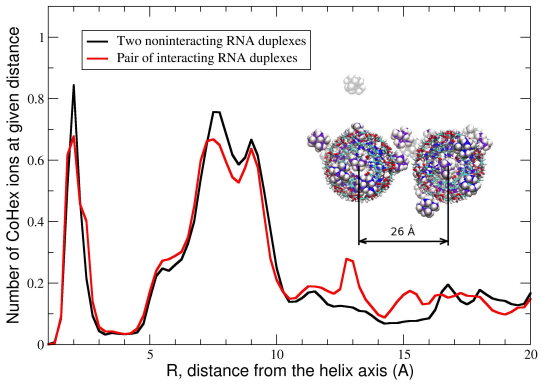

Figure 3.10: Combined CoHex distribution of two independent 25 bp RNA duplexes (black curve) vs. the distribution in a pair of interacting duplexes (red curve) at 26 Å separation. The combined distribution is calculated as a sum of two individual distributions shifted by 26 Å.

3.8 CoHex ion distribution around NA is robust to the choice of water model used in MD simulation

Here we investigate possible dependence of the CoHex ion distribution around a short NA duplex on the choice of the water model used in MD simulations. We have carried out a simulation of the CoHex atmoshere around the homopolymeric 25 bp poly(dA):poly(dT) DNA duplex using the TIP4P-Ew [109] water model instead of the TIP3P [117] model employed in the simulations as described in Section 3.2.4; all other simulation parameters remained the same. The resulting distribution of CoHex counterions is shown in Figure 3.13, where it is compared with the original CoHex distribution obtained using TIP3P water model. The conclusion is that the CoHex distribution is robust to the choice of water model.

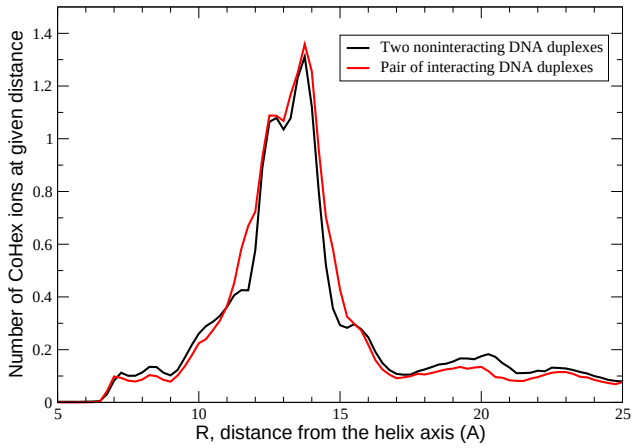

Figure 3.11: Combined CoHex distribution of two independent 25 bp poly(dA):poly(dT) DNA duplexes (black curve) vs. the distribution in a pair of interacting duplexes (red curve) at 26 Å separation. The combined distribution is calculated as a sum of two individual distributions shifted by 26 Å.

3.9 Structural changes in the RNA duplex caused by CoHex binding have relatively small effect on the electrostatic potential around the duplex

As discussed in Section 3.3, the binding of CoHex causes conformational changes in the A-form structures of RNA and DNA:RNA hybrid. To investigate the possible effect of these changes on the electrostatic potential around the RNA duplex, we have repeated the calculation shown in Fig. 3.4, but now for the RNA structure with the CoHex-induced conformational changes, Fig. 3.14. As in the case of RNA before CoHex binding, Fig. 3.4, the strongest potential occurs in the major groove (the maximum value is +21.29 kcal/mol/$|e|$). Just like in the case of canonical RNA structure, the positive electrostatic potential in the internal binding shell (major groove) is at least 10 kcal/mol/$|e|$ stronger than in the external shell. Therefore, the rationale given in Section 3.3 for the preferential internal binding of CoHex in A-like duplexes is robust to CoHex induced conformational changes in the RNA. Given the high similarity between RNA and DNA:RNA hybrid structures in CoHex, we expect the conclusion to hold for the hybrid as well.

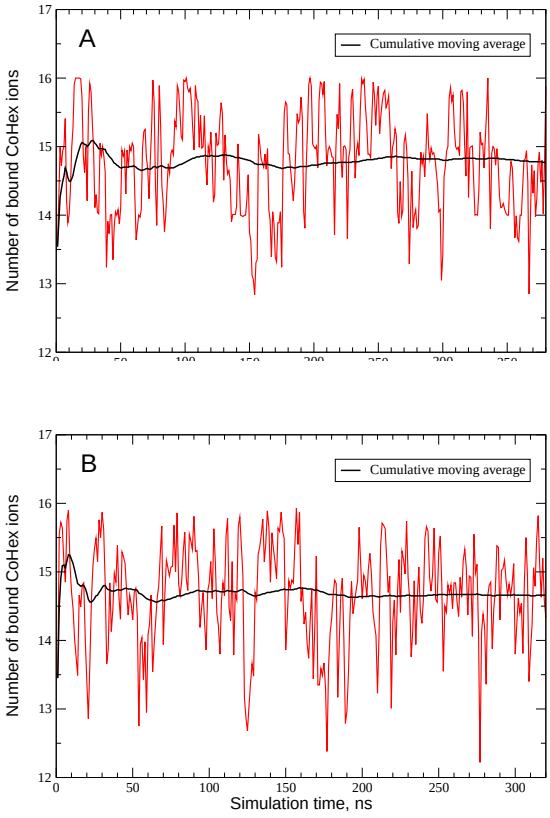

Figure 3.12: Number of bound CoHex ions within 16 Å from the duplex helical axis as a function of simulation time. (A) RNA and (B) poly(dA):poly(dT) DNA. Each point is a time average over 1 ns, cumulative moving averages are shown as well.

3.10 The "shells" model of CoHex distributions around NA duplexes

The all-atom MD simulations of short (25 base pairs) DNA, RNA and DNA:RNA hybrid duplexes with CoHex counterions described in Section 3.2.4 show that, under near neutralization conditions necessary for NA condensation ($\sim 90\%$ of NA charge neutralized by bound counterions), the majority ($\sim 2/3$) of neutralizing CoHex ions bound to B-form DNA are localized at the external surface of the phosphate groups in the cylindrical layer 12 to 16 Å from the helical axis, Fig. 3.15. This layer is defined as the "external" ion binding shell. The smaller fraction of bound CoHex ions around B-DNA ($\sim 1/3$) are localized in the major groove at distances 7 to 12 Å from the helical axis – the corresponding cylindrical layer is defined as the "internal" ion binding shell. In contrast to B-DNA, the majority of CoHex ions ($\sim 2/3$) bound to A-form RNA are localized within the RNA major groove, in the "internal" ion binding shell, Fig. 3.15. This substantial difference in CoHex distributions is explained

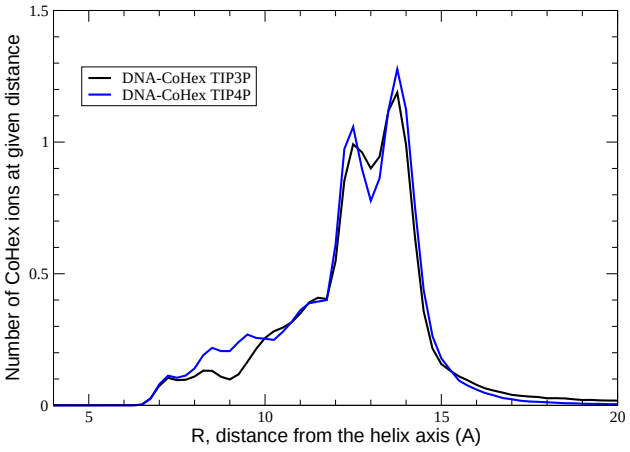

Figure 3.13: Robustness of the computed CoHex distribution around NA to the choice of the water model. Shown are CoHex distributions around 25 bp poly(dA):poly(dT) DNA duplex simulated in TIP3P and TIP4P water models.

by the major difference in the electrostatic potential of B-DNA and A-RNA [52, 257]. The potential in the major groove of A-RNA is about 10 kcal/mol/$|e|$ lower than on the rest of the RNA surface accessible to CoHex, or anywhere on the surface of B-DNA. The resulting much stronger attraction of trivalent CoHex ions overwhelms the ion-ion repulsion, leading to the qualitatively different pictures in CoHex binding.

An analysis of the simulated CoHex ion distributions around 25-bp DNA, RNA and DNA:RNA hybrid duplexes of the equivalent mixed sequence [193] and around homopolymeric poly(dA):poly(dT) DNA led to the following two observations.

First, at the near-neutralizing conditions necessary for NA condensation, the fractions of CoHex ions in the "external" ion binding shells of the four NA duplexes correlate with measured condensation propensities of these duplexes. Second, the "external" CoHex binding shells of NA molecules are the only shells that overlap substantially at the inter-axial duplex separation corresponding to the separations observed in CoHex condensed (aggregated) DNA phases (about 28 Å [206, 202]).

These two observations constitute a basis for the "overlapping ion binding shell" mechanism of NA condensation which was proposed in Section 3.3. According to the mechanism, the fraction of neutralizing multivalent ions bound in the NA "external" ion binding shell (and not the total number of bound ions) is a key parameter for understanding multivalent ion-induced nucleic acid condensation. This fraction reflects a complex interplay between various structural and sequence features of NA helices, and its ions are responsible for most of the attractive interaction between the helices. The proposed mechanism was rationalized by simple and robust electrostatic arguments, and is in excellent qualitative agreement with experimental condensation propensities of various NA duplexes. The mechanism was later

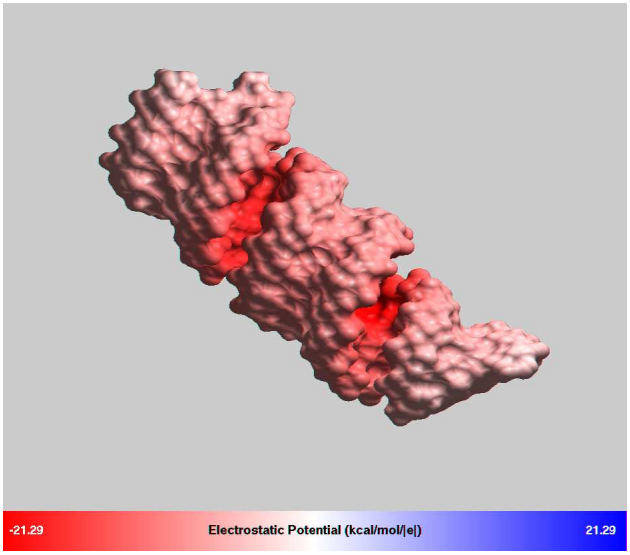

Figure 3.14: Electrostatic potential at the surface of 25 bp RNA duplex after equilibration with CoHex ions in solution. As in Fig. 3.4, CoHex ions were removed prior to calculating the potential. The graphics is by GEM package [90].

illustrated by explicit calculations of the potential of mean force (PMF) between two adjacent NA duplexes [277]. However, no quantitative model of the duplex condensation phenomenon has been presented so far; such a model would be the best illustration for the proposed new mechanism and could lead to novel predictions.

In this section, we develop a semi-quantitative model of CoHex-induced NA duplex condensation that is based on the "overlapping ion binding shell" mechanism. The model provides a quantitative relationship between key characteristics of bound multivalent counterion distributions around NA molecules and the free energy changes upon NA duplex aggregation. We show that the estimated free energy changes for different NA duplexes correlate well with the experimental condensation propensities of these duplexes and explain observed differences in DNA and RNA condensation, and make several predictions.

3.11 The multi-shell model of ion-mediated NA-NA interaction

We consider a transition between a dilute aqueous solution of relatively short nucleic acid duplexes (compared to the DNA persistence length of ∼150 base pairs) and their condensed (aggregated) phase represented by a bundle of parallel, hexagonally packed NA molecules [202] with a distance between neighbors smaller than their length. The solution phase

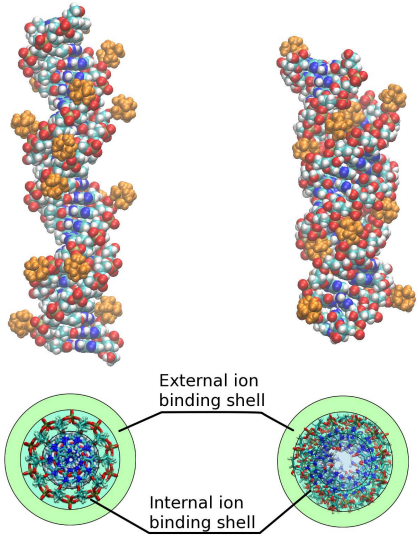

Figure 3.15: The "shells" of counterion distribution around nucleic acids duplexes. Upper panel: representative snapshots of B-form dsDNA (left) and A-form dsRNA (right) structures with bound CoHex ions (orange). Lower panel: the "external" and "internal" cylindrical ion binding shells around B-DNA (left) and A-RNA (right). Most of the ions ($\sim 2/3$) shown in the upper panel are bound in the "external" shell of the B-DNA, and the "internal" shell of the A-RNA.

contains a certain amount of trivalent CoHex counterions sufficient to neutralize all the duplexes. Part of these ions are bound to the duplexes due to Manning-Oosawa condensation [163, 190, 164, 230]. The fraction of NA duplex charge neutralized by bound CoHex ions, Θ, in the solution and aggregated phases is considered to be the same, although the aggregated phase as a whole is assumed to be neutral.

The aggregated phase is stabilized by short-range attractive forces between the duplexes that originate mostly from the electrostatic interactions of multivalent counterions bound to one duplex with the "correlation holes" [230, 146, 181, 182, 145] on the charged surface of another duplex covered by a layer of bound multivalent counterions. I.e., these interactions have a significant non-mean-field component [189, 178, 151, 119] due to correlations between multivalent counterions bound to different neighboring duplexes in the aggregate.

Our goal is to estimate the free energy difference between the aggregated and solution phases of NA duplexes, ΔG_{aggr}, as a function of Θ. The latter can be used as a convenient proxy for bulk CoHex concentration in the solution phase.

The multivalent ion-induced aggregation free energy can be represented as a sum of three

additive components:

$$\Delta G_{aggr} = \Delta G_{attr} + \Delta G_{el-rep} - T\Delta S_{conf} \, . \tag{3.1}$$

The first two terms in Eq. 3.1 describe the changes in the electrostatic interactions upon the transition between the solution and aggregated phases. These interactions are decomposed into the short-range net attractive term, ΔG_{attr}, and the net repulsive term, ΔG_{el-rep} that describes the (residual) repulsion between the duplexes almost neutralized by bound counterions. The last term, $T\Delta S_{conf}$, represents the loss of duplex configurational entropy (translational and rotational) upon the aggregation.

The decomposition of the electrostatic contributions to ΔG_{aggr} in Eq. 3.1 into net attractive and net repulsive terms is not unique. Both terms include contributions from the interactions of multivalent ions bound to one duplex with the bare charges of adjacent duplex, which are favourable for the aggregation, and contributions from the interactions of these same ions with the ions bound to the adjacent duplex, which oppose the aggregation. Due to ion-ion correlations this latter type of the interactions between bound multivalent ions is difficult to estimate. However, grouping these interactions between bound multivalent ions into the contributions where the ion-ion correlations play a substantial role and where they can be neglected allows us to estimate the net attractive ΔG_{attr} and net repulsive ΔG_{el-rep} electrostatic terms through simple analytical expressions.

The last term in Eq. 3.1, $-T\Delta S_{conf}$, can be estimated using a simple coarse-grained approach. This term is usually neglected for long DNA molecules [182], but, as we shall see later, it can contribute appreciably to the destabilization of the aggregated phase in the case of relatively short 25-bp NA duplexes.

Additional underlying assumptions of our model for the aggregation free energy, physical considerations that justify them, and details of how each term in Eq. 3.1 is calculated are presented in Section 3.13. Derivations of the components of Eq. 3.1 are described below.

3.11.1 Short-range attractive component, ΔG_{attr}, of the ion-mediated duplex-duplex interactions

The multivalent ion-mediated short-range attractive forces between the duplexes in the aggregated phase arise from the interactions of the multivalent counterions bound to the surface of one duplex with all the charges on the neighboring duplex (including its bound counterions). These are the same interactions that determine the distribution of counterions around NA duplexes in the solution phase. They are strong when the counterion is inside the duplex ion binding shell and relatively weak when outside [203, 250], and can be characterized by the ion binding affinity to nearly neutralized duplex. In the case of multivalent counterions these interactions can have a significant non-mean-field component [178, 151, 119] due to ion-ion correlations which reduce the ion-ion repulsion.

In our decomposition of the aggregation free energy, Eq. 3.1, we consider that these attractive interactions contribute to ΔG_{attr} when CoHex ion bound to one duplex enters the "external" ion binding shell of another duplex [207], i.e. when the "external" binding shells of the two duplexes overlap upon aggregation. When the bound CoHex ion is outside the ion binding shell of a neighboring duplex we no longer consider its interaction with that duplex contributing to ΔG_{attr}. In this case the bound ion can be treated as part of the averaged neutralizing background [203, 245, 66, 191] that screens the bare duplex charge and reduces the mutual electrostatic repulsion between the duplexes described by ΔG_{el-rep}. The proposed decomposition assumes that the non-mean-field component of the ion interactions with the counterions bound to the neighboring duplex can be significant and accounted for in ΔG_{attr} when the ion enters the "external" ion binding shell of the neighboring duplex but is small and can be neglected when the ion is outside the ion binding shell.

Once the duplexes approach each other upon aggregation, CoHex ions bound in the "external" ion binding shell of one duplex enter the "external" shell of another duplex. The additional interaction energy of each of these ions in the "external" ion binding shell of the neighboring duplex is essentially the binding energy for the additional CoHex ions in the "external" shell of the duplex. Assuming that minimal changes in the CoHex distributions occur when the two duplexes approach each other, this binding energy averaged over the volume the shell can be approximated by CoHex binding affinity, μ_a, to the "external" ion binding shell of an isolated NA duplex. This quantity varies with the degree of duplex neutralization Θ, but for a narrow range of Θ (at near neutralizing conditions, $0.9 \leq \Theta \leq 1.0$) that is of interest to us here, this dependence is small and can be neglected. Ion affinity μ_a reflects the balance between the attraction to the bare NA duplex charges and the repulsion from all other bound ions around the duplex and, therefore, depends on the ion charge, Ze. It also absorbs the contribution from the ion-ion correlations which increases with the ion valency Z.

The above assumptions about the contributions to ΔG_{attr} result in a short-range attractive component of the interactions between the two neighboring NA duplexes that is proportional to two quantities: (1) the average number of CoHex ions in the overlapping region of the ion binding shells of these duplexes, ΔN_s, and (2) the binding affinity μ_a of CoHex ion in the "external" ion binding shell of NA duplex at near neutralizing conditions.

For the hexagonal packing of the duplexes in the aggregated phase [202], the total short-range attractive term in the decomposition of the aggregation free energy, Eq. 3.1, per duplex can therefore be estimated as

$$\Delta G_{attr} = 3\mu_a \times \Delta N_s \,, \tag{3.2}$$

where the factor of 3 accounts for the half of the attractive interactions with the six nearest neighbors in the hexagonally packed aggregate.

To estimate μ_a and ΔN_s we will use the equilibrium properties of CoHex distribution around a single NA duplex in the solution phase obtained from the MD simulation.

CoHex binding affinity μ_a.

The binding affinity of an ion to the ion binding shell of a polymer can be defined as a difference of the excess chemical potentials of this ion in the binding shell and in the bulk. In the case of the "external" ion binding shell of NA duplex where the excluded volume effects for CoHex ions are negligible this difference can be estimated as

$$\mu_a = -k_B T \ln\left(\rho_s/\rho_b\right) , \tag{3.3}$$

where ρ_b and ρ_s are the number densities of CoHex ions in the bulk and in the "external" ion binding shell of the NA duplex, respectively. A more detailed derivation of this equation is presented in Section 3.13.

Because of the finite size of the simulation box used in all-atom MD simulations and the absence of a monovalent salt screening in the simulations, the CoHex concentration at the simulation box boundary, ρ_B, is not equal to the bulk value ρ_b required in Eq. 3.3 for estimating μ_a. To account for the difference, we introduce a long-range correction to μ_a:

$$\mu_a = -k_B T \ln\left(\rho_s/\rho_B\right) + Ze\varphi(r_B) , \tag{3.4}$$

where $Ze\varphi(r_B)$ is the energy of the CoHex ion charge, Ze, in the electrostatic potential of the NA duplex and its bound ions, $\varphi(r_B)$, at the system boundary ($r_B = 31$ Å). Without monovalent ions present, $\varphi(r)$ can be estimated as the electrostatic potential of a uniformly charged non-conducting rod of length H with a linear charge density λ in a solvent with dielectric constant ε,

$$\varphi(r) = \frac{2\lambda}{\varepsilon} \ln\left(\frac{H/2 + \sqrt{(H/2)^2 + r^2}}{r}\right) . \tag{3.5}$$

Here λ corresponds to the linear charge density of 25-bp NA duplex (charge $Q_{na} = -48e$, length H) re-normalized (scaled down) by the charge of all bound CoHex ions ($N_0 Ze$) within the outer boundary of the "external" ion binding shell, $\lambda = (Q_{na} + N_0 Ze)/H = Q_{na}(1 - \Theta_0)/H$, where $\Theta_0 = N_0 Ze/|Q_{na}|$ is a degree of duplex neutralization by all (N_0) bound CoHex ions determined from the results of MD simulations. The potential $\varphi(r)$ is zero at infinity. The values of the potential estimated for the NA duplexes of interest at $r_B = 31$ Å and dielectric constant $\varepsilon = 78.5$ are relatively small, ranging from -0.7 to -1.1 $k_B T/e$.

Number of bound CoHex ions, ΔN_s, in the overlapping volume of ion binding shells.

Fig. 3.16 shows a schematic of the "internal" and "external" ion binding shells around two parallel duplexes at the typical duplex-duplex separation of 28 Å in the DNA aggregate [206, 201, 202]. The shells overlap volume is completely defined by geometries of the shells and mutual orientation of the duplexes. In the hexagonal packing of the duplexes in the

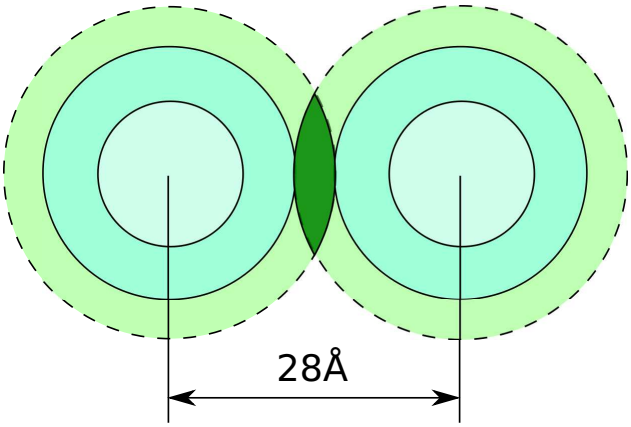

Figure 3.16: Horizontal cross section of the overlap of the "external" CoHex ion binding shells of two NA duplexes at the inter-axial separation $d = 28$ Å. The overlap region (volume) is indicated by dark green color. CoHex ions in this volume element, which is about 80 Å high (length of the NA duplex), strongly interact with both duplexes creating the effective short-range attraction between NA molecules. These ions are excluded from the estimation of the effective duplex-duplex electrostatic repulsion in Eq. 3.10.

aggregated phase [202] considered in our model, the duplexes are parallel to each other and their ends are assumed at the same height.

Under the assumption of minimal changes in the CoHex distributions when the duplexes approach each other the number of ions ΔN_s in the overlap volume of the two "external" ion binding shells ΔV_s can be estimated as

$$\Delta N_s = 2\rho_s \Delta V_s = 2\,\frac{N_s}{V_s}\Delta V_s\,, \tag{3.6}$$

where V_s and N_s are the volume and the average number of CoHex ions bound in the "external" ion binding shell, respectively.

Since we are interested in exploring the dependence of the aggregation free energy ΔG_{aggr} on the degree of NA duplex neutralization Θ at near neutralizing conditions when Θ is close to its value Θ_0 observed in MD simulations, we need to estimate how N_s changes with Θ. Assuming the fractions of the bound CoHex ions in each of the ion binding shell are insensitive to the small changes in the total number of bound CoHex ions, N, we can write

$$N_s = f_s N = f_s \frac{|Q_{na}|}{Ze}\Theta\,, \tag{3.7}$$

where $f_s = N_s^0/N_0$ is the fraction of CoHex ions bound in the "external" shell, N_s^0 and N_0 are the simulation results values.

The above results allow one to rewrite Eq. 3.2 as

$$\Delta G_{attr}(\Theta) = 6\mu_a \frac{\Delta V_s}{V_s} \frac{|Q_{na}|}{Ze} f_s \Theta \,. \tag{3.8}$$

Note that for long duplexes, μ_a is independent of the duplex length H, and so ΔG_{attr} in Eq. 3.8 depends on H only through the bare duplex charge $Q_{na} = \lambda_{na} H$, where λ_{na} is the (constant) linear charge density of the unscreened NA duplex.

3.11.2 Repulsive electrostatic component ΔG_{el-rep} of the ion-mediated duplex-duplex interactions

Our decomposition of the electrostatic interactions between two NA duplexes upon ion-mediated duplex aggregation onto the attractive and repulsive parts is based on the separation of the bound CoHex ions into those which are in the ion binding shell overlap volume ΔV_s, Fig. 3.16, and the rest of the bound ions. The interactions of the former ones with the neighboring duplex and all of its bound counterions have a significant ion-ion correlation contribution and constitute the attractive part of the aggregation energy ΔG_{attr} described above. The interactions of rest of the bound ions with the neighboring duplexes and their bound counterions can be treated at the mean-field level. In this case these counterions can be considered as a neutralizing background which uniformly reduces the bare charge of each duplex. This approximation allows us to use continuum electrostatics in estimating ΔG_{el-rep}.

As discussed above, the NA duplex neutralization by bound CoHex is described by a degree of neutralization Θ which, according to Manning counterion condensation theory [164] applied to NA molecules interacting with trivalent counterions, reaches ~ 0.92 level. Approximately the same level of duplex neutralization by CoHex ions ($\Theta_0 \sim 0.88 - 0.92$) is observed in all-atom MD simulations.

As in the case of ΔG_{attr}, we consider the repulsive interactions between the duplexes at zero monovalent salt condition, neglecting a small monovalent ion screening in the NA duplex condensation experiments [147, 257]. Indeed, for spermine^{4+}, which is similar to CoHex in its condensing power, condensation of 150-bp DNA fragments was shown to be unaffected by monovalent ion concentrations of 20mM and lower if DNA monomer concentration ([P]) is higher than 0.4 mM [205]. In the condensation experiments we discuss [147, 257], NA monomer concentration is about 2 mM.

Additional considerations of why the influence of 20mM of NaCl on CoHex induced duplex condensation can be neglected here include: 1) the interactions of bound CoHex ions with NA molecules are virtually unaffected by monovalent salt concentration below 40 mM [34] and 2) CoHex affinity μ_a to NA duplex is affected by a small amount of monovalent salt through the change of long-range potential in much the same way as the long-range repulsive duplex-duplex interactions. Since each of these two monovalent salt screening effects on $\Delta G_{attr} \sim \mu_a$

and ΔG_{el-rep} are small and opposite, we assume that the combined salt screening effect on the duplex-duplex attractive and repulsive terms is negligibly small and estimate these terms without considering the monovalent screening. However, at physiological monovalent salt concentrations ($\sim$ 100 mM), the difference could increase; the drop in the attraction due to reduction of CoHex affinity may be substantially larger than the drop in the repulsive term, contributing to the onset of NA duplex aggregation by monovalent ions [205].

The above approximations (no ion-ion correlations and zero monovalent salt condition) allow us to estimate the duplex-duplex repulsion as interaction between the neighboring NA duplexes with their charges uniformly re-normalized by bound counterions, with the duplexes being in continuum dielectric without free ions. The interactions beyond the nearest neighbors in the hexagonal packing are neglected (assumed to be screened out by monovalent ions or by unbound CoHex ions in the net neutral duplex aggregated phase).

Within the linear response electrostatics in a continuum dielectric, the interaction between two duplexes with the charges re-normalized uniformly by a factor ξ is equivalent to scaling the interaction between bare duplexes, ΔG_{el-na}, by ξ^2.

To estimate ξ we notice that according to the specific way we decompose the electrostatic interactions, all CoHex ions bound to one NA duplex ($N \sim \Theta$), except those ($\Delta N_s/2$) in the shell overlapping volume, participate in the duplex charge re-normalization. Taking into account Eqs. 3.6 and 3.7, the charge fraction of these $\Delta N_s/2$ ions that has to be excluded from the re-normalization is

$$\Lambda = \frac{\Delta N_s}{2}\frac{Ze}{|Q_{na}|} = \frac{\Delta V_s}{V_s} f_s \Theta \,. \tag{3.9}$$

This fraction reduces the duplex screening and changes the duplex charge re-normalization coefficient from $\xi = 1 - \Theta$ to $\xi = 1 - \Theta + \Lambda$. The resulting estimate for the duplex repulsive contribution to the aggregation free energy at the hexagonal packing of the aggregated phase can be written as

$$\Delta G_{el-rep}(\Theta) = 3\Delta G_{el-na}(1 - \Theta + \Lambda)^2 = 3\Delta G_{el-na}\left(1 - \Theta\left[1 - \frac{\Delta V_s}{V_s} f_s\right]\right)^2 , \tag{3.10}$$

where ΔG_{el-na} is the electrostatic interaction between the two bare duplexes in a dielectric continuum which can be estimated numerically, within the Poisson framework. Here we neglect the interactions with more distant neighbors in the hexagonal packing taking into account that they should be scaled by a much smaller factor ($\xi^2 = (1 - \Theta)^2$ at $\Lambda = 0$) and assuming the overall neutrality of the aggregated phase due to unbound diffuse ions.

Note that the re-normalization coefficient ξ is independent of the duplex length H and so the repulsion term ΔG_{el-rep} in Eq. 3.10 depends on H only through the bare duplexes repulsion ΔG_{el-na}. For long duplexes $\Delta G_{el-na} \sim \lambda_{na}^2 H$ leading to ΔG_{el-rep} linear with H, similar to ΔG_{attr}.

We estimate ΔG_{el-na} by solving the Poisson equation (PE) for the two parallel unscreened duplexes and averaging the interaction energies over 12 different mutual duplex orientations. The details of this estimation are presented in Section 3.13.

3.11.3 Configurational entropy loss upon duplex association

The change in the NA duplex configuration entropy upon duplex aggregation, ΔS_{conf}, is caused by restrictions of the duplex translational and rotational motion in the aggregated phase compared to a free motion in a dilute solution phase. For long NA molecules this entropic contribution to the aggregation free energy is usually neglected compared to other repulsive and attractive contributions [182]. For short 25 base pair NA duplexes, however, the contribution of $T\Delta S_{conf}$ can be substantial and, as we will see, comparable with the electrostatic repulsion between the duplexes in the aggregated phase.

Short NA duplexes of the length smaller than the NA persistence length ($\sim$ 150-bp for dsDNA) can be considered as rigid rods. We estimate their configuration entropy change, $\Delta S_{conf} = \Delta S_{tran} + \Delta S_{rot}$, by following the approach described in Ref. [249],

$$\Delta S_{conf} = k_B \ln(c\Delta X \Delta Y \Delta Z) + k_B \ln(\Delta X \Delta Y / \pi H^2)\,. \tag{3.11}$$

Here, the first term corresponds to the change in the duplex translational entropy S_{tran} due to reduction of volume available for translational motion of the duplex center of mass from $1/c$ in the solution phase to $\Delta X \Delta Y \Delta Z$ in the aggregated phase, where c is a NA duplex concentration in the solution. The second term reflects the loss of the rotational entropy ΔS_{rot} in two of the three duplex rotational degrees of freedom upon aggregation; the corresponding reduction of the available rotational phase space being $\Delta\theta_1 \Delta\theta_2 / 4\pi$ where $\Delta\theta_{1(2)} \approx \Delta X(Y)/(H/2)$, Fig. 3.17. In our estimations we use $\Delta X = \Delta Y = (d - 2a)/2$ where d is inter-axial duplex separation and a is the radius of the NA duplex estimated as $a = 11$ Å. At $d = 28$ Å corresponding to duplex separation observed in CoHex condensed DNA phases [206, 202], $\Delta X = \Delta Y = 3$ Å – a half-width of the nucleic acid effective attraction energy well [129] at $5k_BT$ above the minimum. For simplicity we use the same value for ΔZ ($\Delta Z = 3$ Å), which describes the vertical misalignment of the ends of parallel duplexes in the hexagonal bundle.

From Eq. 3.11 it follows that only ΔS_{rot} part of the duplex configuration entropy change depends on the duplex length H – this part scales logarithmically with H, increasing its contribution to the aggregation energy for longer duplexes.

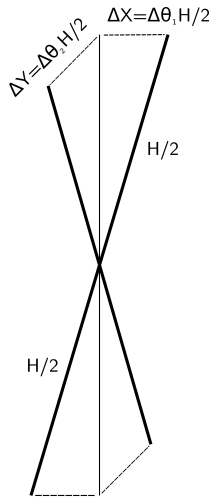

Figure 3.17: Schematic of a rigid rod rotation about the two axes used in the estimate of the rotational entropy change upon aggregation.

3.11.4 Attractive and repulsive interactions at the "internal-external" ion binding shell overlaps

So far, we considered the simplest case of two parallel NA duplexes at such separations in the aggregated phase that only their "external" ion binding shells overlap; we approximated the additional binding free energy for the CoHex ion bound in the "external" ion binding shell of one duplex and entering the "external" ion binding shell of the neighboring duplex by the CoHex binding affinity μ_a to the "external" shell. This affinity, defined as a difference of the CoHex excess chemical potentials, Eq. 3.16, reflects only the change of CoHex ion interactions with its environment and does not depend on CoHex configurational entropy. Since this entropy in either of the two "external" ion binding shells of the two neighboring duplexes is same, the above approximation is reasonable.

If the duplex-duplex separations in the aggregated phase are smaller than 28 Å, in addition to the "external-external" shell overlaps the overlaps of the "external" and "internal" ion binding shells of the adjacent duplexes occur, Fig. 3.18. Below we will argue that the attractive contribution ΔG_{attr} to the aggregation energy can still be estimated by Eq. 3.2 where the quantity ΔN_s is replaced by ΔN, the sum of the numbers of bound CoHex ions in both overlaps.

In equilibrium, the chemical potential of CoHex ions in the "external" and "internal" ion binding shells is the same. Any difference in the excess chemical potentials of the ions in these shells (and, thus, in the ion binding affinities) is compensated by the difference in ion configurational entropies within these shell. The same consideration can be applied to the estimation of the free energy gain for the CoHex ion bound in the "external" shell of

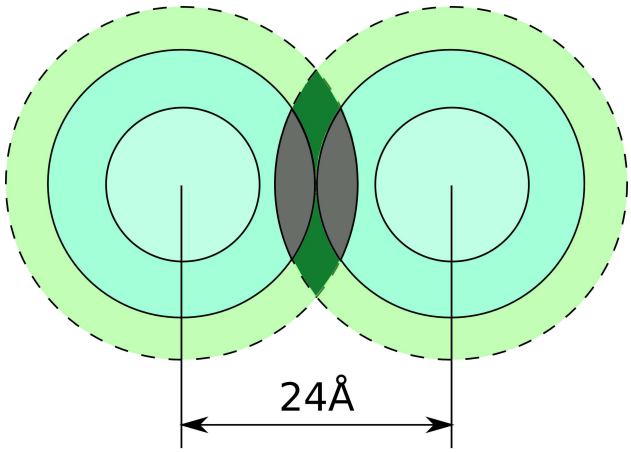

Figure 3.18: Schematic of the overlaps of the "external" and "internal" ion binding shells of two NA duplexes at the inter-axial separation $d = 24$ Å. The regions of the "external-internal" shells overlap are gray. Dark green indicates "external-external" shells overlap as in Fig. 3.16

one duplex once entering the "internal" shell of adjacent duplex when these shells overlap. Whatever correction to μ_a might be applied due to the difference in the CoHex excess chemical potentials in these shells, it should be compensated by the original difference in the ion configurational entropies in these shells. Therefore, the contribution of each CoHex ion in the "external-internal" shell overlaps to ΔG_{attr} can still be approximated by the same value of μ_a as in the case of the CoHex ions in the "external-external" shell overlaps. This approximation allows us to use Eq. 3.2 for the close inter-duplex separation, Fig. 3.18, as well, but now ΔN_s becomes ΔN, the sum of the numbers of ions in two overlapping shell regions, "external-external" and "external-internal" (combined gray and dark green regions in Fig. 3.18). This quantity can be estimated as

$$\Delta N = 2\rho_s \Delta V_s + (\rho_i + \rho_s)\Delta V_{is} = \left[\frac{f_i}{V_i}\Delta V_{is} + \frac{f_s}{V_s}(\Delta V_{is} + 2\Delta V_s)\right]\frac{|Q_{na}|}{Ze}\Theta\,, \tag{3.12}$$

where ΔV_{is} is the volume of the "external-internal" overlap, ρ_i and f_i are the number density and the fraction of CoHex ions bound in the "internal" ion binding shell of the volume V_i. Similar to f_s, the fraction f_i is defined as $f_i = N_i^0/N_0$ where N_i^0 is the number of CoHex ions in the "internal" shell determined from the all-atom MD simulations. Estimations of N_i^0, N_0 are in Table 3.3, while ΔV_{is} and ΔV_s are presented in Section 3.15.

Using Eq. 3.12, the attractive contribution to the aggregation free energy, Eq. 3.8, can be rewritten as

$$\Delta G_{attr}(\Theta) = 3\mu_a\left[\frac{f_i}{V_i}\Delta V_{is} + \frac{f_s}{V_s}(\Delta V_{is} + 2\Delta V_s)\right]\frac{|Q_{na}|}{Ze}\Theta\,. \tag{3.13}$$

We do not consider the "internal-internal" shell overlaps, as these could only occur at such short inter-duplex separations where steric repulsion would become prohibitively large.

In the case of "external-internal" ion binding shell overlap, the quantity ΔN, Eq. 3.12, has to be used instead of ΔN_s in determining the correction Λ, Eq. 3.9, to the duplex charge re-normalization coefficient $\xi = 1 - \Theta + \Lambda$. The latter one can now be written as

$$\xi = 1 - \Theta + \frac{1}{2}\left[\frac{f_i}{V_i}\Delta V_{is} + \frac{f_s}{V_s}(\Delta V_{is} + 2\Delta V_s)\right]\Theta\,, \tag{3.14}$$

resulting in the following duplex-duplex repulsive term in the aggregation free energy,

$$\Delta G_{el-rep}(\Theta) = 3\Delta G_{el-na}\left(1 - \Theta\left[1 - \frac{1}{2}\left(\frac{f_i}{V_i}\Delta V_{is} + \frac{f_s}{V_s}(\Delta V_{is} + 2\Delta V_s)\right)\right]\right)^2. \tag{3.15}$$

As in the case of the duplex separations in the aggregated phase when only the "external" ion binding shells overlap, for long duplexes the attractive and repulsive contributions, Eqs. 3.13 and 3.15, depend linearly on the duplex length H through the bare duplex charge, $Q_{na} = \lambda_{na}H$, and bare duplexes repulsion, $\Delta G_{el-na} \sim \lambda_{na}^2 H$, respectively. The entropic contribution, Eq. 3.11, remains unchanged and scales logarithmically with H.

Note that Eqs. 3.13 and 3.15 reduce to simpler Eqs. 3.8 and 3.10 when the "external-internal" overlap volume ΔV_{is} vanishes for the inter-axial duplex separations $d \geq r_i + r_s = 28$ Å ($r_i + r_s$ is the sum of the outer radii of the "internal" and "external" ion binding shells). Since the typical inter-axial duplex separation in the CoHex induced DNA aggregates is 28 Å [206, 201, 202], for this and larger separations we can use more simple Eqs. 3.8 and 3.10.

3.12 Application of the Multi-Shell Model and Discussion

3.12.1 The stability of nucleic acid aggregates.

We apply our semi-quantitative model, Eq. 3.1, where the terms are determined by Eqs. 3.13, 3.15 and 3.11, to characterize CoHex induced aggregation of four 25-bp long nucleic acid duplexes previously simulated and studied experimentally as described in Section 3.2; these are DNA, RNA and DNA:RNA hybrid duplexes of the equivalent mixed sequence [8, 193] and homopolymeric poly(dA):poly(dT) DNA duplex.

The model presented here allows us to estimate the aggregation free energies of these duplexes, ΔG_{aggr}, as functions of the degree of duplex neutralization by bound CoHex ions, Θ, and compare them with the duplex condensation propensities observed in Section 3.3. The aggregation begins when Θ, which depends on the bulk CoHex concentration in the solution

Table 3.2: Estimated changes of the free energy, $\Delta G_{aggr} = \Delta G_{attr} + \Delta G_{el-rep} - T\Delta S_{conf}$, upon CoHex induced 25-bp NA duplex aggregation under experimentally relevant conditions (CoHex concentration corresponds to 94% of duplex charge neutralization, NA duplex concentrations in the solution phase is 40 μM). All energy components are in k_BT units per one duplex, $T = 300$ K. The calculations assume 28 Å inter-axial duplex separation [206, 202] in the hexagonal aggregated phase that corresponds to "external-external" CoHex binding shells overlap shown in Fig. 3.16.

	dA:dT	DNA	HYB	RNA
CoHex binding affinity, μ_a	-7.66	-7.57	-7.34	-5.79
Fraction of "externally" bound ions, f_s	0.67	0.59	0.24	0.14
Number of ions in single overlap, ΔN_s	2.41	2.12	0.88	0.49
ΔG_{attr}	-55.5	-48.2	-18.6	-8.4
ΔG_{el-rep}	27.6	24.0	11.5	8.7
$-T\Delta S_{conf}$	22.2	22.2	21.9	21.9
ΔG_{aggr}	-5.7	-2.0	14.8	22.2

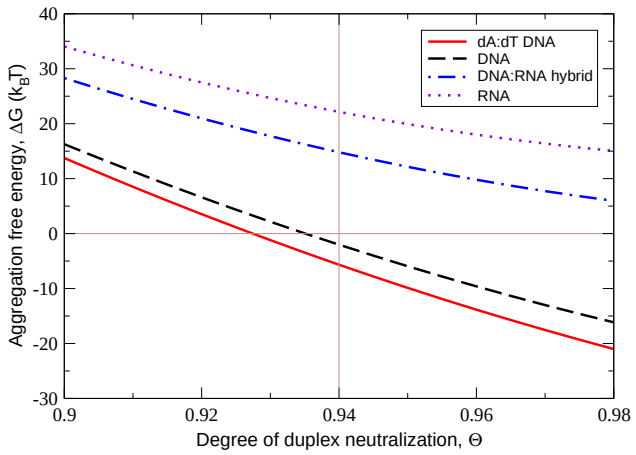

Figure 3.19: Predicted aggregation free energy ΔG_{aggr} for the four 25-bp NA duplexes as a function of the degree of duplex charge neutralization Θ by bound CoHex ions. The calculations assume 28 Å inter-axial duplex separation in the hexagonal aggregated phase and 40 μM duplex concentration in the solution phase. The data in the Table 3.2 correspond to $\Theta = 0.94$ (thin vertical line).

phase, reaches the level at which the NA aggregated phase is more stable than the solution phase, i.e. when $\Delta G_{aggr} < 0$.

Our main results for ΔG_{aggr} and its components at the typical 28 Å inter-axial duplex separation in the hexagonal aggregated phase are summarized in Table 3.2 and Fig. 3.19. The latter shows ΔG_{aggr} as a function of Θ. In Table 3.2, the attractive, repulsive and configurational entropy components of ΔG_{aggr} for the four NA duplexes are compared for the same degree of duplex neutralization $\Theta = 0.94$, at which the aggregated phase for the two considered DNA duplexes is more stable than the solution phase, $\Delta G_{aggr} < 0$. The main factors determining the CoHex ion-mediated duplex attraction, CoHex binding affinity to the "external" ion binding shell μ_a and the fraction of bound CoHex ions in this shell f_s, are presented as well.

The predicted aggregation free energies suggest that at the same degree of duplex neutralization the DNA aggregates are more stable than the DNA:RNA hybrid or RNA aggregates (see Fig. 3.19). Among the two DNA molecules, the mixed sequence DNA duplex requires higher degree of neutralization by CoHex ions for the beginning of the aggregation, $\Delta G_{aggr} < 0$. Provided that the CoHex affinities μ_a for both DNA duplexes are roughly the same (see Table 3.2), the higher Θ at $\Delta G_{aggr} = 0$ for the mixed sequence DNA translates into the higher bulk CoHex concentration in the solution phase at which mixed sequence DNA duplexes condense. At roughly the same μ_a for both DNA duplexes, the major difference in the values of the attractive components stems from the difference in the fractions of bound CoHex ions in the "external" ion binding shells f_s of these duplexes, 0.67 for the homopolymeric DNA and 0.59 for the mixed sequence DNA, leading to different values of the number of CoHex ions ΔN_s in the overlaps of these shells for two adjacent NA duplexes in the aggregated phase.

Comparing the DNA:RNA hybrid and RNA duplexes, the differences in both μ_a and f_s (and, therefore, ΔN_s) contribute to a smaller value of the attractive component in ΔG_{aggr} for the RNA duplex. The same conclusion is valid when comparing DNA and RNA duplexes. However, the major part of the difference in ΔG_{attr} for DNA and RNA duplexes arises from a drastic difference in f_s, 0.59 and 0.14, respectively (see Table 3.2), leading to four fold difference in the number of CoHex ions ΔN_s in the shell overlaps.

3.12.2 Dependence of the NA aggregation free energy on inter-axial duplex separation

A more detailed investigation of the model predictions over a range of duplex-duplex separation distances and degrees of duplex neutralization is presented in Fig. 3.20; the trends are exemplified for the mixed sequence DNA and RNA molecules. The aggregation free energy values are estimated using Eqs. 3.11, 3.13 and 3.15. Several conclusions can be made that further support the validity of our approach.

For both DNA and RNA, a region of $\Delta G_{aggr} < 0$ as a function of inter-duplex separation d in the aggregated phase exists at large enough degrees of duplex neutralization Θ by CoHex ions. For the mixed DNA duplex, the region of the favorable condensation conditions begins

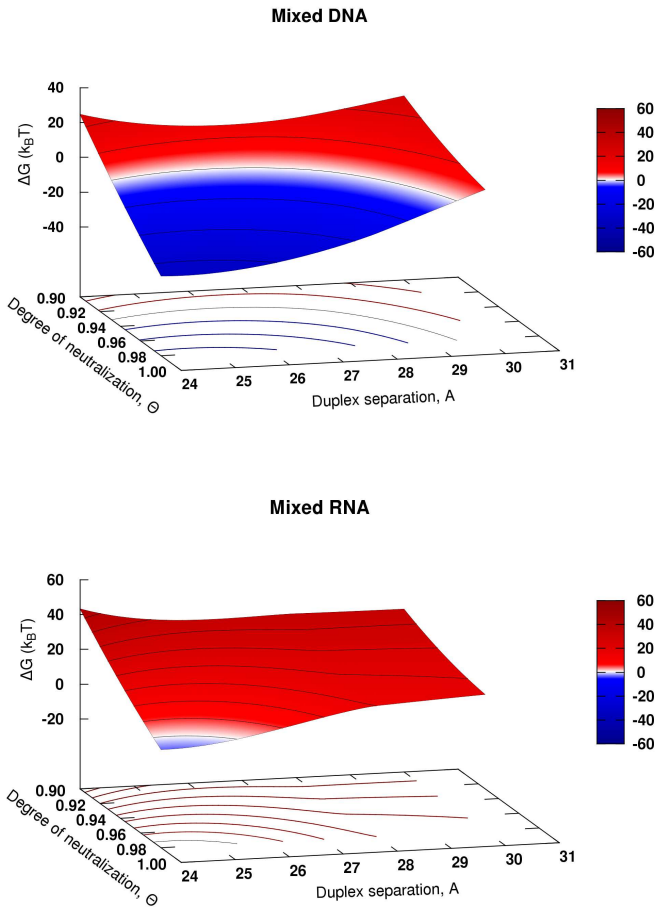

Figure 3.20: Predicted aggregation free energy ΔG_{aggr} of 25-bp long DNA and RNA duplexes as a function of the degree of duplex neutralization Θ and the inter-axial duplex separation, d. Condensation conditions, $\Delta G_{aggr} < 0$, correspond to blue regions of the parameter space, while the white band designates $\Delta G_{aggr} \approx 0$. Lines of equal ΔG_{aggr} are projected onto the $d - \Theta$ plane.

at $\Theta > 0.92$ with a minimum of $\Delta G_{aggr}(d)$ at the inter-duplex separation $d = 26$ Å. This minimum is not too different from the known experimental value $d = 28$ Å measured for long DNA strands [206, 202]. With the increase of Θ, the minimum slightly shifts toward the smaller separations.

The existence of the minimum of $\Delta G_{aggr}(d)$ can be rationalized by considering differences in scaling behavior of the attractive and repulsive components of ΔG_{aggr} within our model. The attractive part ΔG_{attr} scales linearly with the number of CoHex ions in the shells overlap ΔN_s (Eq. 3.2), which increases as the inter-duplex separation d decreases. At the same time, the repulsive contribution ΔG_{el-rep} is quadratic in ΔN_s, Eqs. 3.9 and 3.10.

Also, it is reassuring that at typical DNA condensation conditions, the predicted value of ΔG_{aggr} per base pair is a fraction of k_BT, in agreement with the experimental estimates [29].

Relative to the DNA, the region of the parameter space (d,Θ) where the RNA duplexes may be expected to aggregate is very narrow, suggesting that almost complete neutralization of the RNA duplex by bound CoHex ions is required prior to aggregation. This result is consistent with the notion that RNA is generally harder to condense than the equivalent in sequence DNA [147, 257], requiring a much larger CoHex concentrations in the solution phase.

Several testable predictions can be made directly from Fig. 3.20. A close examination reveals that, compared to the DNA, the aggregated phase of RNA is expected to have shorter inter-duplex distances, $\sim$1.5 Å shorter than for the aggregated DNA phase. Physically, this is because even if the RNA neutralization were 100%, where $\Delta G_{el-rep} \approx 0$, a relatively weak attractive force at 28 Å of inter-duplex separation would not be enough to overcome the entropic cost of aggregation of short 25-bp RNA duplexes, $\sim$22 k_BT per duplex. At inter-duplex distances shorter than 28 Å, an additional attractive contribution comes into play, the one resulting from the counterions in the more populated "internal" ion binding shell of the RNA. These shells begin to overlap with the "external" ion binding shells of the opposite duplexes at the separation $d < 28$ Å, Fig. 3.22. That additional attractive force tips the total attraction-repulsion balance towards RNA condensation at 25 Å inter-duplex separation and near 100% ($\Theta = 0.99$) duplex neutralization. However, as the duplex-duplex separation decreases further in the already tightly packed aggregate, both the configurational entropy loss and the electrostatic repulsion increase, quickly drive the RNA out of the condensation regime.

Another specific prediction that follows directly from the free energy balance within our model is that longer duplexes are expected to condense better than shorter ones. The explanation, based on Eqs. 3.1,3.2,3.10,3.11, is as follows. All of the contributions to ΔG_{aggr}, except the configurational entropy loss, Eq. 3.11, are proportional to the duplex length H (if one neglect the end effects, which is a reasonable approximation for the number of base pairs in a duplex $\gg 1$). In contrast, the unfavorable configurational entropy change scales logarithmically with H, Eq. 3.11. Thus, as H increases, the relative (per unit length) contribution of the entropic term to ΔG_{aggr} diminishes. Since the entropic term resists condensation, we expect longer duplexes to condense better.

3.12.3 Over-all model performance compared to experiment

It is reassuring that our estimations based on MD simulation results for CoHex ion distributions produce CoHex binding affinities of the same magnitude as earlier estimates of CoHex-DNA binding free energies based on equilibrium dialysis and single-molecule magnetic tweezers study [256] (-14.5 k_BT), isothermal titration calorimetry data [170] (-13.7 k_BT), and ^{59}Co chemical shift NMR measurements [34] (-8.6 k_BT). The predicted CoHex affinities to the "external" ion binding shells of the nearly neutralized DNA duplexes are smaller than the above experimental estimates, which makes sense since the latter were derived from the apparent binding constants in the limit of zero binding; i.e., in the absence of the mutual repulsion between bound CoHex ions that decreases the effective affinity in our calculations.

More importantly, the values in Table 3.2 correctly predict the aggregation propensity of different nucleic acid duplexes, ranging from the favorable aggregation energies for the poly(dA):poly(dT) and mixed sequence DNA constructs to the unfavorable energies for the DNA:RNA hybrid and RNA duplex at the same degree of duplex neutralization. The fact that the order of the condensation propensities observed in the condensation experiments (see Section 3.2) are captured by our semi-quantitative model, last line in Table 3.2, provides further support for the robustness of the conceptual picture we have developed.

We have verified that the results for the aggregation free energy and its components presented above are robust with respect to simulation details: water model used for simulations, initial conditions and details of the RNA duplex simulation protocol (see Section 3.17).

3.13 Methodological details

3.13.1 Additional assumptions and considerations.

1. Since the condensation experiments we are trying to explain have been conducted at low (20 mM) concentrations of monovalent ions in solution, when they produce a negligible effect on the bound CoHex ions [34], the effects of monovalent salt are omitted from our semi-quantitative model. Whenever appropriate, we will discuss the consequences of neglecting the presence of small amount of monovalent salt.

2. The fractions of NA duplex charge neutralized by bound CoHex ions, Θ, in the solution and aggregated phases are considered to be the same. That is the aggregation energy we estimate does not include the contributions related the changes of CoHex concentration (and Θ) when CoHex ions are added to the solution phase prior to the aggregation and replace the monovalent counterion atmosphere around the NA duplexes.

3. The aggregated phase as a whole is assumed to be net neutral. In addition to bound

CoHex ions, a small amount of unbound ions between the duplexes in a bundle or between the layers of duplex bundles are assumed in the aggregated phase. This assumption of the overall neutrality of the aggregated phase allow one to neglect the interactions between the duplexes beyond the nearest neighbors.

4. We use the degree of NA duplex neutralization Θ as a convenient proxy for bulk CoHex concentration. Even though CoHex concentration in the bulk is the quantity most relevant for the thermodynamics analysis presented, it is not easily accessible experimentally. Typically, NA condensation experiments simply report the total CoHex concentration in the solution prior to the condensation, which does not equal the bulk CoHex concentration. The relationship between the two and Θ is complex [230, 182], but it is monotonic, which justifies the use of Θ for our purposes: we are interested in the relative NA condensation propensities, not their absolute values. The relation between Θ and the bulk multivalent ion concentration has some non-trivial form [230, 182] different from the original Manning predictions [163, 164] and predictions based on the Poisson-Boltzmann (PB) model [203, 245, 66, 191] due to correlations between bound multivalent ion at the surface of NA molecules. We do not need to consider the exact relation here. The only general property of the layer of bound (condensed) trivalent counterions at the NA duplex surface we will use is that the degree of NA duplex charge neutralization by these counterions, Θ, weakly depends on the bulk ion concentration and monotonically increases with the increase of the latter [230, 182]. At low (mM) trivalent ion bulk concentration the value of Θ is close to the Manning theory prediction for the double-stranded DNA, $\Theta \approx 0.9$.

3.13.2 Estimation of CoHex binding affinity μ_a.

We approximate the average binding energy that a CoHex ion bound in the "external" binding shell of NA duplex 1 gains when it enters the "external" binding shell of the (neighboring) NA duplex 2, by the CoHex binding affinity μ_a to the "external" ion binding shell of an isolated duplex in the solution phase at near neutralizing conditions. This approximation is based on the assumption that minimal changes in the CoHex distributions occur when the two NA duplexes approach each other at the inter-duplex distances relevant to NA condensation.

The above assumption was tested in Section 3.3 for homopolymeric poly(dA):poly(dT) DNA and mixed sequence RNA duplexes. By comparing the superposition of two independent single duplex CoHex distributions with CoHex distribution around a pair of NA duplexes separated by 26 Å, no changes in the number of "internally" bound (inside 12 Å from the helical axis) CoHex ions were observed. A small increase (1.1 ions) in the "external" ion binding shell (12 to 16 Å from the helical axis) of a duplex in both DNA and RNA duplex pairs was seen. The latter was suggested to be due to some redistribution of the bound ions in the external shell toward the shell overlapping region of the pair, which can be neglected

in the case of six neighbors of the hexagonal packing considered in this work.

CoHex binding affinity μ_a to the "external" ion binding shell of NA duplex can be defined as a difference of the excess chemical potentials of CoHex ion in the "external" shell of the duplex, μ^s_{excess}, and in the bulk, μ^b_{excess},

$$\mu_a = \mu^s_{excess} - \mu^b_{excess}\,. \tag{3.16}$$

The quantity reflects the change of the ion interaction with its environment as the ion moves from the bulk towards the surface of charged NA molecule.

The chemical potential of CoHex ion, μ, can be written as

$$\mu = \mu_{excess} + \mu_{ideal}\,, \tag{3.17}$$

where

$$\mu_{ideal} = k_B T \ln(\rho_n C) \tag{3.18}$$

is a portion of the potential that can be treated as a chemical potential of an ideal gas with a particle number density ρ_n corresponding to the CoHex density [271], C is a constant which does not depend on ρ_n. At equilibrium between the bulk (b) and the layer of bound CoHex ions in the "external" ion binding shell (s), μ is constant throughout the system:

$$\mu^s_{excess} + \mu^s_{ideal} = \mu^b_{excess} + \mu^b_{ideal} \tag{3.19}$$

From Eqs. 3.18 and 3.19 and the definition of the ion binding affinity, Eq. 3.16, we arrive at Eq. 3.3, i.e. $\mu_a = -k_B T \ln(\rho_s/\rho_b)$.

CoHex number density in the "external" ion binding shell $\rho_s = N^0_s/V_s$ is estimated using the average number of CoHex ions in this shell N^0_s determined from the MD simulations and presented in Table 3.3. With the height $H = 88$ Å for the two DNA duplexes and $H = 76$ Å for the DNA:RNA hybrid and RNA duplexes, the "external" shell volume, $V_s = \pi(r_s^2 - r_i^2)H$, constitutes 30964 and 26741 Å^3, respectively.

Table 3.3: Average numbers of bound CoHex ions in the "external" and "internal" ion binding shells around the four NA duplexes estimated from the MD simulations.

	DNA(dA:dT)	DNA	HYB	RNA
"External" shell ions, N^0_s	9.8	8.5	3.3	2.0
"Internal" shell ions, N^0_i	4.6	4.2	7.7	9.4
All bound CoHex ions, N_0	14.5	14.4	14.1	14.7

3.13.3 Estimation of repulsive interaction ΔG_{el-na} between the two unscreened NA duplexes.

We define $\Delta G_{el-na}(d)$ as the difference of the electrostatic free energy of the two duplexes in water, ΔG_{el}, at the separations d and ∞. The ΔG_{el} is estimated as [187]

$$\Delta G_{el} = \Delta G_{solv} + E_{Coul}\,. \tag{3.20}$$

Here ΔG_{solv} is the solvation energy of the two duplexes, estimated from the solution of the PE, and E_{Coul} is the Coulomb charge-charge interactions in the system computed with AMBER12 [42]. Both terms in Eq. 3.20 are computed using charge distributions on NA duplexes determined from the AMBER ff99bsc0 nucleic acid force-field [47, 194]. The ΔG_{el} for the two infinitely separated duplexes is calculated as the sum of the electrostatic free energies of the two isolated molecules.

We estimate ΔG_{el} by solving the Poisson equation (PE) for the two parallel unscreened duplexes and averaging the interaction energies over 12 different mutual duplex orientations. The interactions are computed at two different inter-axial separations, $d = 24$ and 28 Å. The orientations were changed by rotating one of the duplexes around its helical axis with 30° increment. The interactions at other inter-axial duplex separations were estimated using an interpolation, see below.

The PE was solved using MEAD solver [17] with three levels of focusing and with 281 grid points in each dimension: the coarsest grid spacing was 5.0 Å, and the finest 0.5 Å. The internal (NA duplex) and the external (water) dielectric constants were 4.0 and 78.5, respectively [116, 246].

The interaction $\Delta G_{el-na}(d)$ between the two parallel NA duplexes in water is then estimated as the difference of the orientationally averaged ΔG_{el} at a separation d and ∞,

$$\Delta G_{el-na}(d) = \langle \Delta G_{el}(d) \rangle - \Delta G_{el}(\infty)\,. \tag{3.21}$$

The value of $\Delta G_{el-na}(d)$ at the inter-duplex separations different from the two distances ($d_1 = 24$ Å and $d_2 = 28$ Å) for which it was directly calculated is then estimated using the logarithmic interpolation,

$$\Delta G_{el-na}(d) = \Delta G_{el-na}(d_1) + (\Delta G_{el-na}(d_2) - \Delta G_{el-na}(d_1)) \frac{\ln(d_1/d)}{\ln(d_1/d_2)}\,. \tag{3.22}$$

At high degree of duplex neutralization, ∼90%, the variation of the scaled $\Delta G_{el-na}(d_2)$ with one of the duplex rotating about its helical axis does not exceed 0.3 k_BT, consistent with an earlier estimate of $\sim 0.5 k_BT$ per 25 base pairs based on a different model [129]. The insignificance of the variation justifies the use of the simple averaging for ΔG_{el-na}. Small ($\Delta Z/H \ll 1$) axial translations of one duplex with respect to another reduce the repulsive

contribution mostly through the reduction of Λ which is proportional to the shell overlapping volume ($\Lambda \sim \Delta N_s \sim \Delta V_s$). At the same time, the change in ΔN_s leads to a reduction of the attractive contribution to the aggregation energy resulting in a destabilization of the duplex bundle.

3.13.4 Atomistic simulations.

The details of CoHex ion distributions and its chemical potential were extracted from MD simulations (see Section 3.2.4). A brief description of the protocols used are presented below. Other details of the duplex structure preparations and simulations are described in Section 3.2.

All-atom MD simulations were used to generate CoHex ions distributions around four 25-bp NA duplexes: homopolymeric poly(dA):poly(dT) DNA, mixed sequence DNA, DNA:RNA hybrid and RNA. Each system contained one 25-bp NA duplex, a neutralizing amount of 16 CoHex ions and 16880 TIP3P [117] water molecules. To avoid uncertainties associated with monovalent ions force field parameters [280], no monovalent salt was added to the simulated systems. The simulated trajectories were 300-380 ns long, generated at 300 K temperature in the canonical (NVT) ensemble. The simulations were performed using AMBER12 package [42] and ff99bsc0 force-field [47, 194]. Since no measurable effects of CoHex on the DNA duplex structures were experimentally observed, the homopolymeric and mixed sequence DNA duplexes were restrained to their B′ [3] and canonical B-form, respectively, during the simulations. These restraints minimized possible structure bias due to the use of imperfect modern force-fields [159]. On the other hand, the addition of CoHex has lead to the experimentally observed changes in RNA helical structure. Therefore, the RNA and DNA:RNA hybrid duplexes were simulated unrestrained allowing the duplex structures to relax when CoHex ions bind to these molecules. The CoHex ion distributions were analyzed using 28,000, 30,000 and 34,000 snapshots extracted from DNAs, DNA:RNA hybrid and RNA duplex trajectories, respectively.

3.14 Long-range correction $\varphi(r)$ to CoHex binding affinity estimation

To account for the finite size of the simulation box used in all-atom MD simulations of CoHex distributions around DNA, RNA and hybrid DNA:RNA duplexes, a long-range correction $Ze\varphi(r_B)$ to CoHex binding affinity μ_a at the system boundary ($r_B = 31$ Å) was introduced. At high degree of nucleic acid (NA) duplex neutralization by bound CoHex ions ($\Theta = 0.90$ – 0.94) and in the absence of monovalent salt in the simulations, the electrostatic potential of

the NA duplex partially neutralized by bound CoHex ions, $\varphi(r_B)$, is estimated (Eq. 3.5) as

$$\varphi(r) = \frac{2\lambda}{\varepsilon} \ln\left(\frac{H/2 + \sqrt{(H/2)^2 + r^2}}{r}\right) . \tag{3.23}$$

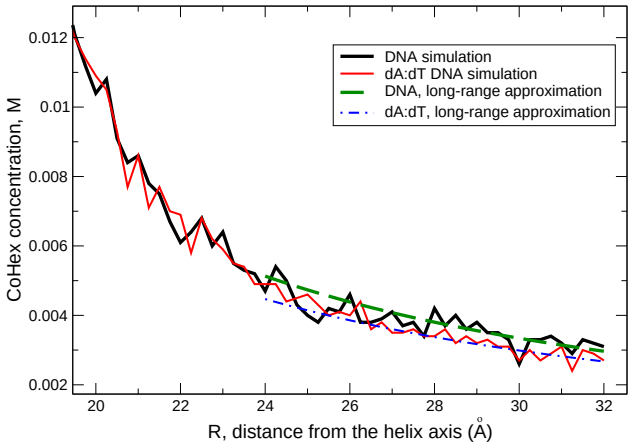

Figure 3.21: Simulated (all-atom MD) and analytically approximated behavior of the CoHex concentration as a function of distance R from the duplex helical axis for homopolymeric DNA (dA:dT) and mixed sequence DNA. The approximation uses Eq. 3.23 for the long-range behavior of the electric potential $\varphi(r)$ around the duplex.

Here λ is a linear charge density of 25-bp NA duplex of length H ($H = 88$ Å for DNA and $H = 76$ Å for RNA and hybrid molecules) and charge $Q_{na} = -48e$ neutralized by bound CoHex ions ($N_0 Ze$), $\lambda = Q_{na}(1 - \Theta_0)/H$, where $\Theta_0 = N_0 Ze/|Q_{na}|$. The potential $\varphi(r)$ is zero at infinity.

To test a proper asymptotic behavior of $\varphi(r)$ estimated by Eq. 3.23, we compare the simulation results for CoHex concentration profile with the concentration profile $\rho(r)$ determined by $\varphi(r)$ using the Boltzmann distribution of the ions in the electrostatic potential,

$$\rho(r) = \rho_b e^{-Ze\varphi(r)/k_B T} . \tag{3.24}$$

The results of the comparison are presented Fig. 3.21, confirming the validity of the approximation Eq. 3.23 for the long-range correction to μ_a in the no-salt conditions. The values of the potential $\varphi(r)$ estimated by Eq. 3.23 are relatively small ranging from -0.7 to -1.1 $k_B T/e$ at $r_B = 31$ Å for the NA duplexes of interest.

3.15 Evaluation of the ion binding shell overlap volumes ΔV_s and ΔV_{is}

The volumes of the "external-external" ΔV_s and "external-internal" ΔV_{is} overlaps of the ion binding shells of two identical parallel NA duplexes, Fig. 3.22, are uniquely determined by the geometries of the shells and the mutual positions of the duplexes. The latter ones are characterized by the duplex-duplex separation d and the vertical misalignment of the ends of the parallel duplexes ΔZ. The cylindrical shell geometries are defined by the duplex height H and by the values of the shell radii – the internal radius of the "internal" shell, $r_0 = 7$ Å, and the outer radii of the "internal", $r_i = 12$ Å, and the "external", $r_s = 16$ Å, shells. With the height $H = 88$ Å for the two DNA duplexes and $H = 76$ Å for the DNA:RNA hybrid and RNA duplexes, the "external" shell volume, $V_s = \pi(r_s^2 - r_i^2)H$, constitutes 30964 and 26741 Å^3, respectively.

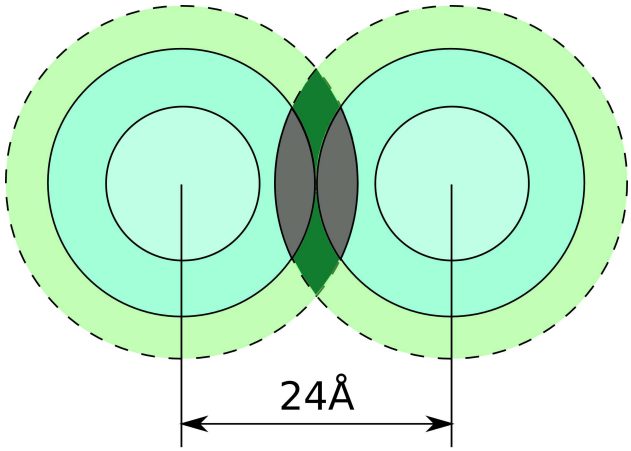

Figure 3.22: Schematic of the overlaps of the "external" and "internal" ion binding shells of two NA duplexes at the inter-axial separation $d = 24$ Å. The regions of the "external-internal" shells overlap are gray. Dark green indicates "external-external" shells overlap.

The "external-internal" overlap volume ΔV_{is} can be written as

$$\begin{aligned} \Delta V_{is} &= 2(H - 2\Delta Z)\left(r_i^2 \arccos(p_1) + r_s^2 \arccos(p_2) - 0.5(p_3)^{1/2}\right) \text{ for } 2r_i \le d \le r_s + r_i \\ \Delta V_{is} &= 0 \text{ for } d \ge r_s + r_i \ . \end{aligned} \tag{3.25}$$

Here arccos() arguments are

$$\begin{aligned} p_1 &= (d^2 + r_i^2 - r_s^2)/(2dr_i)\,, \\ p_2 &= (d^2 + r_s^2 - r_i^2)/(2dr_s)\,, \\ p_3 &= (r_i + r_s - d)(r_i - r_s + d)(r_s - r_i + d)(r_i + r_s + d)\,. \end{aligned} \tag{3.26}$$

The "external-external" overlap volume ΔV_s is estimated as

$$\Delta V_s = (H - 2\Delta Z)\left(2r_s^2 \arccos(u_1) - 0.5(u_2)^{1/2}\right) - \Delta V_{is} \text{ for } d \leq 2r_s. \quad (3.27)$$

and $\Delta V_s = 0$ for $d \geq 2r_s$. Here $u_1 = d/(2r_s)$ and $u_2 = (4r_s^2 - d^2)d^2$.

In our estimations of the attractive and repulsive electrostatic contributions to ΔG_{aggr} which depend on the values of ΔV_{is} and ΔV_s determined by Eqs. 3.25 and 3.27 we neglect small ($\Delta Z/H \ll 1$) misalignments of the duplexes in the aggregated phase as they reduce both of these contributions tending to cancel their effects.

3.16 Electrostatic interaction between the two bare duplexes in water

We estimated the orientationally averaged repulsive interaction $\Delta G_{el-na}(d)$ between the two unscreened NA duplexes in water (Eq. 3.21) for the two inter-axial duplex separations, $d_1 = 24$ Å and $d_2 = 28$ Å. The results are presented in Table 3.4.

Table 3.4: Orientationally averaged repulsive interaction $\Delta G_{el-na}(d)$ between the two parallel unscreened NA duplexes in water calculated by solving the Poisson equation for the solvation energy and by estimating the Coulomb charge-charge interactions in the system using AMBER12 [42] and NA duplex charge distributions determined from the AMBER ff99bsc0 nucleic acid force-field [47, 194]. All energies are in k_BT units per duplex pair. The internal (NA duplex) and the external (water) dielectric constants were 4.0 and 78.5, respectively.

NA duplex pair	Inter-axial distance $d_1 = 24$ Å	Inter-axial distance $d_2 = 28$ Å
dA:dT – dA:dT (DNA)	565.0	502.3
DNA – DNA	565.0	502.3
HYB – HYB	594.4	515.0
RNA – RNA	594.4	515.0

3.17 Robustness to force-field details and initial conditions

To investigate possible influence of force-filed guided structural relaxation of the NA duplexes on the estimated aggregation free energy components, we have repeated MD simulations as described in "Atomistic simulations" for the RNA duplex, but now with all the atoms in

the RNA duplex restrained to their initial canonical A-RNA positions. Being a different simulation, it also automatically examines robustness to initial conditions, including CoHex ion positions at $t = 0$. These results, summarized in Table 3.5 and Fig. 3.23, show only modest differences from our main estimates of the ΔG_{aggr} components based on the CoHex distribution around unrestrained RNA duplex; the estimate of CoHex binding affinity to the "external" ion biding shell, μ_a, differs from that based on the unrestrained simulations by 0.7 k_BT and the aggregation free energy is less favorable by 3 k_BT.

Table 3.5: Robustness of the free energy estimates to RNA force-field and initial conditions. Shown are estimated changes of the free energy, $\Delta G_{aggr} = \Delta G_{attr} + \Delta G_{el-rep} - T\Delta S_{conf}$, upon CoHex induced 25-bp RNA duplex aggregation under experimentally relevant conditions, see Table 3.2. All energy components are in k_BT units per duplex, $T = 300$ K. The values in this table (right column, restrained RNA) are obtained from MD simulation with RNA atoms restrained to canonical A-RNA, while all other conditions remained as for the unrestrained RNA (Table 3.2).

	unrestrained RNA	restrained RNA
CoHex binding affinity, μ_a	-5.79	-5.05
Fraction of "externally" bound ions, f_s	0.14	0.07
Number of ions in single overlap, ΔN_L	0.49	0.26
ΔG_{attr}	-8.4	-4.0
ΔG_{el-rep}	8.7	7.2
$-T\Delta S_{conf}$	21.9	22.2
ΔG_{aggr}	22.2	25.4

Another important component of possible sensitivity to force-field parameters is the choice of specific water model used to represent solvation effects in all-atom MD simulations. To address this sensitivity, we have repeated the DNA-CoHex MD simulations described in the "Atomistic simulations" section using the TIP4P-Ew [109] rather than the TIP3P water model [117] used to obtain the main results. The resulting aggregation free energy estimates, presented in Fig. 3.24, show that the choice of water model has a very small (0.2 – 1.1 k_BT) impact on the estimated ΔG_{aggr} values.

3.18 Robustness to the definition of the "external" ion binding shell

Several basic approximations have been made when constructing the model. Among them are replacements of the ion density profile and the profile of ion excess chemical potential near the duplex by the average ion density and the ion affinity in the ion binding shell,

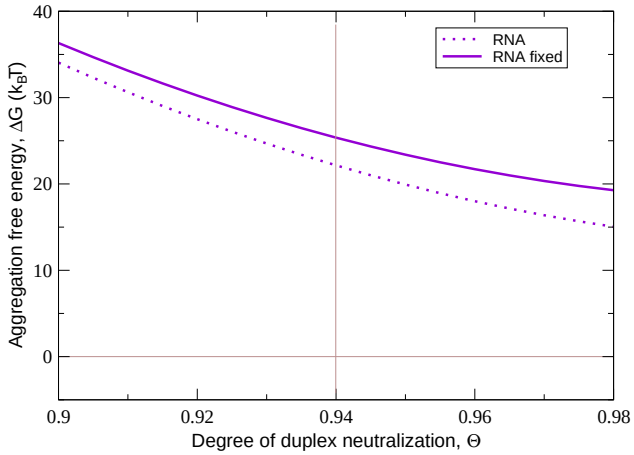

Figure 3.23: Robustness of ΔG_{aggr} to initial conditions and details of RNA duplex simulation protocol. The effect of relaxation of 25-bp RNA structure (unrestrained vs. restrained (fixed)) on ΔG_{aggr} as a function of duplex degree of neutralization. The calculations assume 28 Å inter-axial duplex separation in the hexagonal aggregated phase and 40 μM duplex concentration in the solution phase. The data in the Table 3.5 correspond to $\Theta = 0.94$ (thin vertical line).

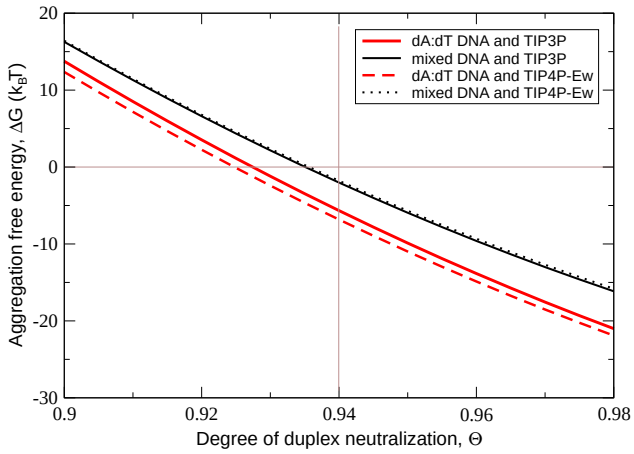

Figure 3.24: Robustness of the aggregation free energy ΔG_{aggr} to simulation details. The effect of water model (TIP3P vs. TIP4P-Ew) on ΔG_{aggr} as a function of duplex degree of neutralization Θ for the 25-bp long homopolymeric poly(dA):poly(dT) and mixed sequence DNA duplexes at $d = 28$ Å inter-axial duplex separation in the hexagonal aggregated phase.

Table 3.6: Robustness to water model. Estimated changes of the free energy, $\Delta G_{aggr} = \Delta G_{attr} + \Delta G_{el-rep} - T\Delta S_{conf}$, upon CoHex induced DNA duplex aggregation under experimentally relevant conditions (CoHex concentration corresponds to 94% of duplex neutralization, ~ 40 μM duplex concentrations). All energy components are in k_BT units per one duplex, $T = 300$ K. The calculations assume 28 Å inter-axial duplex separation in the hexagonal aggregated phase that corresponds to "external-external" CoHex binding shells overlaps. Here, CoHex binding affinity and number of ions in the shell overlaps are estimated from MD simulations in TIP4P-Ew water; all other conditions are the same as in Table 3.2 where TIP3P water was used for the estimates.

	dA:dT DNA	DNA
CoHex binding affinity, μ_a	-7.95	-7.57
Number of ions in single overlap, ΔN_L	2.33	2.10
ΔG_{attr}	-55.5	-47.7
ΔG_{el-rep}	26.7	23.8
$-T\Delta S_{conf}$	22.2	22.2
ΔG_{aggr}	-6.8	-1.8

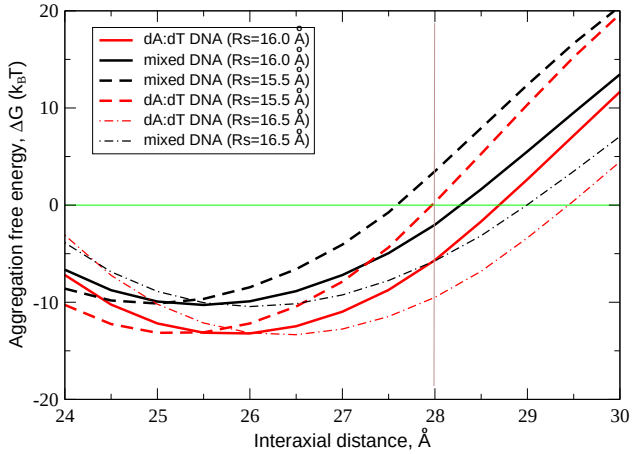

Figure 3.25: Robustness of the aggregation free energy ΔG_{aggr} to the definition of the "external" ion binding shell. Predicted ΔG_{aggr} for the two DNA duplexes at the degree of duplex charge neutralization $\Theta = 0.94$ is calculated as a function of the inter-axial duplex separation for the outer boundary of the "external" shell, $R_s = 16$ Å, shifted by ± 0.5 Å.

respectively. To comply with these approximations, the definitions of the "internal" and "external" ion binding shells allow only small degree of flexibility of their boundaries. The definitions are tied to the behavior of the ion density profiles around DNA and RNA duplexes as a function of distance to the helical axis – the profiles are rather steep at the boundary between the shells and vanish at the outer boundary of the "external" shell (see Fig. 3.3). The latter was defined in Section 3.3 as the distance from the helical axis (R_s = 16 Å) at which CoHex ions can still make direct contacts (hydrogen bonds) with the oxygen atoms of the phosphate groups, *i.e.* are not fully hydrated. Thus, a rational definition of the shells consistent with the model allows for only relatively small variations of the positions of the shell boundaries, about ±0.5 Å. Within this margin, the results are rather insensitive to the geometrical parameters of the ion binding shells (see Fig. 3.25). Specifically, the relative aggregation propensities of the duplexes remain the same, and predicted ΔG_{aggr} at the free energy minima do not change. At 28 Å inter-duplex separation, ΔG_{aggr} changes by only ±5 k_BT.

3.19 Conclusion

In this chapter we combine experiment and atomistic modeling to propose a mechanistic picture of counterion-induced condensation of nucleic acids (NA) and resolves differences in their condensation propensities. The picture can explain an unexpected experimental finding that NA condensation induced by trivalent counterion cobalt hexammine (CoHex) varies markedly between various short 25 bp double-helical structures, from RNA which largely resists condensation, to a homopolymer DNA which is most susceptible to it. A unique feature of our proposed mechanism is the relationship between NA duplex condensation and the location of preferential counterion binding relative to the NA helical axis. Specifically, condensation propensity is determined by the fraction of counterions bound to the external (outermost) surface of the double-helix. We found a significant difference, up to 5-fold, in the fraction of CoHex ions bound to the external surfaces of our different NA constructs. For example, in the simulated poly(dA):poly(dT) DNA structure 68% of the bound CoHex ions bind externally and can contribute to the NA-NA attraction and hence condensation, while in the mixed sequence double-stranded RNA of the same size, most CoHex ions (87%) are bound internally (inside the major groove), which explains the observed resistance to condensation. In contrast to many previous studies, we make a clear distinction between the total number of counterions bound to NA and the fraction of counterions bound to the outermost surface. The overall strength of counterion-NA binding does not solely determine the condensation propensity; instead, the specific structural feature of bound ion distributions is of the most important influence. Multivalent ion binding to RNA is stronger than to DNA, yet the former resists condensation. Importantly, it is not simply variation in NA helical structure parameters that ultimately affects the condensation propensity: we find that duplexes with the same helical parameters can also have different condensation propensities.

It is beyond the scope of the current study to model condensed DNA states using MD simulations and to suggest specific modes of NA-NA association (e.g., "bridge" vs. "zipper" etc.). Nevertheless, it is useful to examine our findings in the light of previous theoretical studies that attempt to explain the differences in DNA and RNA condensation [129] based on the differeces in the helical structure. The assumptions [129] made about the preferential placement of bound CoHex ions and the purely geometric interpretation are inconsistent with our experimental data and the results of atomistic simulations of CoHex binding. Our results show that in DNA the majority of bound CoHex are outside of the major groove and that the peaks of CoHex distributions overlap at the NA-NA distances that correspond to the condensed phase. While NA condensation depends on a complex interplay between various structural and sequence features, our coupled experimental and theoretical results suggest a new model in which a single parameter – the fraction of externally bound multivalent counterions – connects the NA condensation propensity with geometry and sequence dependence of CoHex binding.

In the second half of this chapter, we have developed the first semi-quantitative model of nucleic acid aggregation (condensation) induced by trivalent CoHex ($Co(NH_3)_6^{+3}$) counterions in which the radial distribution of the bound counterions is the key ingredient. Namely, the counterions bound to a NA duplex are partitioned into an "external" and "internal" ion binding shells; the fraction of ions in each shell, and their affinity to the nearly neutralized duplex can be accurately quantified from converged ion distributions obtained from all-atom molecular dynamics simulations in explicit solvent. Within the model, the source of the short-range attraction between two approaching NA duplexes are the oppositely charged CoHex ions from the overlapping region of the ion binding shells of these duplexes. Basic thermodynamic arguments are then used to estimate the aggregation free energy components corresponding to CoHex-mediated attraction between the duplexes and the opposing residual repulsion of the nearly neutralized duplexes and configurational entropy loss upon aggregation of the short duplexes. Importantly, robustness of the model estimates to simulation details has been thoroughly verified.

The key conclusion from this chapter is that our semi-quantitative model based on the "ion binding shells" framework is able to reproduce the correct order of condensation propensities seen in experiment for various NA duplexes, some of which differ in fairly subtle way, e.g., poly(dA):poly(dT) DNA vs. mixed sequence DNA. The key role of the fraction of multivalent ions bound in the "external" ion binding shell of NA duplexes in nucleic acid condensation is clarified. The larger the fraction of the externally bound multivalent ions, the larger the condensation propensity of a nucleic acid molecule.

Another conclusion is a relatively small value of the attractive interactions between the RNA duplexes at the inter-axial distances and degree of neutralization at which DNA duplexes begin to aggregate. To overcome the relatively large entropic cost of the aggregation of short RNA duplexes it is necessary both to increase the attractive component and to minimize the electrostatic repulsive contribution by a more complete duplex neutralization compare to the case of the DNA. We therefore predict that RNA condensation occurs at inter-helical

distances smaller than ones experimentally found in DNA aggregates. Consistent with experiment, the model predicts that higher degree of RNA neutralization is needed as well (in experiment, RNA condensation starts at considerably higher CoHex concentrations than for the equivalent DNA molecules).

In contrast to some of the available models of counterion-induced NA condensation, ours is simple enough to allow for analytical estimates and analyses of trends. For example, we show why longer NA fragments are expected to condense easier than shorter ones. At the same time, the apparent simplicity of the model does not come at the expense of its physical realism: the key component of the model – counterion affinity to the "external" ion binding shell and the fraction of ions in this shell are computed from explicit solvent atomistic simulations, which is arguably the most accurate practical approach to such estimates to-date. The details of the NA duplex geometry and sequence are implicitly included in the values of these parameters. The atomistic level of details "built into" the basics physics model greatly enhances versatility of the latter. In particular, in the future we are interested in exploring a wider range of multivalent ion types, and look more closely into sequence dependence of NA condensation.

Chapter 4

Opposing effects of multivalent ions on flexibility of DNA and RNA

4.1 Introduction

Mechanical properties of nucleic acids (NA) are known to be strongly affected by ionic conditions of the solutions they are in [19, 164, 270, 252, 175, 154, 16]. Experiments show that increasing the concentration of monovalent salt reduces DNA bending persistence length [19, 270] from its classical value of ~ 500 Å at physiological NaCl concentration of ~ 0.145 M. An even more dramatic decrease of DNA persistence length with salt concentration has been observed when monovalent cations were replaced by multivalent ones [19]. The common explanation is that counterion atmosphere around a NA molecule screens the repulsion between the negatively charged phosphates along the polymer, thus increasing its bending flexibility [186, 236, 80, 223, 198].

In contrast to DNA, mechanical properties of double-stranded RNA (referred to simply as RNA for the remainder of the text) have not been investigated as extensively. From the similar charge density and overall structure of DNA and RNA double helices, one might expect that they should exhibit similar elastic responses under applied forces. Indeed, recent single-molecule force and torque measurements of RNA bending flexibilities [1, 106, 150] have shown analogous to DNA decrease of bending persistence length of RNA with increasing monovalent salt concentration [106].

Here we compare the effects of multivalent counterions on RNA bending flexibility using atomistic molecular dynamics simulations; the study was motivated by a recent finding that binding patterns of multivalent counterions to DNA and RNA are drastically different [257, 193, 147]. We investigate how this difference in multivalent ion binding affects RNA bending persistence length when monovalent counterions are replaced by ions with charges $+3e$ and $+4e$, and propose an explanation for the opposing effects on bending flexibilities of

DNA and RNA.

4.2 Methods

DNA and RNA duplexes of 25 base pairs (bp) with the same mixed sequence described in [257] and a homopolymeric poly(rA)·poly(rU) RNA fragment of the same length were generated in canonical B-form for DNA and A-form for RNA using Nucleic Acid Builder [158]. The NA molecules were then neutralized with three different combinations of monovalent Na^+ and trivalent Cobalt(III) Hexammine (CoHex): 48 Na^+ ("no CoHex"), 8 $CoHex^{3+}$ and 24 Na^+ (0.17 $CoHex^{3+}$ per phosphate, indicated by P^-) or 16 $CoHex^{3+}$ (0.33 $CoHex^{3+}/P^-$). In addition, mixed sequence RNA molecules were also neutralized with two other counterion types: 16 hypothetical "Na^{3+}" ions and 12 $Spermine^{4+}$ ions. "Na^{3+}" was simulated by increasing threefold the Na^+ charge in Amber topology file. Each system was then solvated with ∼16800 TIP3P water molecules in a periodic box. To account for monovalent salt background roughly equivalent to physiological conditions, 24 Na^+Cl^- ion pairs were added to all systems. All MD simulations were carried out using AMBER 12 [43] and ff99bsc0 force field [47, 194] at $T = 300$ K. Each system was first simulated for at least 100 ns while holding the duplex restrained to allow the ionic atmosphere to equilibrate around the molecule. Then, positional restraints were removed for production MD runs: 300 ns for $CoHex^{3+}$ and "no CoHex" systems, and 400 ns for $Spermine^{4+}$ and "Na^{3+}" systems.

NA helical axis representations for each frame were generated from MD trajectories using Curves+ [141]. We then calculate, for each MD snapshot, the bending angle θ between average vectors of consecutive helical repeats (10 bp in our case), which should reduce the possibility of error from the non-symmetric nature of the double-helix[172]. The result of the calculation is the angular probability distribution $P(\theta, l)$, where l is the average contour length of the 10 bp helical repeat (calculated for each NA/ion system individually by summing all base-pair rise parameters). For $l \ll L_p$ we can use the inextensible worm-like chain [41, 274, 173, 273] approximation for the bending energy:

$$E(\theta, l) = \frac{1}{2} k_B T \frac{L_p}{l} \theta^2 \tag{4.1}$$

For spherically isotropic distribution in 3D space[173]

$$P(\theta, l) = \sin\theta \exp(-E(\theta, l)/k_B T) \tag{4.2}$$

which leads to

$$-\ln\left(\frac{P(\theta, l)}{\sin\theta}\right) = \frac{L_p}{2l}\theta^2 \tag{4.3}$$

DNA w/ CoHex^{3+}	$\Delta L_p/L_p$ (%)	$\Delta l/l$ (%)
0.17 CoHex^{3+}/P$^-$	-31	-1.8
0.33 CoHex^{3+}/P$^-$	-37	-1.8

Table 4.1: Effect of CoHex^{3+} counterions on DNA flexibility. Shown are relative changes in bending persistence length (L_p) and length of helical repeat (10 bp) segment (l), relative to the same system with no CoHex^{3+} ions (multivalent ions replaced by Na$^+$ to maintain over-all neutrality).

RNA w/ CoHex^{3+}	$\Delta L_p/L_p$ (%)	$\Delta l/l$ (%)
0.17 CoHex^{3+}/P$^-$	$+68$	-12
0.33 CoHex^{3+}/P$^-$	$+90$	-14

Table 4.2: Effect of CoHex^{3+} on RNA flexibility. Persistence length (L_p) of RNA increases with CoHex^{3+} concentration – an effect that is opposite to the DNA response to CoHex^{3+}. Bound CoHex^{3+} also significantly shrinks RNA duplex along its helical axis.

used to estimate L_p, Fig. 4.1. While this method may not provide the most accurate assessment of the absolute value of L_p for such small fragments (which could depend on the total length of the polymer), the qualitative effects of multivalent ions on NA persistence length observed in simulations are strong enough that we are confident in our findings. For all calculated values of L_p in this work the relative statistical error never exceeded 2%, and is not reported below.

4.3 Results

4.3.1 DNA L_p with and without CoHex

Despite known imperfections of modern force-fields used for all simulations in this study [74], our calculations predict the main quantity of interest – the relative change of persistence length $\Delta L_p/L_p$ – in acceptable agreement with experiment. For mixed sequence DNA the addition of neutralizing amount of CoHex yields the predicted relative change $\Delta L_p/L_p = -37\%$ (Table 4.1, bottom row) which is within the reported error margin of the corresponding experimental estimate[19] of $-44 \pm 8\%$. Furthermore, recent WAXS [192] and CD spectra [257] measurements have not detected any significant changes in DNA structure upon addition of CoHex, consistent with our results that the molecule contracts by no more than 2%, Table 4.1. Previous all-atom [32] and even coarse-grained models derived from the same atomistic potentials [131, 218] also reproduce correct quantitative dependence of DNA L_p on ionic strength.

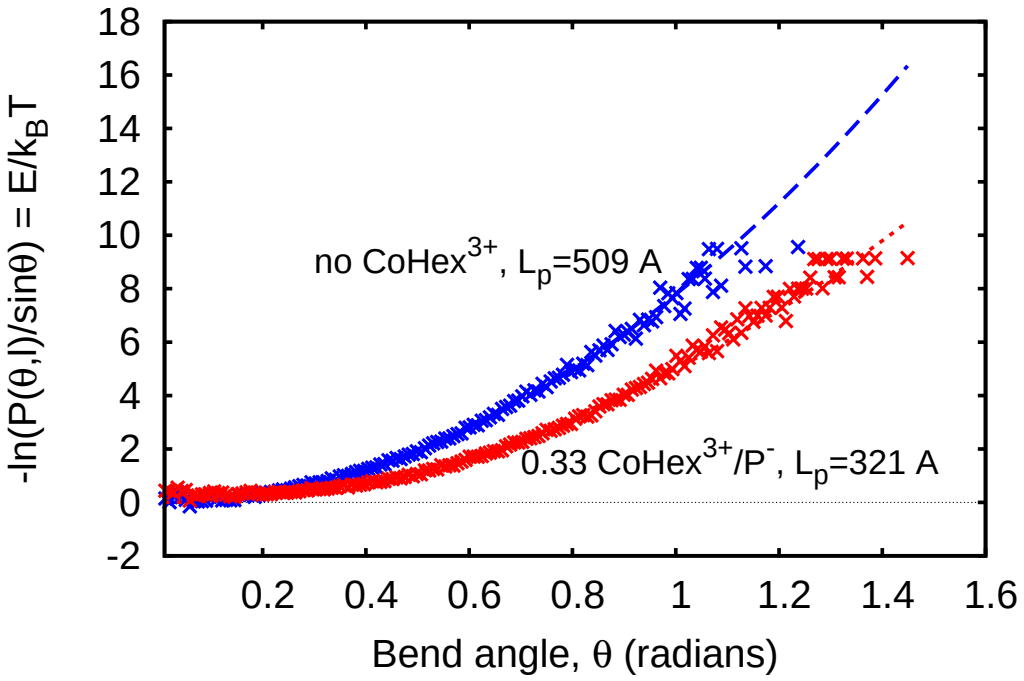

Figure 4.1: Estimation of bending persistence length L_p of DNA duplex with (red) and without (blue) $CoHex^{3+}$ counterions. L_p is estimated by fitting (dashed lines) the data points to Eq. 4.3; each point represents a value of $-\ln\left(\frac{P(\theta,l)}{\sin\theta}\right)$ averaged over ∼0.008 interval of θ.

4.3.2 RNA L_p increases with CoHex concentration

Unexpectedly, the effect of CoHex on RNA is the opposite of its effect on the DNA: the addition of $CoHex^{3+}$ results in a significantly higher L_p of the RNA, Table 4.2. The counter-intuitive increase in RNA bending rigidity caused by CoHex does not appear to be sequence-dependent: the same strong effect of similar magnitude is seen when we replace the mixed sequence RNA with a homopolymeric poly(rA)·poly(rU) fragment (first row of Table 4.3). This effect on RNA persistence length is not limited to $CoHex^{3+}$ counterions: tetravalent $Spermine^{4+}$ (a linear polyamine ion) and hypothetical trivalent "Na^{3+}" also stiffen the RNA, Table 4.3. It is also insensitive to water model and force-field choice; simulations with TIP4P water and the latest chiOL3 modifications for RNA to ff99bsc0 force-field[283] result in a negligible 3% difference in CoHex ions bound to RNA.

RNA w/ +3 and +4-valent ions	$\Delta L_p/L_p$ (%)	$\Delta l/l$ (%)
0.33 $CoHex^{3+}/P^-$ (poly(rA)·poly(rU))	+56	−14
0.33 "Na^{3+}"/P^- (mixed RNA)	+85	−16
0.25 $Spermine^{4+}/P^-$ (mixed RNA)	+29	−14

Table 4.3: $CoHex^{3+}$ increases persistence length (L_p) of homopolymeric poly(rA)·poly(rU) RNA molecules. Other tri- and tetravalent ions have the same qualitative effect: $Spermine^{4+}$ and hypothetical "Na^{3+}" increase L_p and decrease the fragment length (l) of RNA duplexes.

4.3.3 Qualitative model for the counter-intuitive effect

We propose the following qualitative explanation for this novel effect. The persistence length of a charged polymer in the presence of counterions can be conceptually decomposed into two distinct contributions: $L_p \simeq L_{int} + \Delta L_{scr}$[165]. The "base" contribution L_{int} depends on the internal structure (short-range interactions) and charge distribution of the polymer, while ΔL_{scr} represents a correction to L_{int} due to the screening by the counterions of the long-range charge-charge interactions along the polymer. Both of these contributions can significantly affect the total persistence length L_p[216]. In the case of NA, ΔL_{scr} is always negative because the screening results in a reduced effective electrostatic repulsion between the negatively charged phosphate groups. In general, DNA and RNA double helices have very similar overall structure and electrostatic properties, including linear charge density and other parameters relevant to L_p according to existing models [165, 235]. Thus, given the same amount of reduction in the charge-charge repulsion along the NA polymer due to partial neutralization by the counterions, the ΔL_{scr} term is expected to be roughly equal for DNA and RNA surrounded by the same type and amount of counterions. Indeed, increasing the concentration of monovalent ions (which are known to form loose distributions around NA molecules [217, 124, 123]) leads to similar changes in observed L_p of both DNA and RNA. However, the situation is very different with multivalent counterions. For DNA, these ions mostly bind to the externally exposed surface of the phosphate groups [257] (Fig. 4.2), leading to their partial neutralization, which mainly affects ΔL_{scr} in the expected way. In contrast, subtle structural differences between DNA and RNA helices (B- vs. A-form) lead to a more negative electrostatic potential in the major groove of RNA, which causes CoHex to bind preferentially inside the molecule, closer to the helical axis [257, 193], see Fig. 4.2. The internally and strongly bound CoHex ions can be considered as part of the internal structure of RNA itself, mainly affecting L_{int}. As sufficient number of CoHex counterions bind into the major grove of RNA, they pull the oppositely charged phosphate groups closer together. This results in the contraction of the duplex along its helical axis (Tables 4.2, 4.3), which has been experimentally confirmed [192]. In contrast, virtually no structural change occurs in DNA upon binding of the same multivalent counterions (see Table 4.1). Note that even though "Na^{3+}" is sterically about 3 times smaller than $CoHex^{3+}$, the relative contraction of the RNA double helix caused by "Na^{3+}" remains almost the same, Tables 4.2 and 4.3, suggesting that any further contraction of the RNA double helix becomes highly unfavorable energetically.

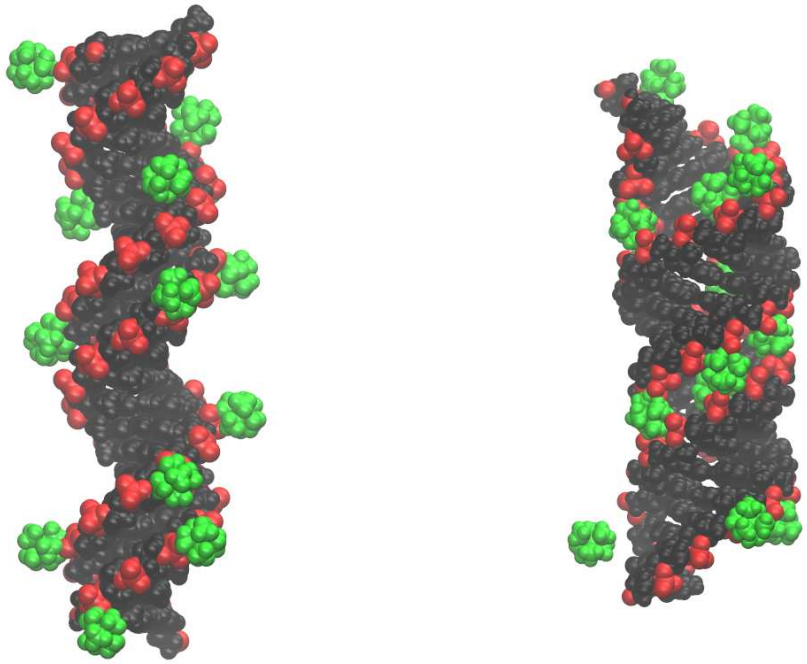

Figure 4.2: In DNA (left), multivalent counterions (illustrated for $CoHex^{3+}$, green) bind mostly externally [257] onto the negatively charged phosphates (red), with little effect on the DNA structure. The binding reduces the effective electrostatic repulsion along the helix, which in turn decreases the helix bending rigidity. In contrast, the same ions bind deep inside the RNA major groove (right) [257] causing the double helix to contract and significantly stiffen its internal structure. This leads to an over-all increase of the RNA bending rigidity – the pull of the ions works as taut bicycle spokes that tighten the wheel. The distributions (not shown) of bound $Spermine^{4+}$ around DNA and RNA are similar to the above.

The strong electrostatic pull between phosphates and buried counterions is balanced by the short-range interactions that maintain the internal structure of the double helix. Bending of a polymer implies some combination of contraction and stretching (of the opposite side of the chain), and both of these deformations require more energy in the already contracted state of the RNA compared to the original A-form conformation, which ultimately results in higher bending modulus of the polymer. Thus, the binding of multivalent counterions to the RNA helix increases L_{int} significantly, resulting in a net increase of its L_p, while in the DNA the ions mostly reduce L_p via ΔL_{scr}.

4.4 Conclusion

Our main result – that multivalent counterions can significantly increase RNA bending persistence length – is unexpected from the perspective of the completely opposite and well-known effect of increasing salt concentration on RNA's closest cousin, the DNA. The physics behind the salt effect on bending rigidity of long charged polymers appeared well understood: the charge-charge repulsion along the polymer is screened out by the counterions, making it easier to bend. But apparently, this is only one side of the story: small differences in structure between RNA and DNA can make dominant a previously unexplored consequence of counterion binding, changing the sign of the over-all salt effect on the polymer's persistence length. While the magnitude of the predicted effect might have some dependence on details of the methodology, it is clear that the effect is strong, persists over a range of ion concentrations, and is robust to ion type and sequence details, all of which should facilitate direct experimental verification. We believe that single-molecule measurements[19, 106] are best suited for studying this phenomenon. AFM studies can also shed some light on this effect, provided the interactions of the surface with the NA duplexes do not affect the binding of ions to the molecule. The new effect may manifest itself in other scenarios: mechanical properties of DNA with sequences and/or structure that have relatively stronger affinity for multivalent ions compared to canonical B-form DNA considered here may also show unexpected, or even counter-intuitive response. It is also known that under low hydration[171, 174], as well as with increasing concentrations of CoHex[10, 50], DNA can spontaneously transition from B- to A-form, and if the binding pattern of CoHex in A-form DNA is similar to that of RNA, it could be enough to offset any decrease in persistence length due to electrostatic screening and make the duplex more rigid (like in the case of RNA). Given the importance of nucleic acid stiffness in genome packing, these effects may have significant biological consequences. The influence of divalent ions such as Mg^{2+}, known to bind to nucleic acids, is also worth exploring. Perhaps most importantly, the physics of the relationship between mechanical properties of charged polymers and counterion binding is worth revisiting.

Chapter 5

Conclusion

The work presented in this thesis is focused on non-trivial mechanical properties involving nucleic acids.

While there have been many previous attempts at reconciling the classical picture of harmonic DNA bending with recent experimental evidence of increased flexibility at shorter length-scales (fragments less than 100 bp), none of them provide a general framework for polymer bending from a physical point of view. Our CHC model successfully explains the unexpectedly high cyclization probability of short DNA fragments, without redefining the well-accepted hWLC in the weak bending regime. The CHC model can be applied to other polymers besides DNA, and can even incorporate complex details such as sequence-dependence and anisotropy. Furthermore, we were able to deduce that the major contribution to non-convexity of the effective DNA bending energy function comes from the Van-der Walls interactions in the backbone of the molecule. When the polymer is bent, the atoms in the backbone that approach each other reduce their total energy from VDW interactions more than the combined increase from the other atom pairs, which counters the effects of electrostatic and bonded energy terms, that increase with bending angle. This is possible because of the non-convexity of the VDW energy function at distance greater than the VDW radii.

Nucleic acid condensation is another very important phenomenon relevant to biological processes. It has become evident recently that current models are not capable of explaining differences in condensation propensities between DNA and RNA. Even models that take into account the geometric structure of the double-helix cannot provide a complete picture; specifically, they cannot explain how similar sequence and geometry of NA duplexes can result in different condensation propensities. In this thesis we have shown that condensation propensity correlates well with the number of externally bound counterions, and outlined a quantitative model that uses the concept of internal and external binding shells to effectively explain this phenomenon.

Finally, when studying NA condensation, we have discovered an interesting and unexpected

effect: increasing concentrations of multivalent CoHex ions actually stiffen double-stranded RNA, as opposed to DNA that becomes less rigid, as predicted by electrostatics. We decided to further investigate this effect by looking at various multivalent ions, different sequences, and force-field details. We have found this counter-intuitive effect to be quite strong, and robust. A simple qualitative model was then developed to attempt to explain these results. We are quite confident in our prediction, and hope that one day another research group conducts an experiment that can help validate or disprove this claim.

Bibliography

[1] J. Abels, F. Moreno-Herrero, T. Van der Heijden, C. Dekker, and N. Dekker. Single-molecule measurements of the persistence length of double-stranded rna. *Biophysical journal*, 88(4):2737–2744, 2005.

[2] N. Agrawal, P. V. N. Dasaradhi, A. Mohmmed, P. Malhotra, R. K. Bhatnagar, and S. K. Mukherjee. Rna interference: Biology, mechanism, and applications. *Microbiology and Molecular Biology Review*, 67:657–685, 2003.

[3] D. G. Alexeev, A. A. Lipanov, and I. Y. Skuratovskii. Poly(dA).poly(dT) is a B-type double helix with a distinctively narrow minor groove. *Nature*, 325:821–823, 1987.

[4] C. Alhambra, F. J. Luque, F. Gago, and M. Orozco. Ab initio study of stacking interactions in A- and B-DNA. *J. Phys. Chem. B*, 101:3846–3853, 1997.

[5] E. Allahyarov, G. Gompper, and H. Löwen. Attraction between DNA molecules mediated by multivalent ions. *Physical Review E*, 69:041904+, 2004.

[6] H. T. Allawi, M. W. Kaiser, A. V. Onufriev, W. P. Ma, A. E. Brogaard, D. A. Case, B. P. Neri, and V. I. Lyamichev. Modeling of flap endonuclease interactions with DNA substrate. *J Mol Biol*, 328(3):537–554, May 2003.

[7] J. F. Allemand, S. Cocco, N. Douarche, and G. Lia. Loops in DNA: An overview of experimental and theoretical approaches. 19(3):293–302, 2006.

[8] K. Andresen, X. Qiu, S. A. Pabit, J. S. Lamb, H. Y. Park, L. W. Kwok, and L. Pollack. Mono- and trivalent ions around DNA: A small-angle scattering study of competition and interactions. *Biophysical Journal*, 95:287–295, 2008.

[9] A. Arnold, O. Lenz, S. Kesselheim, R. Weeber, F. Fahrenberger, D. Roehm, P. Košovan, and C. Holm. ESPResSo 3.1 — Molecular Dynamics Software for Coarse-Grained Models. In M. Griebel and M. A. Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VI*, volume 89 of *Lecture Notes in Computational Science and Engineering*, pages 1–23. Springer, 2013.

[10] P. G. Arscott, C. Ma, J. R. Wenner, and V. A. Bloomfield. Dna condensation by cobalt hexaammine (iii) in alcohol–water mixtures: dielectric constant and other solvent effects. *Biopolymers*, 36(3):345–364, 1995.

[11] N. W. Ashcroft and N. D. Mermin. *Solid State physics*. WB Saunders, Philadelphia, 1976.

[12] P. Auffinger and E. Westhof. Water and ion binding around RNA and DNA (C,G) oligomers. *J Mol Biol*, 300(5):1113–1131, Jul 2000.

[13] R. I. Babicheva, K. A. Bukreeva, S. V. Dmitriev, and K. Zhou. Discontinuous elastic strain observed during stretching of nial single crystal nanofilms. *Computational Materials Science*, 79:52–55, 2013.

[14] A. Balaeff, M. E. Churchill, and K. Schulten. Structure prediction of a complex between the chromosomal protein HMG-D and DNA. *Proteins*, 30(2):113–135, Feb 1998.

[15] A. Balaeff, L. Mahadevan, and K. Schulten. Modeling DNA loops using the theory of elasticity. *Physical Review E*, 73(3):1–23, Mar. 2006.

[16] L. Bao, X. Zhang, L. Jin, and Z.-J. Tan. Flexibility of nucleic acids: From dna to rna. *Chinese Physics B*, 25(1):18703, 2016.

[17] D. Bashford. *An object-oriented programming suite for electrostatic effects in biological molecules*, volume 1343 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin/Heidelberg, 1997.

[18] D. Bashford and D. Case. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, 51:129–152, 2000.

[19] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single dna molecules. *Proceedings of the National Academy of Sciences*, 94(12):6185–6190, 1997.

[20] V. A. Belyi and M. Muthukumar. Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academy of Sciences*, 103(46):17174–17178, 2006.

[21] P. Beroza and D. A. Case. Calculation of Proton Binding Thermodynamics in Proteins. *Methods Enzymol.*, 295:170–189, 1998.

[22] D. L. Beveridge, G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J*, 87(6):3799–3813, Dec 2004.

[23] D. L. Beveridge, P. V. Maye, B. Jayaram, G. Ravishanker, and M. Mezei. Aqueous hydration of nucleic acid constituents: Monte Carlo computer simulation studies. *J Biomol Struct Dyn*, 2(2):261–270, Oct 1984.

[24] D. L. Beveridge, K. J. McConell, R. Nirmala, M. A. Young, S. Vijayakumar, and G. Ravishanker. Molecular dynamics simulations of DNA and a protein-DNA complex including solvent. In A. Pullman, editor, *Modelling of Biomolecular Structures and Mechanisms*, pages 409–423. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995.

[25] D. L. Beveridge, M. A. Young, and D. Sprous. Modeling of DNA via molecular dynamics simulation: Structure, bending, and conformational transitions. In N. Leontis, editor, *Molecular Modeling of Nucleic Acids*, pages 260–284. American Chemical Society: Washington, D.C., 1997.

[26] M. Biswas, T. Wocjan, J. Langowski, and J. C. Smith. DNA bending potentials for loop-mediated nucleosome repositioning. *EPL (Europhysics Letters)*, 97(3):38004, Feb. 2012.

[27] V. A. Bloomfield. Condensation of DNA by multivalent cations: Considerations on mechanism. *Biopolymers*, 31(13):1471–1481, 1991.

[28] V. A. Bloomfield. DNA condensation. *Current Opinion in Structural Biology*, 6(3):334–341, 1996.

[29] V. A. Bloomfield. DNA condensation by multivalent cations. *Biopolymers*, 44(3):269–282, 1997.

[30] V. A. Bloomfield, D. M. Crothers, and I. Tinoco Jr. *Nucleic acids: structures, properties, and functions*. University Science Books, 2000.

[31] K. Boehncke, M. Nonella, K. Schulten, and A. H. Wang. Molecular dynamics investigation of the interaction between DNA and distamycin. *Biochemistry*, 30(22):5465–5475, Jun 1991.

[32] Y. J. Bomble and D. A. Case. Multiscale modeling of nucleic acids: insights into dna flexibility. *Biopolymers*, 89(9):722–731, 2008.

[33] A. Borodavka, R. Tuma, and P. G. Stockley. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences*, 109(39):15769–15774, 2012.

[34] W. H. Braunlin, C. F. Anderson, and M. T. Record. Competitive interactions of cobalt(3+)hexamine and sodium with helical B-DNA probed by cobalt-59 and sodium-23 NMR. *Biochemistry*, 26(24):7724–7731, 1987.

[35] W. H. Braunlin and Q. Xu. Hexaamminecobalt(III) binding environments on double-helical DNA. *Biopolymers*, 32(12):1703–1711, 1992.

[36] S. L. Brenner and V. A. Parsegian. A physical method for deriving the electrostatic interaction between rod-like polyions at all mutual angles. *Biophysical Journal*, 14(4):327–334, 1974.

[37] K. Bukreeva, R. Babicheva, S. V. Dmitriev, K. Zhou, and R. R. Mulyukov. Inhomogeneous elastic deformation of nanofilms and nanowires of nial and feal alloys. *JETP letters*, 98(2):91–95, 2013.

[38] D. Bumcrot, M. Manoharan, V. Koteliansky, and D. W. Sah. Rnai therapeutics: a potential new class of pharmaceutical drugs. *Nature chemical biology*, 2(12):711–719, 2006.

[39] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky. Mechanical processes in biochemistry. *Annual Review of Biochemistry*, 73(1):705–748, 2004.

[40] N. Calimet, M. Schaefer, and T. Simonson. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins*, 45:144–158, 2001.

[41] C. R. Cantor and P. R. Schimmel. *Biophysical chemistry: Part III: the behavior of biological macromolecules*. Macmillan, 1980.

[42] D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossvry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. Kollman. Amber 12. *University of California, San Francisco*, 2012.

[43] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[44] R. Chandrasekaran, A. Radha, and H.-S. Park. Sodium ions and water molecules in the structure of poly(dA).poly(dT). *Acta Crystallographica Section D: Biological Crystallography*, 51(6):1025–1035, 1995.

[45] B. Chattopadhyay and M. Mukherjee. Molecular dynamics study and electronic structure evolution of a DNA duplex $d(CCCGATCGGG)_2$. *The Journal of Physical Chemistry B*, 115(8):1760–1766, Feb. 2011.

[46] T. E. Cheatham and D. A. Case. Twenty-five years of nucleic acid simulations. *Biopolymers*, 99(12):969–977, 2013.

[47] T. E. Cheatham, P. Cieplak, and P. A. Kollman. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *Journal of Biomolecular Structure and Dynamics*, 16(4):845–862, Feb. 1999.

[48] T. E. Cheatham, M. F. Crowley, T. Fox, and P. A. Kollman. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc Natl Acad Sci U S A*, 94(18):9626–9630, Sep 1997.

[49] T. E. Cheatham and P. A. Kollman. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J Mol Biol*, 259(3):434–444, Jun 1996.

[50] T. E. Cheatham and P. A. Kollman. Insight into the stabilization of a-dna by specific ion association: spontaneous b-dna to a-dna transitions observed in molecular dynamics simulations of d[acccgcgggt] 2 in the presence of hexaamminecobalt (iii). *Structure*, 5(10):1297–1311, 1997.

[51] T. E. Cheatham and P. A. Kollman. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaamminecobalt(III). *Structure*, 5(10):1297–1311, Oct 1997.

[52] K. Chin, K. A. Sharp, B. Honig, and A. M. Pyle. Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nature Structural & Molecular Biology*, 6(11):1055–1061, Nov. 1999.

[53] J. Chocholousová and M. Feig. Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment. *J. Phys. Chem. B*, 110(34):17240–17251, Aug 2006.

[54] T. Cloutier and J. Widom. Spontaneous sharp beinding of double-stranded DNA. *Molecular Cell*, 14:355–362, 2004.

[55] T. E. Cloutier and J. Widom. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10):3645–50, Mar. 2005.

[56] B. D. Coleman, W. K. Olson, and D. Swigon. Theory of sequence-dependent DNA elasticity. *The Journal of Chemical Physics*, 118(15):7127, 2003.

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

[58] C. Cramer and D. Truhlar. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.*, 99:2161–2200, 1999.

[59] D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene. [1] DNA bending, flexibility, and helical repeat by cyclization kinetics. In J. E. D. David M.J. Lilley, editor, *DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA*, volume 212 of *Methods in Enzymology*, pages 3–29. Academic Press, 1992.

[60] L. Dai, Y. Mu, L. Nordenskiöld, and J. R. C. van der Maarel. Molecular dynamics simulation of multivalent-ion mediated attraction between DNA molecules. *Physical Review Letters*, 100(11):118301+, Mar. 2008.

[61] L. David, R. Luo, and M. Gilson. Comparison of generalized Born and Poisson models: energetics and dynamics of HIV protease. *J. Comp. Chem.*, 21:295–309, 2000.

[62] L. F. De Castro and M. Zacharias. DAPI binding to the DNA minor groove: a continuum solvent analysis. *J Mol Recognit*, 15(4):209–220, Jul-Aug 2002.

[63] H. Deng and V. A. Bloomfield. Structural effects of cobalt-amine compounds on DNA condensation. *Biophysical Journal*, 77(3):1556 – 1561, 1999.

[64] J. DeRouchey, V. A. Parsegian, and D. C. Rau. Cation charge dependence of the forces driving DNA assembly. *Biophysical journal*, 99(8):2608–2615, Oct. 2010.

[65] M. Deserno, A. Arnold, and C. Holm. Attraction and Ionic Correlations between Charged Stiff Polyelectrolytes. *Macromolecules*, 36(1):249–259, Dec. 2002.

[66] M. Deserno, C. Holm, and S. May. Fraction of Condensed Counterions around a Charged Rod: Comparison of PoissonBoltzmann Theory and Computer Simulations. *Macromolecules*, 33(1):199–206, Dec. 1999.

[67] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of dna oligonucleotides. ii: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophysical Journal*, 89(6):3721–3740, 2005.

[68] A. V. Dobrynin. Electrostatic persistence length of semiflexible and flexible polyelectrolytes. *Macromolecules*, 38(2):9304–9314, 2005.

[69] B. Dominy and C. Brooks. Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B*, 103:3765–3773, 1999.

[70] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15):5397–402, Apr. 2005.

[71] S. Edinger, C. Cortis, P. Shenkin, and R. Friesner. Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of Poisson–Boltzman equation. *J. Phys. Chem. B*, 101:1190–1197, 1997.

[72] G. Edwards, D. Hochberg, and T. W. Kephart. Structure in the electric potential emanating from DNA. *Phys. Rev. E*, 50:R698–R701, Aug 1994.

[73] E. Ennifar, P. Walter, and P. Dumas. A crystallographic study of the binding of 13 metal ions to two related RNA duplexes. *Nucleic Acids Research*, 31(10):2671–2682, May 2003.

[74] E. Fadrná, N. Špačková, R. Štefl, J. Koča, T. E. Cheatham, and J. Šponer. Molecular dynamics simulations of guanine quadruplex loops: Advances and force field limitations. *Biophysical Journal*, 87(1):227 – 242, 2004.

[75] M. Feig and B. M. Pettitt. Experiment vs force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem. B*, 101:7361–7363, 1997.

[76] M. Feig and B. M. Pettitt. A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers*, 48(4):199–209, 1998.

[77] M. Feig and B. M. Pettitt. Sodium and chlorine ions as part of the DNA solvation shell. *Biophys J*, 77(4):1769–1781, Oct 1999.

[78] A. Felts, Y. Harano, E. Gallicchio, and R. Levy. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins*, 56:310–321, 2004.

[79] A. Fenley, D. Adams, and A. Onufriev. Charge state of the globular histone core controls stability of the nucleosome. 99:1577–1585, 2010.

[80] M. O. Fenley, G. S. Manning, and W. K. Olson. Electrostatic persistence length of a smoothly bending polyion computed by numerical counterion condensation theory. *The Journal of Physical Chemistry*, 96(10):3963–3969, 1992.

[81] N. Foloppe and A. D. Mackerel. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21:86–104, 2000.

[82] M. Frank-Kamenetskii, A. Lukashin, V. Anshelevich, and A. Vologodskii. Torsional and bending rigidity of the double helix from data on small dna rings. *Journal of Biomolecular Structure and Dynamics*, 2(5):1005–1012, 1985.

[83] S. Fujii, H. Kono, S. Takenaka, N. Go, and A. Sarai. Sequence-dependent dna deformability studied using molecular dynamics simulations. *Nucleic acids research*, 35(18):6063–6074, 2007.

[84] E. Gallicchio and R. Levy. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comp. Chem.*, 25:479–499, 2004.

[85] H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins. Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers*, 85(2):115–130, 2007.

[86] R. F. Garmann, M. Comas-Garcia, A. Gopal, C. M. Knobler, and W. M. Gelbart. The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *Journal of Molecular Biology*, 426(5):1050–1060, Mar. 2014.

[87] A. Ghosh, C. Rapp, and R. Friesner. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B*, 102:10983–10990, 1998.

[88] M. K. Gilson. Theory of Electrostatic Interactions in Macromolecules. *Curr. Opin. Struct. Biol.*, 5:216–223, 1995.

[89] E. Giudice and R. Lavery. Simulations of nucleic acids and their complexes. *Acc Chem Res*, 35(6):350–357, Jun 2002.

[90] J. C. Gordon, A. T. Fenley, and A. Onufriev. An analytical approach to computing biomolecular electrostatic potential. II. Validation and applications. *The Journal of Chemical Physics*, 129:075102, 2008.

[91] J. Gosline, M. Lillie, E. Carrington, P. Guerette, C. Ortlepp, and K. Savage. Elastic proteins: biological roles and mechanical properties. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1418):121–132, 2002.

[92] L. Gosule and J. Schellman. Compact form of DNA induced by spermidine. *Nature*, 259(5541):333–335, 1976.

[93] L. Gosule and J. Schellman. DNA condensation with polyamines. I. Spectroscopic studies. *Journal of Molecular Biology*, 121(3):311–326, 1978.

[94] J. D. Gralla. Transcriptional controllessons from an e. coli promoter data base. *Cell*, 66(3):415–418, 1991.

[95] D. M. Gray, K. H. Johnson, M. R. Vaughan, P. A. Morris, J. C. Sutherland, and R. L. Ratliff. The vacuum UV CD bands of repeating DNA sequences are dependent on sequence and conformation. *Biopolymers*, 29(2):317–323, 1990.

[96] A. Y. Grosberg, A. R. Khokhlov, and L. W. Jelinski. Giant molecules: Here, there, and everywhere. *American Journal of Physics*, 65(12):1218–1219, 1997.

[97] L. Guldbrand, L. G. Nilsson, and L. Nordenskiöld. A Monte Carlo simulation study of electrostatic forces between hexagonally packed DNA double helices. *The Journal of Chemical Physics*, 85(11):6686–6698, 1986.

[98] P. Guo. The emerging field of rna nanotechnology. *Nature nanotechnology*, 5(12):833–842, 2010.

[99] P. J. Hagerman. Flexibility of dna. *Annual review of biophysics and biophysical chemistry*, 17(1):265–286, 1988.

[100] P. J. Hagerman and V. Ramadevi. Application of the method of phage t4 dna ligase-catalyzed ring-closure to the study of dna structure: I. computational analysis. *Journal of molecular biology*, 212(2):351–362, 1990.

[101] D. Hamelberg, L. McFail-Isom, L. Williams, and W. Wilson. Flexible structure of DNA: ion dependence of minor-groove structure and dynamics. *J. Am. Chem. Soc.*, 122:10513–10520, 2000.

[102] D. Hamelberg, L. D. Williams, and W. D. Wilson. Effect of a neutralized phosphate backbone on the minor groove of b-dna: molecular dynamics simulation studies. *Nucleic acids research*, 30(16):3615–3623, 2002.

[103] W. Han, M. Dlakic, Y. J. Zhu, S. M. Lindsay, and R. E. Harrington. Strained DNA is kinked by low concentrations of Zn2+. *Proceedings of the National Academy of Sciences of the United States of America*, 94(20):10565–70, Sept. 1997.

[104] G. Hawkins, C. Cramer, and D. Truhlar. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett*, 246:122–129, 1995.

[105] G. Hawkins, C. Cramer, and D. Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100:19824–19836, 1996.

[106] E. Herrero-Galán, M. E. Fuentes-Perez, C. Carrasco, J. M. Valpuesta, J. L. Carrascosa, F. Moreno-Herrero, and J. R. Arias-Gonzalez. Mechanical identities of rna and dna double helices unveiled at the single-molecule level. *Journal of the American Chemical Society*, 135(1):122–131, 2012.

[107] P. Hobza, M. Kabelac, J. Sponer, P. Mejzlik, and J. Vondrasek. Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO / M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree-Fock results. *J. Comp. Chem.*, 18:11361150, 1997.

[108] B. Honig and A. Nicholls. Classical Electrostatics in Biology and Chemistry. *Science*, 268:1144–1149, 1995.

[109] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*, 120(20):9665–9678, May 2004.

[110] D. Huang, N. Korolev, K. D. Eom, J. P. Tam, and L. Nordenskiöld. Design and biophysical characterization of novel polycationic ϵ-peptides for DNA compaction and delivery. *Biomacromolecules*, 9(1):321–330, 2008.

[111] J. Jackson. *Classical Electrodynamics*. J. Wiley & Sons, New York, 1975.

[112] H. Jacobson and W. H. Stockmayer. Intramolecular reaction in polycondensations. i. the theory of linear systems. *The Journal of Chemical Physics*, 18(12):1600–1606, 1950.

[113] P. A. Janmey. Mechanical properties of cytoskeletal polymers. *Current opinion in cell biology*, 3(1):4–11, 1991.

[114] B. Jayaram, Y. Liu, and D. Beveridge. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.*, 109:1465–1470, 1998.

[115] B. Jayaram, K. McConnell, S. B. Dixit, A. Das, and D. L. Beveridge. Free-energy component analysis of 40 protein-DNA complexes: a consensus view on the thermodynamics of binding at the molecular level. *J Comput Chem*, 23(1):1–14, Jan 2002.

[116] B. Jayaram, K. A. Sharp, and B. Honig. The electrostatic potential of B-DNA. *Biopolymers*, 28(5):975–993, May 1989.

[117] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[118] M. Kanduc, J. Dobnikar, and R. Podgornik. Counterion-mediated electrostatic interactions between helical molecules. *Soft Matter*, 5(4):868–877, 2009.

[119] M. Kanduc, A. Naji, and R. Podgornik. Counterion-mediated weak and strong coupling electrostatic interaction between like-charged cylindrical dielectrics. *Journal of Chemical Physics*, 132:224703+, 2010.

[120] S. Kasas, A. Kis, B. M. Riederer, L. Forró, G. Dietler, and S. Catsicas. Mechanical properties of microtubules explored using the finite elements method. *ChemPhysChem*, 5(2):252–257, 2004.

[121] C. Keslo and C. Simmerling. Enhanced Sampling Methods for Atomistic Simulations of Nucleic Acids. In J. Sponer and F. Lankas, editors, *Computational Studies of RNA and DNA*, volume 2 of *Challenges and Advances in Computational Chemistry and Physics*, pages 147–167, Dordrecht, The Netherlands, 2006. Springer.

[122] J. S. Kieft and I. Tinoco Jr. Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine. *Structure*, 5(5):713 – 721, 1997.

[123] S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack, and R. Elber. RNA and its ionic cloud: Solution scattering experiments and atomically detailed simulations. *Biophysical Journal*, 102(4):819–828, Feb. 2012.

[124] S. Kirmizialtin, A. R. J. Silalahi, R. Elber, and M. O. Fenley. The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophysical Journal*, 102(4):829–838, Feb. 2012.

[125] A. Kornyshev and S. Leikin. Theory of interaction between helical molecules. *The Journal of Chemical Physics*, 107:3656–3674, 1997.

[126] A. Kornyshev and S. Leikin. Electrostatic interaction between helical macromolecules in dense aggregates: An impetus for DNA poly- and mesomorphism. *Proceedings of the National Academy of Sciences of the United States of America*, 95:13579–13584, 1998.

[127] A. A. Kornyshev, D. J. Lee, S. Leikin, and A. Wynveen. Structure and interactions of biological helices. *Reviews of Modern Physics*, 79(3):943–996, Aug. 2007.

[128] A. A. Kornyshev and S. Leikin. Electrostatic zipper motif for DNA aggregation. *Physical Review Letters*, 82(20):4138–4141, 1999.

[129] A. A. Kornyshev and S. Leikin. Helical structure determines different susceptibilities of dsDNA, dsRNA, and tsDNA to counterion-induced condensation. *Biophysical Journal*, 104(9):2031–2041, May 2013.

[130] N. Korolev, N. V. Berezhnoy, K. D. Eom, J. P. Tam, and L. Nordenskiöld. A universal description for the experimental behavior of salt-(in)dependent oligocation-induced DNA condensation. *Nucleic Acids Research*, 37(21):7137–7150, Nov. 2009.

[131] N. Korolev, D. Luo, A. P. Lyubartsev, and L. Nordenskiöld. A coarse-grained dna model parameterized from atomistic simulations by inverse monte carlo. *Polymers*, 6(6):1655–1675, 2014.

[132] N. Korolev, A. P. Lyubartsev, A. Laaksonen, and L. Nordenskiöld. On the Competition between Water, Sodium Ions, and Spermine in Binding to DNA: A Molecular Dynamics Computer Simulation Study. *Biophys J*, 82(6):2860–2875, June 2002.

[133] N. Korolev, A. P. Lyubartsev, A. Laaksonen, and L. Nordenskiöld. A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations. *Nucleic Acids Research*, 31(20):5971–5981, 2003.

[134] K. M. Kosikov, A. A. Gorin, X.-J. Lu, W. K. Olson, and G. S. Manning. Bending of dna by asymmetric charge neutralization: all-atom energy simulations. *Journal of the American Chemical Society*, 124(17):4838–4847, 2002.

[135] D. Kosztin, T. C. Bishop, and K. Schulten. Binding of the estrogen receptor to DNA. The role of waters. *Biophys J*, 73(2):557–570, Aug 1997.

[136] O. Kratky and G. Porod. Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122, 1949.

[137] J. Kypr, I. Kejnovská, D. Renciuk, and M. Vorlícková. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research*, 37(6):1713–1725, 2009.

[138] F. Lankas, R. Lavery, and J. H. Maddocks. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure (London, England : 1993)*, 14(10):1527–34, Oct. 2006.

[139] F. Lankas, J. Sponer, P. Hobza, and J. Langowski. Sequence-dependent elastic properties of DNA. *Journal of molecular biology*, 299(3):695–709, June 2000.

[140] F. Lankaš, J. Šponer, J. Langowski, and T. E. Cheatham. Dna basepair step deformability inferred from molecular dynamics simulations. *Biophysical journal*, 85(5):2872–2883, 2003.

[141] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research*, 37(17):5917–5929, 2009.

[142] R. Lavery, K. Zakrzewska, D. Beveridge, T. C. Bishop, D. A. Case, T. Cheatham, S. Dixit, B. Jayaram, F. Lankas, C. Laughton, et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in b-dna. *Nucleic acids research*, 38(1):299–313, 2010.

[143] M. Lee and Y. Duan. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized Born solvent model. *Proteins*, 55:620–634, 2004.

[144] M. Lee, F. Salsbury, Jr., and C. Brooks, III. Novel generalized Born methods. *J. Chem. Phys.*, 116:10606–10614, 2002.

[145] Y. Levin. Electrostatic correlations: from plasma to biology. *Reports on Progress in Physics*, 65(11):1577, 2002.

[146] Y. Levin, J. J. Arenzon, and J. F. Stilck. The nature of attraction between like-charged rods. *Physical Review Letters*, 83(13):2680–2680, 1999.

[147] L. Li, S. A. Pabit, S. P. Meisburger, and L. Pollack. Double-stranded RNA resists condensation. *Physical Review Letters*, 106:108101, 2011.

[148] H. J. Limbach, A. Arnold, B. A. Mann, and C. Holm. ESPResSo – an extensible simulation package for research on soft matter systems. *Comp. Phys. Comm.*, 174(9):704–727, May 2006.

[149] T. A. Lionberger, D. Demurtas, G. Witz, J. Dorier, T. Lillian, E. Meyhöfer, and A. Stasiak. Cooperative kinking at distant sites in mechanically stressed dna. *Nucleic acids research*, 39(22):9820–9832, 2011.

[150] J. Lipfert, G. M. Skinner, J. M. Keegstra, T. Hensgens, T. Jager, D. Dulin, M. Köber, Z. Yu, S. P. Donkers, F.-C. Chou, et al. Double-stranded rna under force and torque: Similarities to and striking differences from double-stranded dna. *Proceedings of the National Academy of Sciences*, 111(43):15408–15413, 2014.

[151] M. S. Loth and B. I. Shklovskii. Non-mean-field screening by multivalent counterions. *Journal of Physics: Condensed Matter*, 21(42):424104+, Oct. 2009.

[152] X.-J. Lu and W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nuc. Acids Res.*, 31:5108–5121, 2003.

[153] B. Luan and A. Aksimentiev. DNA attraction in monovalent and divalent electrolytes. *Journal of the American Chemical Society*, 130(47):15754–15755, 2008.

[154] B. Luan and A. Aksimentiev. Strain softening in stretched dna. *Physical review letters*, 101(11):118101, 2008.

[155] K. Luger, A. Mäder, R. Richmond, D. Sargent, and T. Richmond. Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature*, 389:251–260, 1997.

[156] R. Luo, L. David, and M. Gilson. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comp. Chem.*, 23:1244–1253, 2002.

[157] A. P. Lyubartsev and L. Nordenskiöld. Monte carlo simulation study of ion distribution and osmotic pressure in hexagonally oriented DNA. *The Journal of Physical Chemistry*, 99(25):10373–10382, 1995.

[158] T. Macke and D. Case. Modeling unusual nucleic acid structures. Molecular Modeling of Nucleic Acids, pages 379–393. American Chemical Society, 1998.

[159] A. D. Mackerell. Empirical force fields for biological macromolecules: Overview and issues. *Journa of Computational Chemistry*, 25(13):1584–1604, Oct. 2004.

[160] J. MacKerell, A. D., J. Wiorkiewicz-Juczera, and M. Karplus. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc*, 117:11946–11975, 1995.

[161] J. D. Madura, M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon. Biological Applications of Electrostatic Calculations and Brownian Dynamics. *Rev. Comp. Chem.*, 5:229–267, 1994.

[162] V. Makarov, B. M. Pettitt, and M. Feig. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. *Acc Chem Res*, 35(6):376–384, Jun 2002.

[163] G. S. Manning. Limiting laws and counterion condensation in polyelectrolyte solutions. i. colligative properties. *The Journal of Chemical Physics*, 51(3):924–933, 1969.

[164] G. S. Manning. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly reviews of biophysics*, 11(02):179–246, 1978.

[165] G. S. Manning. The persistence length of dna is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophysical journal*, 91(10):3607–3616, 2006.

[166] B. Mansoori, S. Sandoghchian Shotorbani, and B. Baradaran. Rna interference and its role in cancer therapy. *Advanced Pharmaceutical Bulletin*, 4(4):313–321, 2014.

[167] D. Marenduzzo, E. Orlandini, A. Stasiak, D. W. Sumners, L. Tubiana, and C. Micheletti. DNADNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *Proceedings of the National Academy of Sciences*, 106(52):22269–22274, Dec. 2009.

[168] J. Marko and E. Siggia. Bending and twisting elasticity of dna. *Macromolecules*, 27(4):981–988, 1994.

[169] J. F. Marko and E. D. Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995.

[170] D. Matulis, I. Rouzina, and V. A. Bloomfield. Thermodynamics of DNA binding and condensation: isothermal titration calorimetry and electrostatic mechanism. *Journal of Molecular Biology*, 296(4):1053–1063, Mar. 2000.

[171] A. K. Mazur. Titration in s ilico of reversible b a transitions in dna. *Journal of the American Chemical Society*, 125(26):7849–7859, 2003.

[172] A. K. Mazur. Evaluation of elastic properties of atomistic dna models. *Biophysical journal*, 91(12):4507–4518, 2006.

[173] A. K. Mazur. Wormlike chain theory and bending of short dna. *Physical review letters*, 98(21):218102, 2007.

[174] A. K. Mazur. The electrostatic origin of low-hydration polymorphism in dna. *ChemPhysChem*, 9(18):2691–2694, 2008.

[175] A. K. Mazur. Local elasticity of strained dna studied by all-atom simulations. *Physical Review E*, 84(2):021903, 2011.

[176] A. K. Mazur and M. Maaloum. DNA Flexibility on Short Length Scales Probed by Atomic Force Microscopy. *Physical Review Letters*, 112(6):068104, Feb. 2014.

[177] K. J. McConnell and D. L. Beveridge. DNA structure: what's in charge? *J Mol Biol*, 304(5):803–820, Dec 2000.

[178] A. Naji, A. Arnold, C. Holm, and R. R. Netz. Attraction and unbinding of like-charged rods. *Europhysics Letters*, 67(1):130, 2004.

[179] A. Naji, S. Jungblut, A. G. Moreira, and R. R. Netz. Electrostatic interactions in strongly coupled soft matter. *Physica A: Statistical Mechanics and its Applications*, 352(1):131–170, July 2005.

[180] P. Nelson. Transport of torsional stress in DNA. *Proceedings of the National Academy of Sciences*, 96(25):14342–14347, Dec. 1999.

[181] T. T. Nguyen, A. Y. Grosberg, and B. I. Shklovskii. Screening of a charged particle by multivalent counterions in salty water: Strong charge inversion. *The Journal of Chemical Physics*, 113:1110–1125, 2000.

[182] T. T. Nguyen, I. Rouzina, and B. I. Shklovskii. Reentrant condensation of DNA induced by multivalent counterions. *The Journal of Chemical Physics*, 112(5):2562–2568, 2000.

[183] J. Norberg and L. Nilsson. NMR relaxation times, dynamics, and hydration of a nucleic acid fragment from molecular dynamics simulations. *J. Phys. Chem.*, 99:14876–14884, 1995.

[184] A. Noy, A. Pérez, M. Márquez, F. J. Luque, and M. Orozco. Structure, recognition properties, and flexibility of the DNA-RNA hybrid. *J. Am. Chem. Soc.*, 127(13):4910–4920, Mar. 2005.

[185] H. Nymeyer and A. Garcia. Free in PMC simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U. S. A.*, 100:13934–13949, 2003.

[186] T. Odijk. Polyelectrolytes near the rod limit. *Journal of Polymer Science: Polymer Physics Edition*, 15(3):477–483, 1977.

[187] A. Onufriev. Continuum electrostatics solvent modeling with the generalized Born model. In M. Feig, editor, *Modeling Solvent Environments*, pages 127–165. Wiley, USA, 2010. Chapter: public policy not applicable.

[188] A. Onufriev, D. Bashford, and D. Case. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins*, 55:383–394, 2004.

[189] F. Oosawa. Interaction between parallel rodlike macroions. *Biopolymers*, 6(11):1633–1647, 1968.

[190] F. Oosawa. *Polyelectrolytes*. Marcel Dekker, New York, 1971.

[191] B. O'Shaughnessy and Q. Yang. Manning-Oosawa counterion condensation. *Phys. Rev. Lett.*, 94:048302, Feb 2005.

[192] S. A. Pabit, A. M. Katz, I. S. Tolokh, A. Drozdetski, N. Baker, A. V. Onufriev, and L. Pollack. Understanding nucleic acid structural changes by comparing wide-angle x-ray scattering (waxs) experiments to molecular dynamics simulations. *arXiv preprint arXiv:1512.08074*, 2015.

[193] S. A. Pabit, X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack. Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. *Nucleic Acids Research*, 37(12):3887–3896, July 2009.

[194] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical Journal*, 92(11):3817–3829, June 2007.

[195] A. Pérez, J. Sponer, P. Jurecka, P. Hobza, F. J. Luque, and M. Orozco. Are the hydrogen bonds of RNA (AU) stronger than those of DNA (AT)? A quantum mechanics study. *Chemistry*, 11(17):5062–5066, Aug 2005.

[196] J. P. Peters and L. J. Maher. *DNA curvature and flexibility in vitro and in vivo.*, volume 43. Feb. 2010.

[197] G. E. Plum and V. A. Bloomfield. Equilibrium dialysis study of binding of hexammine cobalt(III) to DNA. *Biopolymers*, 27(6):1045–1051, 1988.

[198] R. Podgornik, P. L. Hansen, and V. A. Parsegian. Elastic moduli renormalization in self-interacting stretchable polyelectrolytes. *The Journal of Chemical Physics*, 113(20):9343–9350, 2000.

[199] C. Prévost, M. Takahashi, and R. Lavery. Deforming DNA: from physics to biology. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 10(9-10):1399–404, July 2009.

[200] D. Qiu, P. Shenkin, F. Hollinger, and W. Still. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, 101:3005–3014, 1997.

[201] X. Qiu, K. Andresen, J. S. Lamb, L. W. Kwok, and L. Pollack. Abrupt transition from a free, repulsive to a condensed, attractive DNA phase, induced by multivalent polyamine cations. *Phys. Rev. Lett.*, 101:228101, Nov 2008.

[202] X. Qiu, J. Giannini, S. C. Howell, Q. Xia, F. Ke, and K. Andresen. Ion Competition in Condensed DNA Arrays in the Attractive Regime. *Biophysical Journal*, 105(4):984–992, Aug. 2013.

[203] G. V. Ramanathan. Statistical mechanics of electrolytes and polyelectrolytes. III. The cylindrical PoissonBoltzmann equation. *The Journal of Chemical Physics*, 78(6):3223–3232, Mar. 1983.

[204] K. Range, E. Mayaan, L. J. Maher, and D. M. York. The contribution of phosphate-phosphate repulsions to the free energy of DNA bending. *Nucleic acids research*, 33(4):1257–68, Jan. 2005.

[205] E. Raspaud, M. Olvera de la Cruz, J. L. Sikorav, and F. Livolant. Precipitation of DNA by Polyamines: A Polyelectrolyte Behavior. *Biophysical Journal*, 74(1):381–393, Jan. 1998.

[206] D. C. Rau and V. A. Parsegian. Direct measurement of the intermolecular forces between counterion-condensed DNA double helices. evidence for long range attractive hydration forces. *Biophysical Journal*, 61(1):246–259, 1992.

[207] J. Ray and G. S. Manning. An attractive force between two rodlike polyions mediated by the sharing of condensed counterions. *Langmuir*, 10(7):2450–2461, July 1994.

[208] S. Y. Reddy, F. Leclerc, and M. Karplus. DNA polymorphism: a comparison of force fields for nucleic acids. *Biophys J*, 84(3):1421–1449, Mar 2003.

[209] T. Richmond and C. Davey. The structure of DNA in the nucleosome core. *Nature*, 423:145–150, 2003.

[210] O. Romainczyk, B. Endeward, T. F. Prisner, and J. Engels. The RNA-DNA hybrid structure determined by EPR, CD and RNase H1. *Molecular Biosystems*, 7:1050–1052, 2011.

[211] I. Rouzina and V. A. Bloomfield. Macroion attraction due to electrostatic correlation between screening counterions. 1. Mobile surface-adsorbed ions and diffuse ion cloud. *The Journal of Physical Chemistry*, 100(23):9977–9989, Jan. 1996.

[212] I. Rouzina and V. A. Bloomfield. Competitive electrostatic binding of charged ligands to polyelectrolytes: practical approach using the non-linear Poisson-Boltzmann equation. *Biophysical Chemistry*, 64(1-3):139–155, Feb. 1997.

[213] J. Z. Ruscio and A. Onufriev. A computational study of nucleosomal DNA flexibility. *Biophys J*, 91(11):4121–4132, Dec 2006.

[214] W. Saenger. *Principles of nucleic acid structure*. Springer-Verlag, New York, 1984.

[215] A. Savelyev and A. D. MacKerell. Differential Impact of the Monovalent Ions Li + , Na + , K + , and Rb + on DNA Conformational Properties. *The Journal of Physical Chemistry Letters*, pages 212–216, Dec. 2014.

[216] A. Savelyev, C. K. Materese, and G. A. Papoian. Is dnas rigidity dominated by electrostatic or nonelectrostatic interactions? *Journal of the American Chemical Society*, 133(48):19290–19293, 2011.

[217] A. Savelyev and G. A. Papoian. Electrostatic, steric, and hydration interactions favor na+ condensation around dna compared with k+. *Journal of the American Chemical Society*, 128(45):14506–14518, 2006.

[218] A. Savelyev and G. A. Papoian. Chemically accurate coarse graining of double-stranded dna. *Proceedings of the National Academy of Sciences*, 107(47):20340–20345, 2010.

[219] A. V. Savin, I. P. Kikot, M. A. Mazo, and A. V. Onufriev. Two-phase stretching of molecular chains. *Proceedings of the National Academy of Sciences*, 110(8):2816–2821, 2013.

[220] M. Scarsi, J. Apostolakis, and A. Caflisch. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A*, 101:8098–8106, 1997.

[221] M. Schaefer and M. Karplus. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.*, 100:1578–1599, 1996.

[222] R. Schleif. Dna looping. *Annual review of biochemistry*, 61(1):199–223, 1992.

[223] T. Schlick, B. Li, and W. K. Olson. The influence of salt on the structure and energetics of supercoiled dna. *Biophysical journal*, 67(6):2146, 1994.

[224] B. Schneider and H. M. Berman. Hydration of the DNA bases is local. *Biophys J*, 69(6):2661–2669, Dec 1995.

[225] Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, and M. D. Betterton. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6-7 microm. *Biophysical journal*, 93(12):4360–73, Dec. 2007.

[226] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51(7):91–97, 2008.

[227] D. E. Shaw, R. O. Dror, J. K. Salmon, J. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, et al. Millisecond-scale molecular dynamics simulations on anton. In *High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on*, pages 1–11. IEEE, 2009.

[228] G. C. Shields, C. A. Laughton, and M. Orozco. Molecular dynamics simulations of the d(tat) triple helix. *J. Am. Chem. Soc.*, 119:74637469, 1997.

[229] J. Shimado and H. Yamakawa. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules*, 17:689–698, 1984.

[230] B. I. Shklovskii. Screening of a macroion by multivalent ions: Correlation-induced inversion of charge. *Physical Review E*, 60(5):5802–5811, 1999.

[231] B. I. Shklovskii. Wigner Crystal Model of Counterion Induced Bundle Formation of Rodlike Polyelectrolytes. *Physical Review Letters*, 82:3268–3271, Apr. 1999.

[232] D. Shore, J. Langowski, and R. L. Baldwin. Dna flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences*, 78(8):4833–4837, 1981.

[233] D. Shore, J. Langowski, and R. L. Baldwin. Dna flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences*, 78(8):4833–4837, 1981.

[234] C. Simmerling, B. Strockbine, and A. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, 124:11258–11259, 2002.

[235] A. Sivolob and S. N. Khrapunov. Electrostatic contribution to the bending of dna. *Biophysical chemistry*, 67(1):85–96, 1997.

[236] J. Skolnick and M. Fixman. Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules*, 10(5):944–948, 1977.

[237] R. Soliva, F. J. Luque, C. Alhambra, and M. Orozco. Role of sugar re-puckering in the transition of A and B forms of DNA in solution. A molecular dynamics study. *J Biomol Struct Dyn*, 17(1):89–99, Aug 1999.

[238] E. Sorin, Y. Rhee, B. Nakatani, and V. Pande. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophys J*, 85(2):790–803, Aug 2003. Force field: AMBER95 Allen's stochastic (Langevin) integrator TINKER package Folding@Home GB/SA (Qui et al., 1997) TINKER.

[239] V. Spassov, L. Yan, and S. Szalma. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B*, 106:8726–8738, 2002.

[240] J. Sponer, P. Jurecka, and P. Hobza. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J Am Chem Soc*, 126(32):10142–10151, Aug 2004.

[241] J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco, and P. Hobza. Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chemistry*, 12(10):2854–2865, Mar 2006.

[242] J. E. Sponer, N. Spackova, J. Leszczynski, and J. Sponer. Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs. *J Phys Chem B*, 109(22):11399–11410, Jun 2005.

[243] D. Sprous, M. A. Young, and D. L. Beveridge. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: Theoretical considerations of the A double left right arrow B transition. *J. Phys. Chem. B*, 102:4658–4667, 1998.

[244] D. Stigter. Interactions of highly charged colloidal cylinders with applications to double-stranded DNA. *Biopolymers*, 16(7):1435–1448, 1977.

[245] D. Stigter. Evaluation of the counterion condensation theory of polyelectrolytes. *Biophys J*, 69(2):380–388, Aug. 1995.

[246] D. Stigter. An electrostatic model for the dielectric effects, the adsorption of multivalent ions, and the bending of B-DNA. *Biopolymers*, 46(7):503–516, Dec. 1998.

[247] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[248] J. Strauss and L. J. Maher. DNA bending by asymmetric phosphate neutralization. *Science*, 266:1829–1834, 1994.

[249] J. M. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys J*, 86(1 Pt 1):67–74, Jan. 2004.

[250] Z.-J. Tan and S.-J. Chen. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *The Journal of Chemical Physics*, 122(4):044903, 2005.

[251] Z.-J. Tan and S.-J. Chen. Ion-mediated nucleic acid helix-helix interactions. *Biophysical Journal*, 91(2):518–536, 2006.

[252] Z.-J. Tan and S.-J. Chen. Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length. *Biophysical Journal*, 90(4):1175–1190, 2006.

[253] Y. Tang and L. Nilsson. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys J*, 77(3):1284–1305, Sep 1999.

[254] W. H. Taylor and P. J. Hagerman. Application of the method of phage t4 dna ligase-catalyzed ring-closure to the study of dna structure: Ii. nacl-dependence of dna flexibility and helical repeat. *Journal of molecular biology*, 212(2):363–376, 1990.

[255] L. Timmons and A. Fire. Specific interference by ingested dsrna. *Nature*, 395(6705):854–854, 1998.

[256] B. A. Todd and D. C. Rau. Interplay of ion binding and attraction in DNA condensed by multivalent cations. *Nucleic Acids Research*, 36(2):501–510, Feb. 2008.

[257] I. S. Tolokh, S. A. Pabit, A. M. Katz, Y. Chen, A. Drozdetski, N. Baker, L. Pollack, and A. V. Onufriev. Why double-stranded rna resists condensation. *Nucleic Acids Research*, 42(16):10823–10831, 2014.

[258] A. Travers and J. Thompson. An introduction to the mechanics of dna. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 362(1820):1265–1279, 2004.

[259] V. Tsui and D. Case. Molecular dynamics simulations of nucleic acids using a generalized Born solvation model. *J. Am. Chem. Soc.*, 122:2489–2498, 2000.

[260] V. Tsui and D. Case. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.*, 122(11):2489–2498, 2000.

[261] V. Tsui and D. Case. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, 56:275–291, 2001.

[262] R. Vafabakhsh and T. Ha. Extreme bendability of DNA less than 100 base pairs long revealed by Single-Molecule cyclization. *Science*, 337(6098):1097–1101, 2012.

[263] K. Van de Velde and P. Kiekens. Biopolymers: overview of several properties and consequences on their applications. *Polymer Testing*, 21(4):433–442, 2002.

[264] M. van Dijk and A. M. Bonvin. 3D-DART: a DNA structure modelling server. *Nucleic Acids Research*, 37(suppl 2):W235–W239, 2009.

[265] M. Vologodskaia and A. Vologodskii. Contribution of the intrinsic curvature to measured dna persistence length. *Journal of molecular biology*, 317(2):205–213, 2002.

[266] A. Vologodskii and M. D. Frank-Kamenetskii. Strong bending of the dna double helix. *Nucleic acids research*, 41(14):6785–6792, 2013.

[267] L. Wang, B. E. Hingerty, A. R. Srinivasan, W. K. Olson, and S. Broyde. Accurate representation of B-DNA double helical structure with implicit solvent and counterions. *Biophys J*, 83(1):382–406, Jul 2002.

[268] T. Wang and R. Wade. Implicit solvent models for flexible protein-protein docking by molecular dynamics simulation. *Proteins*, 50:158–169, 2003.

[269] J. D. Watson, F. H. Crick, et al. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[270] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield. Salt dependence of the elasticity and overstretching transition of single dna molecules. *Biophysical journal*, 82(6):3160–3169, 2002.

[271] B. Widom. *Statistical Mechanics*. Cambridge U. P., Cambridge, 2002.

[272] J. Widom and R. L. Baldwin. Cation-iduced toroidal condensation of DNA. Studies with $Co^{3+}(NH_3)_6$. *Journal of Molecular Biology*, 144:431–453, 1980.

[273] P. Wiggins and P. Nelson. Generalized theory of semiflexible polymers. *Physical Review E*, 73(3):1–13, Mar. 2006.

[274] P. A. Wiggins, T. Van Der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. High flexibility of dna on short length scales probed by atomic force microscopy. *Nature nanotechnology*, 1(2):137–141, 2006.

[275] D. P. Wilson, A. V. Tkachenko, and J. C. Meiners. A generalized theory of DNA looping and cyclization. *EPL (Europhysics Letters)*, pages 58005+, 2010.

[276] R. W. Wilson and V. A. Bloomfield. Counterion-induced condensation of deoxyribonucleic acid. A light-scattering study. *Biochemistry*, 18(11):2192–2196, May 1979.

[277] Y.-Y. Wu, Z.-L. Zhang, J.-S. Zhang, X.-L. Zhu, and Z.-J. Tan. Multivalent ion-mediated nucleic acid helix-helix interactions: RNA versus DNA. *Nucleic Acids Research*, 43(12):6156–6165, May 2015.

[278] B. Xia, V. Tsui, D. Case, H. Dyson, and P. Wright. Comparison of protein solution structures refined by molecular dynamics simultion in vacuum, with a generalized Born model, and with explicit water. *J. Biomol. NMR*, 22:317–331, 2002.

[279] J. Yoo and A. Aksimentiev. Competitive binding of cations to duplex DNA revealed through molecular dynamics simulations. *The Journal of Physical Chemistry B*, 116(43):12946–12954, Sept. 2012.

[280] J. Yoo and A. Aksimentiev. Improved parametrization of Li+, Na+, K+, and Mg2+ ions for all-atom molecular dynamics simulations of nucleic acid systems. *The Journal of Physical Chemistry Letters*, 3(1):45–50, 2012.

[281] M. Zacharias. Continuum Solvent Models To Study The Structure and Dynamics of Nucleic Acids And Complexes with Ligands. In J. Sponer and F. Lankas, editors, *Computational Studies of RNA and DNA*, volume 2 of *Challenges and Advances in Computational Chemistry and Physics*, pages 95–119, Dordrecht, The Netherlands, 2006. Springer.

[282] K. Zakraewska, A. Madami, and R. Lavery. Poisson-Boltzmann calculations for nucleic acids and nucleic acid complexes. *Chem. Phys.*, 204:263–269, 1996.

[283] M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham III, and P. Jurecka. Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of chemical theory and computation*, 7(9):2886–2902, 2011.

Appendix A

AMBER MD protocols

A.1 Simulating restrained DNA loops of various sizes

For a loop of 100 bp:

```
minimize the structure
 &cntrl
   imin=1, maxcyc=1000, ncyc=500, cut=999.0, igb=1, saltcon=0.28, ntb=0, ntr=1,
 &end
 restrain all P atoms
 1.0
FIND
P * * *
SEARCH
RES 1 200
END
END

run restrained heating simulation
 &cntrl
   imin=0,ntb=0,ntf=2,ntc=2,tol=0.000001,ntr=1,igb=1,saltcon=0.28,ntpr=50,ntwx=500,
   ntt=3,gamma_ln=1.0,tempi=0.0,temp0=300.0,nstlim=50000,dt=0.002,cut=999,
 &end
 restrain all P atoms
 1.0
FIND
P * * *
SEARCH
```

```
RES 1 200
END

run dynamics simulation
 &cntrl
   imin=0,ntb=0,ntf=2,ntc=2,tol=0.000001,ntr=1,igb=1,saltcon=0.28,ntpr=10,ntwx=500,
   ntt=3,gamma_ln=2.0,tempi=300.0,temp0=300.0,nstlim=500000,dt=0.002,cut=999,
 &end
 restrain all P atoms
 0.1
FIND
P * * *
SEARCH
RES 1 200
END
END
```

A.2 Unrestrained long DNA fragments in implicit solvent

For a 150 bp long DNA with $\gamma_{ln} = 0.001$:

```
minimize the structure
 &cntrl
   imin=1, maxcyc=1000, ncyc=500, cut=999.0, igb=1, saltcon=0.28, ntb=0, ntr=1,
 &end
 fix the dna
 1.0
RES 1 300
END
END

run restrained heating simulation
 &cntrl
   imin=0,ntb=0,ntf=2,ntc=2,tol=0.000001,
   ntx=5,ig=-1,
   ntr=1,igb=1,saltcon=0.28,ntpr=50,ntwx=500,
   ntt=3,gamma_ln=1.0,tempi=0.0,temp0=300.0,
   nstlim=55555,dt=0.0018,cut=999,
```

```
 &end
 fix the dna
 0.1
RES 1 300
END
END

run dynamics simulation
 &cntrl
   imin=0,ntb=0,ntf=2,ntc=2,tol=0.000001,
   ntr=0,igb=1,saltcon=0.28,
   ntpr=5000,ntwx=5000,ig=-1,ntx=5,
   ntt=3,gamma_ln=0.001,tempi=300.0,temp0=300.0,
   nstlim=55555555,dt=0.0018,cut=999,
 &end
```

Appendix B

Coarse-grained MD using Espresso

B.1 DNA loops with experimentally-derived bending potential

At T=300K:

```
#!/bin/sh
# tricking... the line after a these comments are interpreted as standard shell script \
    exec $ESPRESSO_SOURCE/Espresso $0 $*

set tcl_precision 6
source ./subs.tcl
#source ../scripts/ABHmath.tcl

#############################################################
# system name
#############################################################

set n_part 80
set name  "kbt_tab_ring$n_part"

# integration
set int_steps    100
set int_n_times  100000

puts " "
puts "=========================================================="
puts "        $name.tcl                                         "
```

```
puts "=========================================================="
puts " "

############################################################
#output files
############################################################
set obsv_file [open "$name.obs" "w"]
set traj_file [open "$name.vtf" "w"]
set cord_file [open "$name.vcf" "w"]
set stru_file [open "$name.vsf" "w"]

puts "opened output files\n"

############################################################
#  Setup System Box
############################################################
set  box_l      100
setmd box_l     $box_l $box_l $box_l
setmd periodic  0 0 0
#setmd periodic  1 1 1

puts "set up simulation box with length $box_l\n"
############################################################
#particles and constraints
############################################################

set PI 3.141592653589793
set b_length 1
set r_circ [expr $n_part*$b_length/(2*$PI)]
set d_theta [expr 2*$PI/$n_part]

# Create the particles
for {set i 0} { $i < $n_part } {incr i} {
set cur_angle [expr $d_theta*$i]
set posx [expr $r_circ*cos($cur_angle)]
set posy [expr $r_circ*sin($cur_angle)]
set posz 0.0
part $i pos $posx $posy $posz type 0 mass 1
}

puts "placing particles and constraints"
puts "particle [part 0]"
```

```
puts "particle [part 1]"
puts "...\n"

#############################################################
#interactions between all types of objects
#############################################################

#set pos_tol 0.01
#set vel_tol 0.01
set a_coef 18.6
set b_coef 10000

# The harmonic bond potential
inter 0 harmonic $b_coef $b_length
#inter 0 rigid_bond b_length pos_tol vel_tol

# The harmonic angle potential
#inter 1 angle $a_coef

# Tabulated angle potential
inter 1 tabulated angle "vol_yscale.dat"

# Fix COM
inter 0 0 comfixed 1

puts "Defined Interactions:\n[inter]\n"

#############################################################
#define bonds between specific objects
#############################################################

# Bond the particles
for {set i 0} { $i < [expr $n_part-1] } { incr i } {
part $i bond 0 [expr $i+1]
}
part [expr $n_part-1] bond 0 0

# Add the angle interactions
for {set i 1} { $i < [expr $n_part-1] } { incr i } {
part $i bond 1 [expr $i-1] [expr $i+1]
}
part 0 bond 1 [expr $n_part-1] 1
```

```
part [expr $n_part-1] bond 1 [expr $n_part-2] 0

puts "Defined bonds\n"

#############################################################
# integration parameters
#############################################################

#md parameteres
setmd time_step 0.0045
setmd skin      .5
thermostat langevin 0.596 0.5
setmd max_num_cells 2744

puts "Set md integration parameters\n"

#############################################################
#simulation integration
#############################################################

puts "\n\n____________Starting simulation integration: outputting [expr $int_n_times/$in

writevsf $traj_file
writevsf $stru_file
writevcf $traj_file

#inter ljforcecap 0

for {set i 0} { $i < $int_n_times } { incr i} {

    integrate $int_steps

    writevcf $traj_file
    writevcf $cord_file

    puts -nonewline "integration step [expr $i+1] out of $int_n_times\r"

    puts $obsv_file "{ time [setmd time] } [analyze energy]"

}

puts "\n\n_________________________finished simulation integration_____________________
```

```
#################################################################
#end simulation
#################################################################

# close files
close $obsv_file
close $traj_file
close $cord_file
close $stru_file

# terminate program
puts "\n\n=================================Finished=================================
exit
```

Energy-minimized (example for $L = 60$ bp):

```
#!/bin/sh
# tricking... the line after a these comments are interpreted as standard shell script \
    exec $ESPRESSO_SOURCE/Espresso $0 $*

set tcl_precision 6

#################################################################
# system name
#################################################################

set n_part 60
set name  "ann_ring$n_part"

# integration
set int_n_times  100
set heat_steps   1000
set cool_steps  100000
set int_steps  5

puts " "
puts "======================================================="
puts "        $name.tcl                                      "
puts "======================================================="
puts " "
```

```
#############################################################
#output files
#############################################################
set obsv_file [open "$name.obs" "w"]
set traj_file [open "$name.vtf" "w"]
set cord_file [open "$name.vcf" "w"]
set stru_file [open "$name.vsf" "w"]

puts "opened output files\n"

#############################################################
#  Setup System Box
#############################################################
set  box_l      500
setmd box_l     $box_l $box_l $box_l
setmd periodic  0 0 0
#setmd periodic  1 1 1

puts "set up simulation box with length $box_l\n"
#############################################################
#particles and constraints
#############################################################

set PI 3.141592653589793
set b_length 1
set r_circ [expr $n_part*$b_length/(2*$PI)]
set d_theta [expr 2*$PI/$n_part]

# Create the particles
for {set i 0} { $i < $n_part } {incr i} {
set cur_angle [expr $d_theta*$i]
set posx [expr $r_circ*cos($cur_angle)]
set posy [expr $r_circ*sin($cur_angle)]
set posz 0.0
part $i pos $posx $posy $posz type 0 mass 1
}

puts "placing particles and constraints"
puts "particle [part 0]"
puts "particle [part 1]"
puts "...\n"
```

```
############################################################
#interactions between all types of objects
############################################################

#set pos_tol 0.01
#set vel_tol 0.01
set a_coef 18.6
set b_coef 1000

# The harmonic bond potential
inter 0 harmonic $b_coef $b_length
#inter 0 rigid_bond b_length pos_tol vel_tol

# The harmonic angle potential
#inter 1 angle $a_coef

# Tabulated angle potential
inter 1 tabulated angle "vol_yscale.dat"

# Fix COM
inter 0 0 comfixed 1

puts "Defined Interactions:\n[inter]\n"

############################################################
#define bonds between specific objects
############################################################

# Bond the particles
for {set i 0} { $i < [expr $n_part-1] } { incr i } {
part $i bond 0 [expr $i+1]
}
part [expr $n_part-1] bond 0 0

# Add the angle interactions
for {set i 1} { $i < [expr $n_part-1] } { incr i } {
part $i bond 1 [expr $i-1] [expr $i+1]
}
part 0 bond 1 [expr $n_part-1] 1
part [expr $n_part-1] bond 1 [expr $n_part-2] 0
```

```
puts "Defined bonds\n"

#################################################################
# integration parameters
#################################################################

#md parameteres
setmd time_step 5e-5
setmd skin      .5

set temp 1.0
set gamma 0.01

thermostat langevin $temp $gamma
setmd max_num_cells 2744

puts "Set md integration parameters\n"

#################################################################
#simulation integration
#################################################################

puts "\n\n____________Starting simulation integration: outputting [expr $int_n_times/$in

writevsf $traj_file
writevsf $stru_file
writevcf $traj_file

#inter ljforcecap 0

set cool_step [expr $temp/$cool_steps]
set heat_step [expr $temp/$heat_steps]

for {set i 0} { $i < $int_n_times } { incr i} {

    # Warm-up steps
    for {set j 0} { $j < $heat_steps } {incr j} {
        set ntemp [expr $j*$heat_step]
        thermostat langevin $ntemp $gamma

        integrate 10
    }
```

```
    # Cooling steps
    for {set j 0} { $j < $cool_steps } {incr j} {
        set ntemp [expr $temp - $j*$cool_step]
        thermostat langevin $ntemp $gamma

        integrate $int_steps
    }

    writevcf $traj_file
    writevcf $cord_file

    puts -nonewline "annealing run [expr $i+1] out of $int_n_times\r"

    puts $obsv_file "{ time [setmd time] } [analyze energy]"

}

puts "\n\n_________________________finished simulation integration_______________________


############################################################
#end simulation
############################################################

# close files
close $obsv_file
close $traj_file
close $cord_file
close $stru_file

# terminate program
puts "\n\n==================================Finished==================================
exit
```

B.2 Confinement of charged polymer to an oppositely charged cylinder

```
#!/bin/sh
# tricking... the line after a these comments are interpreted as standard shell script \
    exec $ESPRESSO_SOURCE/Espresso $0 $*
```

```
set tcl_precision 6
#source ./subs.tcl
#source ../scripts/ABHmath.tcl

#################################################################
# system name
#################################################################

set PI 3.141592653589793

set wlc  0
set angon 1

set n_part  149

set qlarge 70.0
set qdelta 10.0
set qint 10
set qsmall -1

set turns 1.7
set ttw [expr 2*$turns*$PI]

set equil_steps 1000000

if { $wlc == 1 } {
set tname "wlc"
} elseif { $wlc == 0 } {
set tname "chc"
} else { return 0 }
set name  "${tname}_line"

set pot_file_name "vol_coul_scaled.dat"

# integration
set int_steps    1000
set int_n_times  5000

puts " "
puts "======================================================="
puts "        $name.tcl                                      "
```

```
puts "============================================================"
puts " "

#############################################################
#output files
#############################################################
#set obsv_file [open "$name.obs" "w"]
set traj_file [open "$name.vtf" "w"]
#set cord_file [open "$name.vcf" "w"]
set stru_file [open "$name.vsf" "w"]

puts "opened output files\n"

#############################################################
#  Setup System Box
#############################################################
set  box_l      500
setmd box_l     $box_l $box_l $box_l
setmd periodic  0 0 0

puts "set up simulation box with length $box_l\n"
#############################################################
#particles and constraints
#############################################################

set b_length 1.0
set r_sphere [expr $n_part*$b_length/($ttw)]
set z_init [expr -$n_part*$b_length/2.0]
set xy_init [expr $r_sphere + $b_length]
set c_scale 18.0
set c_length [expr $c_scale*$b_length/2.0]

# Create the particles
for {set i 0} { $i < $n_part } {incr i} {
set posx 0.0
set posy $xy_init
set posz [expr $z_init + $b_length*$i]
part $i pos $posx $posy $posz q $qsmall type 0 mass 1
if { $i > 1 } {
part $i exclude [expr $i-1] }
# if { $i < $n_part-1 } {
# part $i exclude [expr $i+1] }
```

```
# for {set j 0} {$j < $i} {incr j} {
# part $i exclude $j
# }
}

part $n_part pos 0.0 0.0 0.0 q $qlarge type 1 mass 100

#constraint sphere center 0.0 0.0 0.0 radius $r_sphere direction 1 type 2 penetrable 0 r
constraint cylinder center 0.0 0.0 0.0 axis 1.0 0.0 0.0 radius $r_sphere length [expr 1.
constraint wall normal -1.0 0.0 0.0 dist [expr -$c_length] type 3 penetrable 0 reflectin
constraint wall normal 1.0 0.0 0.0 dist [expr -$c_length] type 3 penetrable 0 reflecting

puts "placing particles and constraints"
puts "particle [part 0]"
puts "particle [part 1]"
puts "...\n"

############################################################
#interactions between all types of objects
############################################################

#set pos_tol 0.01
#set vel_tol 0.01

set a_coef 92.0
set b_coef  500.0
set dna_r 1.0
set eps 2.0
set sig  [expr 0.89*$r_sphere]
set csig 1.0
set l_b  2.0
set kappa  0.05
set r_cut  [expr 2.0**(1/6.0)*$csig]
#[expr $box_l/3.0]
#[expr 2.1*$r_sphere]
set alpha 1.0
set kdih  1.0

# The harmonic bond potential
inter 0 harmonic $b_coef $b_length

# The harmonic angle potential
```

```
if { $wlc == 1 } {
inter 1 angle $a_coef }

# Tabulated angle potential
if { $wlc == 0 } {
inter 1 tabulated angle $pot_file_name }

# Dihedral
#inter 2 dihedral 2 $kdih 0.0

# Fix COM
inter 1 1 comfixed 1

# Lennard-Jones
inter 0 2 lennard-jones 0.0 $csig $r_cut
inter 0 3 lennard-jones 0.0 $csig $r_cut
#inter 0 0 lennard-jones 0.0 $dna_r 8.0
#auto 0.0 0.0 [expr $dna_r-$b_length]

# Soft Sphere
#inter 0 1 soft-sphere $asphere -2 [expr $box_l/10] $r_sphere

# Morse interaction
#inter 0 1 morse $eps $asphere $r_sphere $r_cut
#inter 0 2 morse $eps $alpha 0.0 $r_cut

# Coulomb
inter coulomb $l_b dh $kappa [expr 2.0*$r_sphere]

puts "Defined Interactions:\n[inter]\n"

#############################################################
#define bonds between specific objects
#############################################################

# Bond the particles
for {set i 0} { $i < [expr $n_part-1] } { incr i } {
part $i bond 0 [expr $i+1]
}

# Add the angle interactions
if { $angon == 1} {
```

```
for {set i 1} { $i < [expr $n_part-1] } { incr i } {
part $i bond 1 [expr $i-1] [expr $i+1]
}
}

# Dihedral bonds
#for {set i 1} { $i < [expr $n_part-2] } { incr i } {
# part $i bond 2 [expr $i-1] [expr $i+1] [expr $i+2]
#}
#part [expr $n_part-2] bond 2 [expr $n_part-4] [expr $n_part-3] [expr $n_part-1]

puts "Defined bonds\n"

#############################################################
# integration parameters
#############################################################

#md parameteres
setmd time_step 1e-4
setmd skin      .5
thermostat langevin 1.0 0.01
setmd max_num_cells 2744

puts "Set md integration parameters\n"

#############################################################
#  Warmup Integration
#############################################################

set warm_steps   100
set warm_n_times 300
set min_dist     0.1
set act_min_dist [analyze mindist]

puts "\n\n_____________starting warmup with $warm_steps integration steps $warm_n_times

set LJ cap
set cap 1
inter forcecap $cap

# Warmup Integration Loop
set i 0
```

```
while { $i < $warm_n_times && $act_min_dist < $min_dist } {

    integrate $warm_steps

    # Warmup criterion
    set act_min_dist [analyze mindist]
    puts "run $i at time=[setmd time] min dist = $act_min_dist\r"
    flush stdout

    # write observables
#    puts $obsv_file "{ time [setmd time] } [analyze energy]"

    # Increase LJ cap
    set cap [expr $cap+10]
    inter forcecap $cap
    incr i
}

inter forcecap 0
#e4

puts "\n_____________________________Warmed up system_____________________________\

# Equilibration
integrate [expr $equil_steps]

puts "\n ____________________Equilibrated system_____________________________\n"

#############################################################
#simulation integration
#############################################################

puts "\n\n___________Starting simulation integration: outputting $int_n_times frames___

writevsf $traj_file
writevsf $stru_file
close $stru_file
writevcf $traj_file
#writevcf $cord_file


for {set j 0} { $j < $qint + 1 } { incr j } {
```

```
set qnow [expr $qlarge + $j*$qdelta]
part $n_part q $qnow

set obsv_file [open "$name.q$qnow.obs" "w"]
set cord_file [open "$name.q$qnow.vcf" "w"]

for {set i 0} { $i < $int_n_times } { incr i } {

   puts -nonewline "Running (run [expr $i+1]/$int_n_times)\r"
   integrate $int_steps

   writevcf $cord_file
   writevcf $traj_file

   puts $obsv_file "{ time [setmd time] } [analyze energy]"
}

close $obsv_file
close $cord_file
}

puts "\n\n_____________________finished simulation integration_____________________


############################################################
#end simulation
############################################################

# close files
#close $obsv_file
close $traj_file
#close $cord_file
#close $stru_file

# terminate program
puts "\n\n==================================Finished==================================
exit
```