A coarse grained model of DNA and supercoiling dynamics

by

Murat Öztürk

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computational Sciences & Engineering

Koç University

August, 2013

"It is a mistake to think you can solve any major problems just with potatoes."

Douglas Adams, 'Life, the Universe and Everything'

ACKNOWLEDGMENTS

I would like to take the opportunity to thank my supervisors Mehmet Sayar and Alkan Kabakçıoğlu for inspiring and guiding me throughout this work, my colleagues Cahit Dalgıçdır and Beytullah Özgür for fruitful conversations, and my love, Özge, for being there.

My work for this thesis was thankfully funded by TÜBİTAK under project code TBAG-110T618.

ABSTRACT

DNA polymers are the primary containers of information in all living organisms. In addition to its fundamental role as a biomolecule, DNA has technological applications as a building block for nanostructures. Even though molecular structure of DNA was established as a double helix in 1953, very little about its dynamical properties were discovered during the following decades. With the advent of single molecule experiments, probing dynamical processes in single DNA molecules has been possible in the last decades. Although recent experiments provide unprecedented insight, they have limited resolution in both time and length domains. Molecular simulations, on the other hand, provide complete knowledge of the system but are limited by computational cost. Coarse grained models aim to decrease the cost of molecular simulations while preserving most of the resolution. In this work, we built a coarse grained model to represent both single and double stranded DNA strands with high enough performance to simulate biologically relevant time and length scales.

Supercoiling is a mechanism where a twist storing polymer such as DNA expresses torsional stress through conformational changes. Supercoiling is of interest from a biological point of view as it is readily observed in genome packing schemes and presumably affects gene expression patterns. A recent single molecule experiment provided indirect observations on the supercoil dynamics of overwound DNA. In this experiment, supercoiled regions along the length of DNA demonstrated fast hopping behaviour in addition to diffusive motion. In this work, we reproduced this experimental scenario in silico using a coarse grained model. We simulated up to 1 kilobase long overwound DNA strands under tension and observed time evolution of the conformation of the molecule. We produced supercoiled conformations associated with local density peaks observed in the experiment and time evolution of curvature along the length. We varied the degree of overwinding (σ) and the amount of tension and observed a buckling transition at different threshold σ values. We reproduced the experimentally observed hopping behaviour in addition to diffusive motion near the threshold σ values.

ÖZETÇE

DNA polimerleri tüm canlılardaki temel bilgi saklama mekanizmasını oluşturmaktadır. Biyolojik fonksiyonunun yanında, DNA moleküllerinin teknolojik olarak nano yapıların üretiminde yapıtaşı olarak kullanılması da mümkündür. DNA'nın çift sarmallı moleküler yapısı 1953'de keşfedilmesine rağmen, zamana bağlı davranışı hakkında çok az şey bilinmektedir. Son yıllarda, tek molekül deneylerinin de gelişmesiyle bu alanda örneği görülmemiş çalışmalar yapılmasina rağmen, deneylerin çözünürlüğü sınırlı kalmaktadır. DNA üzerinde moleküler simulasyonlar yapmak mümkün olsa da, bunlar hesaplama maliyeti nedeniyle zaman ve uzunluk ölçeklerinde sınırlı kalmaktadır. Düşük çözünürlüklü modeller moleküler simulasyonların hesaplama maliyetini düşürürken, gerçekçiliğini korumayı amaçlamaktadır. Biz de bu çalışmda hem tek hem de çift iplikli DNA moleküllerini temsil edebilecek düşük çözünürlüklü bir DNA modeli öneriyoruz.

Süpersarmallanma, burkulabilir bir polimer olan DNA'nın üzerindeki burkulmaya yönelik gerilimi saklamak için kullandığı bir mekanizmadır. Süpersarmallanma, birçok organizmanın genom paketleme esnasında kullandığı bir mekanizma olup, gen ifadesinde de etkili bir rol oynadığı düsünülmektedir. Yakın zamanda yapılan bir tek molekül deneyi, fazla burkulmuş bir DNA molekülünün süpersarmallanması üzerine yeni gözlemler önermiştir. Bu deneyde, fazladan burkulmuş bir DNA molekülünun üzerindeki süpersarmallı bölgelerin difüzyonun yanında daha hızlı hoplama hareketleri de sergilediği gözlenmiştir. Biz de bu çalısmada bu deneyi düşük çözünürlüklü bir DNA modeli kullanarak bilgisayar ortamında yeniden canlandırdık. Böylece deneyde gözlemlenmesi mümkün olmayan süpersarmallanmış DNA yapılarını gözlemledik. Bunun yanında, denyde gözlemlenen hoplama hareketlerini yeniden canlandırıp deneylerde ölcülemeyen bazı çoklukları da ölcerek, bir hoplama mekanizması önermeyi amaçladık.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

DNA	deoxyribonucleic acid
ssDNA	single stranded DNA
dsDNA	double stranded DNA
MD	molecular dynamics
CG	coarse grained
topology	connectivity of particles represented in a model
plectoneme	a loop of helix, a unit geometrical structure in a supercoiled ribbon
bp	base pairs, the DNA monomer
n	length of DNA in base pairs
$\vec{r},\ \dot{r},\ \ddot{r}$	position, velocity, acceleration vectors, $\dot{r} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}$, $\ddot{r} = \frac{\mathrm{d}\dot{r}}{\mathrm{d}t}$
$\theta,\ \phi$	angle, dihedral angle
$H(x)$	histogram for random variable x
$P(x)$	probability density function for random variable x
$< x >$	time average for random variable x
$C(\tau)$	correlation function with respect to τ
k_B	Boltzmann constant, 0.008314 $\frac{kJ}{mol \cdot K}$
T	temperature in Kelvins
$V(x)$	potential function for variable x
$\lambda,\ \lambda_0$	pitch, reference λ for a relaxed system
l_p	persistence length
R_{e2e}	end-to-end distance
$Ln,\ Ln_0$	linking number, reference Ln for a relaxed system
σ	a measure of excess link, $\frac{\Delta Ln}{Ln_0} = \frac{Ln}{Ln_0} - 1$
$Tw,\ \omega$	twist, angular twist rate
δTw	a measure of excess twist , $\frac{\Delta Tw}{Ln_0} = \frac{Tw}{Ln_0} - 1$
$\delta Tw'$	normalized δTw with respect to writhe bias of the model
Wr	writhe
κ	curvature

INTRODUCTION

DNA is a polymer that is the primary container of information in all living organisms. The molecular structure of this polymer was discovered in 1953 [1] and its dynamics has been a subject of interest ever since. There are a myriad of biological processes that involve manipulation of DNA molecules, yet very little is known about the actual dynamics. Fundamental biological questions such as gene regulation and genome packing remain open and directly related to DNA dynamics.

DNA molecules have a molecular recognition mechanism embedded in their structure that is both precise and reliable. A great selection of molecular machinery to synthesize and manipulate DNA is also readily available from biological sources in the form of enzymes. Such a system being extremely infeasible to design and implement from scratch, DNA has become a popular building block for nanostructures. Self assembly of both two and three dimensional nanostructures, in addition to functional nanomachines have been realised through the use of DNA in recent years [2, 3, 4]. From an informatics point of view, proof of concept molecular computers were proposed and implemented using DNA molecules [5]. Development of such technological applications would benefit greatly from further understanding of dynamics of DNA.

The last two decades has brought the advent of apparatus such as scanning probe microscopes, optical traps and magnetic tweezers, resulting in the possibility of experiments involving single DNA molecules. It has been possible for example to measure mechanical properties of a single DNA molecule, or forces involved in a transcription event [6]. In a recent experiment, supercoil formation and diffusion processes on a single DNA molecule were observed with millisecond time resolution [7]. Such experiments, despite providing unprecedented insight into DNA dynamics, remain limited in resolution as they typically yield indirect observations.

DNA has also been a subject of interest for theoretical studies. Simpler analytical models such as the Worm Like Chain model represented DNA with considerable success, predicting properties like entropic elasticity in good agreement with experiments. More complicated analytical models were also developed to yield further insight into statistics

of the molecule under various conditions [8]. Numerical methods such as molecular simulations are however indispensable for studying dynamical properties of DNA as they provide time evolution data for a given system. Molecular Dynamics (MD) simulations of atomistic DNA models are straightforward but severely limited by computational cost. Coarse grained models of DNA were therefore built by our group [9] and elsewhere [10, 11, 12, 13, 14] attempting to reduce the computational cost of such simulations.

This document is divided into two Chapters, each representing roughly a year's work done in partial fulfillment of the requirements of a Master of Science degree. In the first Chapter, we attempted to build a coarse grained model of DNA based on structural modeling principles. We aimed to build a model that faithfully represents both single and double stranded DNA molecules, with high enough performance to simulate biologically and technologically relevant time and length scales.

In the second Chapter, reverting to a simpler model that was previously built within our group [9], we reproduced a recent single molecule experiment [7] in silico. The original experiment was focused on supercoil dynamics and relied on indirect observations, resulting in limited resolution. We reproduced the experimental observations and extracted information that is not available to the experimentalist, yielding unique insight into supercoil dynamics.

Chapter 1

A COARSE GRAINED MODEL OF DNA

1.1 Introduction

Molecular dynamics (MD) is an established method for numerical simulation of molecular systems. Availability of open source software packages and refined forcefields make this method a valuable tool in the study of biomolecules. The computational cost however restricts both time and length scales that are accessible by MD simulation. Coarse graining (CG) is a method in which clusters of atoms are represented by single particles called beads and simulated using a molecular simulation scheme such as MD. When properly implemented this method can significantly decrease computational cost while preserving most of the resolution and accuracy of an atomistic simulation.

The particular computational bottleneck in the case of DNA simulations is the number of solvent molecules required to fill the simulation box. As nucleic acids do not necessarily assume stable compact forms as proteins do, a simulation box big enough to accommodate the polymer in its fully extended conformation is required. As a result, number of nucleic acid particles scales linearly with chain length, while the box volume, therefore the number of solvent molecules, scales with third order. Such a system is quickly dominated by solvent particles as chain size grows and most of the computational power is lost to calculating solvent-solvent interactions. In this work, we attempted to overcome this bottleneck by building a coarse grained model with implicit solvent representation.

An interesting property of DNA is its stiffness which increases by an order of magnitude upon hybridization. Many biological processes involve partially molten or nicked DNA, presumably taking advantage of this property. Models used to study such processes therefore need to be able to represent both single stranded and double stranded DNA faithfully. Atomistic models satisfy this requirement, but they are limited due to computational costs. Continuum models have reduced computational cost, but they take stiffness as a parameter, which intrinsically disables them from representing the stiffness change upon hybridization. We suggest that a properly parameterized coarse

grained model might be able to represent DNA in both single and double stranded form, while omitting solvent particles, thus bridging the gap between atomistic and continuum models.

In this work we built a coarse grained model of DNA using a structural modeling approach. We used atomistic MD simulations of smaller DNA molecules to extract structural information which we used in building the CG model. We then characterized the CG model and compared the results to the atomistic reference through several iterations. Our main goal was to reproduce the stiffness difference between ssDNA and dsDNA via a combination of steric effects and topological constraints resulting from inter-strand interactions. We also aimed that the CG model feasibly simulate kilobase long DNA segments for several microseconds, which are the lower limits of biologically and technologically relevant time and length scales, not accessible through atomistic simulations. Despite the limited success of the CG model in function, the process of building the model on a structural basis provided valuable insight into driving forces behind the double helical form and exceptional stiffness of dsDNA.

1.2 Methods

1.2.1 Molecular Dynamics

Molecular dynamics (MD) is a method where dynamics of a number of particles are simulated by numerically integrating an equation of motion (Eq. 1.1) for discrete timesteps to yield a trajectory of positions $r_i(t)$ and velocities $\dot{r}_i(t)$. Initial conditions of the simulation describe the positions and velocities for each particle at $t = 0$. Net force $F_i(t)$ acting on each particle is then calculated by summing all the forces acting on the particle, which are derived from interaction potentials describing the system. Forces are then used to accelerate each particle with respect to their mass m_i, updating the position and velocities for the next timestep.

$$\ddot{r}_i(t) = \frac{F_i(t)}{m_i} \quad \text{where} \quad F_i(t) = \sum_j -\frac{\mathrm{d}V_j}{\mathrm{d}r}(r_i) \tag{1.1}$$

Potential functions for bonds, angles, dihedrals and non-bonded interactions can be determined empirically from experiments or analytically from quantum mechanical calculations. Collections of such potentials describing commonly studied molecules exist in the form of force fields. Initial velocities are usually randomly assigned with a Maxwell distribution, and initial positions are chosen appropriately for the system of interest.

Various methods to generate isothermal and isobaric conditions exist in the form of modifications to the equation of motion.

We simulated atomistic models of single and double stranded DNA molecules dissolved in water for reference purposes. We used the GROMACS MD simulation package [15] with AMBER99 forcefield [16] with a velocity rescaling thermostat [17] and Berendsen barostat [18] for atomistic MD simulations. We prepared 7 base-long single stranded and 30 base-long double stranded DNA molecules dissolved in simple point charge (SPC) water molecules and neutralized by positive ions. After minimization using the steepest descent algorithm, we simulated the single stranded system for 60 nanoseconds and the double stranded system for 30 nanoseconds at 300K and 1 bar conditions, yielding NPT trajectories. Resulting total energy trajectories (Fig. 1.1) suggested that systems were properly equilibrated.

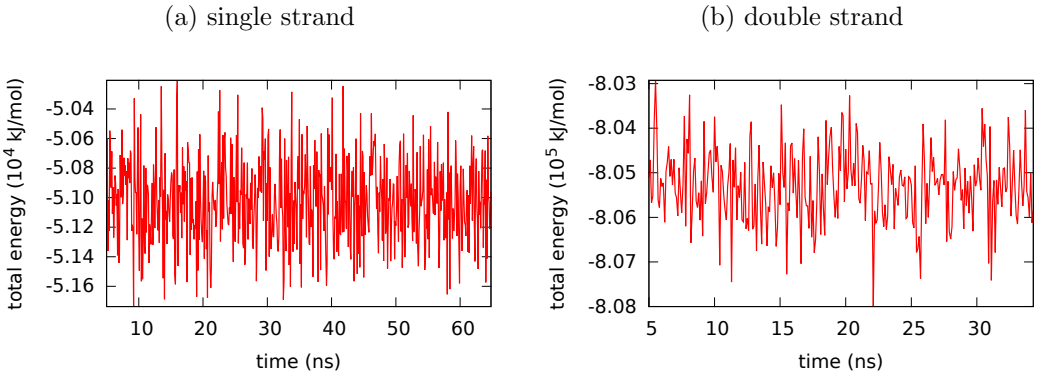


Figure 1.1: Total energies from atomistic simulations fluctuated without drifting, suggesting equilibrated systems. Total energies were not directly comparable due to the difference in system size. Normalization with respect to system size was not trivial due to solvent interactions.

1.2.2 Coarse graining

Coarse graining (CG) is a method where several atoms are represented by a single particle, also called a bead, attempting to reduce the computational cost of MD simulations. A DNA monomer, namely a nucleotide, consists of three components: a sugar, a phosphate, and a nucleic base. Such chemically distinct domains within a monomer serve as a natural basis for our CG mapping scheme. We defined three types of beads that are called P, S and B beads representing the center of mass of the sugar, phosphate and

nucleic base atoms respectively.

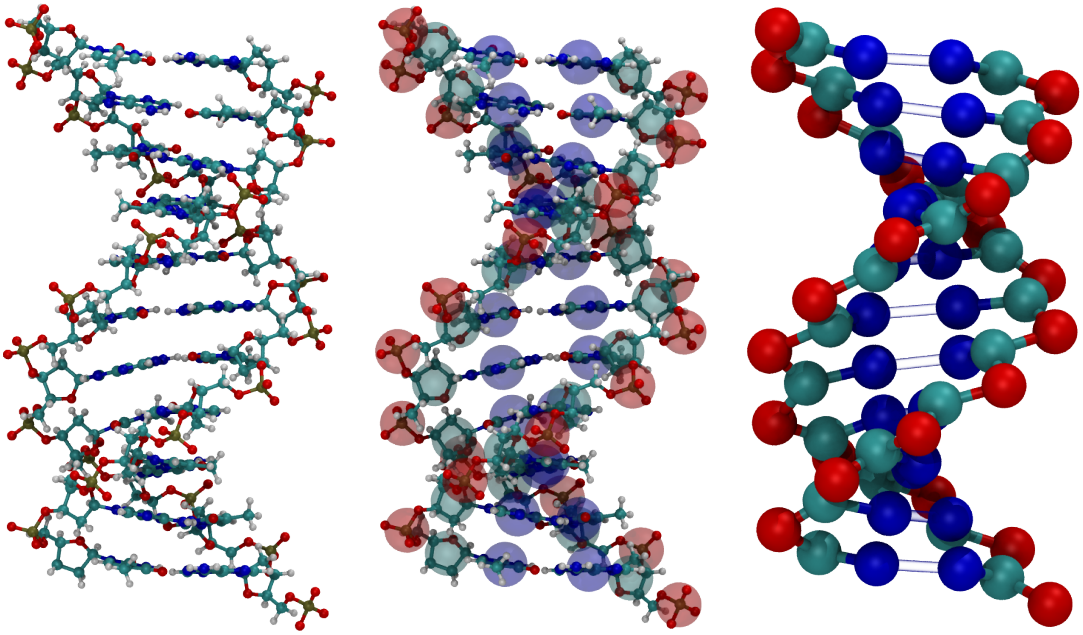


Figure 1.2: Coarse grained representation of double stranded DNA. **P**hosphate (red), **S**ugar (cyan) and **B**ase (blue) beads are mapped onto the center of mass of the respective chemical domains. Inter-strand coupling and the double helical structure is achieved through a combination of inter-strand bonds and steric effects.

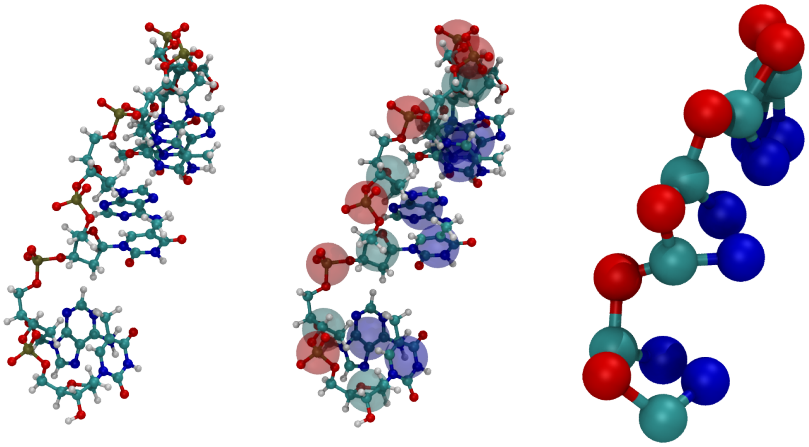


Figure 1.3: Coarse grained representation of single stranded DNA. A much wider conformational space is available to ssDNA due to lack of inter-strand interactions. Base-base stacking is effectively weaker without complementary hydrogen bonds, enabling the backbone to bend more freely.

Considering chain directionality and purine-pyrimidine asymmetry, this scheme provided us with a basic topology for our model. In addition to elementary bonds, angles and dihedrals that we derived from connectivity of the original system, we also defined an improper dihedral to describe the direction of bases with respect to the backbone.

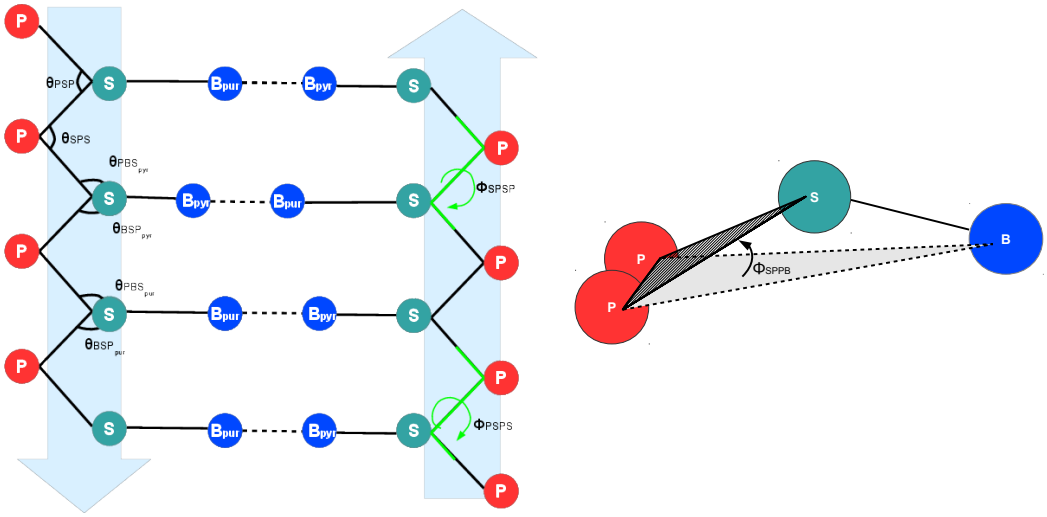


Figure 1.4: Diagram for topology of the CG model on left, diagram for the improper dihedral on right

Interaction potentials can be derived and tuned by various methods. We used the Boltzmann inversion method for a structural modeling approach. In this method, interaction potentials are derived using the Boltzmann relation (Eq. 1.3) to reproduce distributions observed in the original system. We extracted bond, angle and dihedral histograms from atomistic trajectories using VOTCA coarse graining package [19]. We normalized the histograms for density of states using appropriate functions (Eq. 1.2) and renormalized them to unity to acquire true probability density functions.

$$P_{bond}(r) = \frac{H(r)}{4\pi r^2}, \quad P_{angle}(\theta) = \frac{H(\theta)}{\sin(\theta)}, \quad P_{dihedral}(\phi) = H(\phi) \tag{1.2}$$

We then used the Boltzmann relation (Eq. 1.3) to produce effective potentials from histograms. Effective potentials extracted by Boltzmann inversion implicitly include the solvent effects. Upon extracting and smoothing effective potentials, we observed irregular features with magnitudes around k_BT and decided to approximate the potentials using analytical forms for better control in tuning.

$$V(x) = -k_b T \ln(P(x)) \tag{1.3}$$

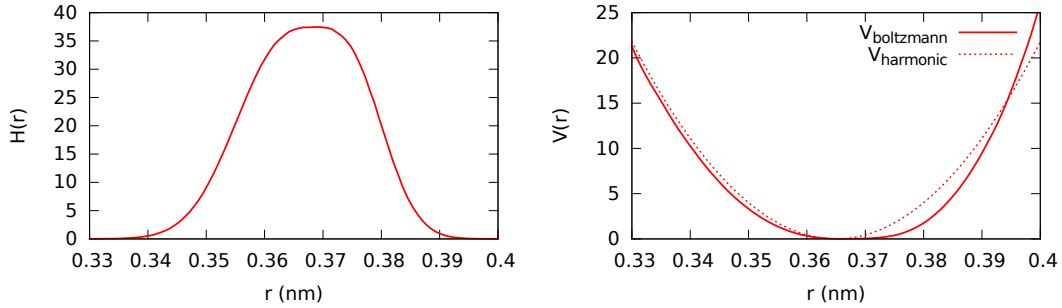


Figure 1.5: Boltzmann inversion and approximation with harmonic function (Eq. 1.4)

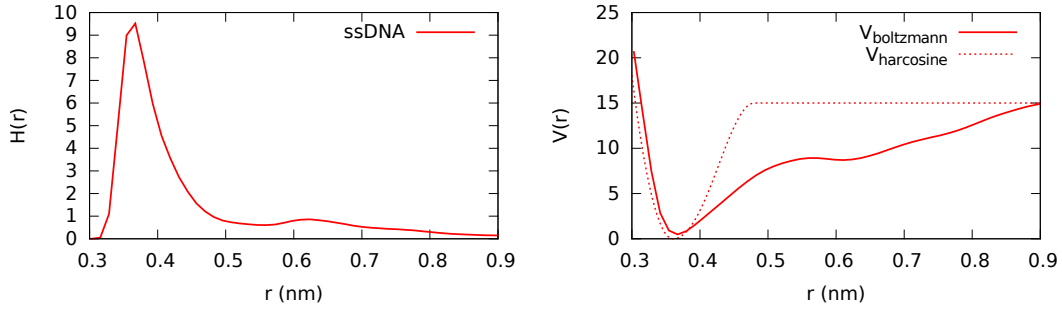


Figure 1.6: Boltzmann inversion and approximation with harmonic cosine function (Eq. 1.5)

1.2.2.1 Bonds and angles

We approximated all intra-strand bonds and angles except for base-base stacking with harmonic potentials (Eq. 1.4).

$$V_{harmonic}(x) = \frac{1}{2}\, k\, (x - x_0)^2 \tag{1.4}$$

We approximated breakable bonds such as base-base stacking and inter strand bonds using harmonic cosine functions (Eq. 1.5). Harmonic cosine functions are harmonic functions with a cosine tail that smoothly go to zero at a given cutoff value; they therefore require two extra parameters ϵ and r_{cutoff} in addition to harmonic parameters x_0 and k.

$$V_{harcos}(r) = \begin{cases} \frac{1}{2}\ k\ (r - r_0)^2 & r \leq r_0 \\ \frac{1}{2}\ \epsilon\ [\cos(\alpha x^2 + \beta) - 1] + \epsilon & r_0 < r \leq r_{cutoff} \\ \epsilon & r > r_{cutoff} \end{cases} \tag{1.5}$$

$$\text{where} \quad \alpha = \frac{\pi}{r_{cutoff}^2 - r_0^2}\ ,\ \beta = \pi - \alpha r_0^2$$

1.2.2.2 Dihedrals

We approximated proper dihedrals using periodic functions (Eq. 1.6) natively used by GROMACS. We used harmonic functions (Eq. 1.4) to approximate improper dihedrals.

$$V_{dihedral}(\phi) = k\ (1 + \cos(\phi - \phi_s) \tag{1.6}$$

1.2.2.3 Non-bonded interactions

We used a hard-core repulsive potential (Eq. 1.7) as the default non-bonded interaction for all particles. Hard core potential was derived from the repulsive part of a 9-6 Lennard-Jones potential. Such a repulsive interaction was necessary to represent the excluded volume for each bead and prevent bond crossing.

$$V_{hardcore}(r) = \begin{cases} \epsilon \left[2\left(\frac{r_{cutoff}}{r}\right)^9 - 3\left(\frac{r_{cutoff}}{r}\right)^6\right] + \epsilon & r \leq r_{cutoff} \\ 0 & r > r_{cutoff} \end{cases} \tag{1.7}$$

We expected P beads, being charged in atomistic systems, to interact with a stronger repulsion, which was confirmed by radial distribution functions. Considering the order of the pairwise coulomb potential, we derived a longer range soft core repulsive potential for P-P non-bonded interactions. We derived he soft core potential (Eq. 1.8) from the repulsive part of a 2-1 Mie function, which is the generalized form of n-m Lennard-Jones potentials.

$$V_{softcore}(r) = \begin{cases} 4\epsilon \left[\frac{1}{4}\left(\frac{r_{cutoff}}{r}\right)^2 - \frac{1}{2}\left(\frac{r_{cutoff}}{r}\right)\right] + \epsilon & r \leq r_{cutoff} \\ 0 & r > r_{cutoff} \end{cases} \tag{1.8}$$

1.2.2.4 Exclusions

We excluded atoms that are topologically within 3 bonds distance from interacting through non-bonded interactions. Base-base stacking and inter-strand bons were not considered exclusion generating in this sense, i.e. consecutive or complementary bases

were not considered 'bound' in the topology despite the bonded interactions between them. So we explicitly excluded consecutive and complementary bases from non-bonded interactions to avoid competition between the generic repulsion and spesific interactions. We also used the same exclusion scheme while calculating the radial distribution functions from atomistic systems.

1.2.2.5 Iterations

Finding the set of interactions to most faithfully describe a system is the fundamental problem of coarse graining. Topology provided a set of interactions to include in our model, this however was a necessary but not sufficient set to represent our original system. As we did not have the resources to completely span the parameter space, we iterated our model while varying interaction parameters intuitively to understand relative contributions of various interactions to behaviour of the system. We also experimented with a number of interactions not described by basic topology.

We simulated the CG systems at 300 K using a stochastic dynamic integrator with a modified (Langevin) equation of motion (Eq. 1.9). This integrator was reported to be suitable for reproducing the dynamics of implicit solvent models [20].

$$\ddot{r}_i(t) = \frac{F_i(t)}{m_i} - \gamma \dot{r}_i(t) + \frac{R_i(t)}{m_i} \tag{1.9}$$

where γ is a positive friction coefficient and $R_i(t)$ are stochastic forces

We simulated 50 base-long chains for 500 nanoseconds for each iteration. Simulations took only a few hours and resulted in well equilibrated systems big enough to calculate global properties. We then characterized our model with respect to local and global metrics and compared the results to the atomistic reference.

We used bond, angle and dihedral histograms as local metrics. We expected bond and angle histograms to reproduce atomistic histograms minus the irregularities we omitted by analytical approximations. This was a sanity check as such behaviour is expected due to the Boltzmann inversion method we used. We expected the backbone dihedrals to stiffen upon hybridization. This was the main challenge as such behaviour is not explicitly represented in the model.

We used pitch and persistence length as global metrics. We calculated these by fitting an analytical form (Eq. 1.10) to the correlation of bond vectors along the backbone of

the polymer. We expected an order of magnitude increase in persistence length upon hybridization, as observed in the atomistic reference.

$$C(\tau) \;=\; \sum^{n} \frac{\vec{r}_{(i+\tau)} \cdot \vec{r}_i}{\vec{r}_i \cdot \vec{r}_i} \;\approx\; \exp\left(-\tau/l_p\right)\left[a + (1-a)\cos\left(\frac{2\pi\tau}{\lambda}\right)\right] \tag{1.10}$$

where n is chain length, λ is pitch and l_P is persistence length

1.3 Results

1.3.1 Atomistic systems

1.3.1.1 Bonds and angles

We calculated histograms for intra-strand bonds and angles from both single and double stranded atomistic simulations and inter-strand bond histograms from double stranded atomistic simulation. Backbone bonds (Fig. 1.7) and angles (Fig. 1.8) demonstrated nearly harmonic distributions and behaved similarly in ssDNA and dsDNA. Bonds (Fig. 1.9) and angles (Fig. 1.8) involving base-backbone interaction demonstrated more complicated behaviour, as they were strongly coupled with inter-strand interactions. We approximated the effective potentials for these interactions using harmonic functions (Eq. 1.4), yielding the parameters summarized in Table 1.1

Base-base stacking (Fig. 1.11) behaved qualitatively different in ssDNA and dsDNA; bases were constrained in a stacked configuration in dsDNA whereas they had a significant probability of dissociation in ssDNA. For this reason, we approximated the base-base stacking interaction potential using a harmonic cosine function (Eq. 1.5), yielding the parameters summarized in Table 1.2.

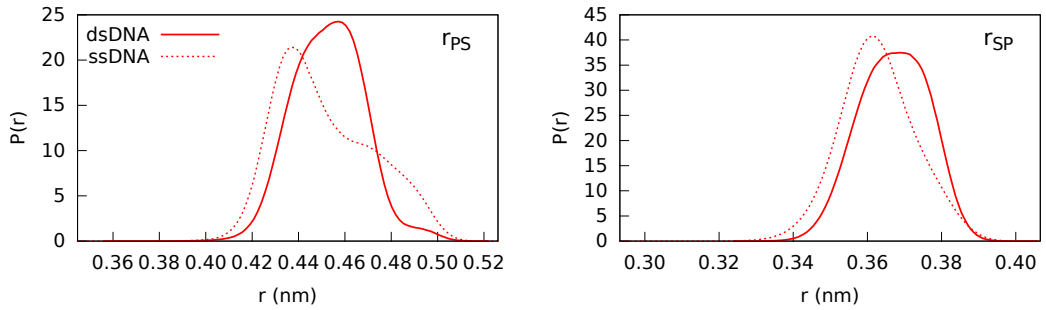


Figure 1.7: Backbone bond histograms calculated from atomistic simulations. PS and SP bonds make up the sugar-phosphate backbone. SP bond represents a single rotatable bond while PS bond represents two rotatable bonds joined by a methyl group in the original system. This results in an extra degree of freedom that is not explicitly represented in the CG model.

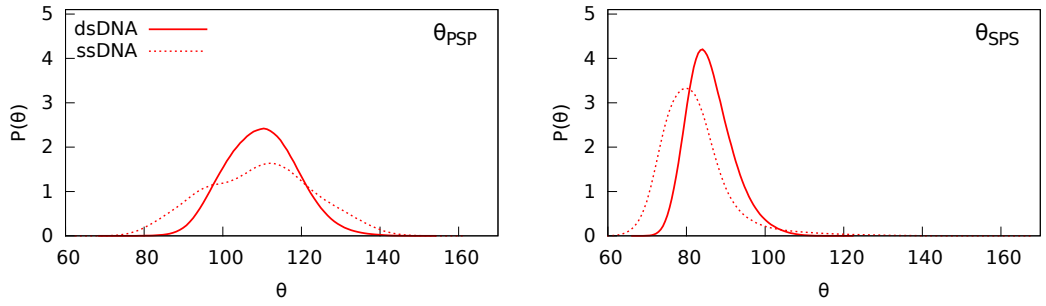


Figure 1.8: Backbone angle histograms calculated from atomistic simulations. Backbone directionality results in two distinct angle types.

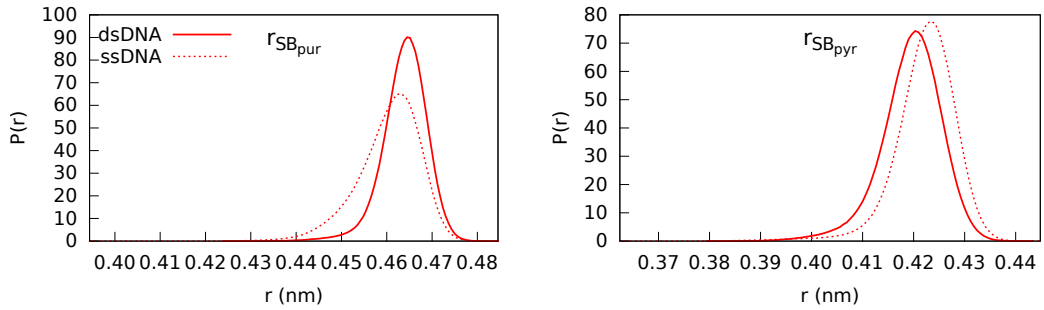


Figure 1.9: Histograms for base-backbone bonds calculated from atomistic simulations. Bases are connected to the backbone through sugars. Purine-pyrimidine assymmetry yields two distinct bond types.

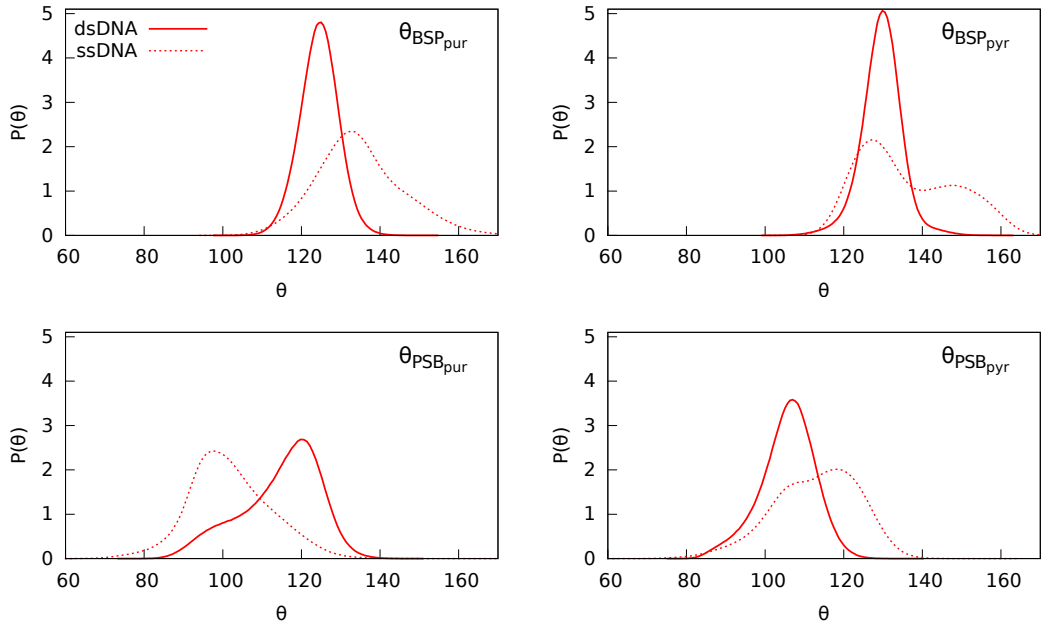


Figure 1.10: Histograms for base-backbone angles calculated from atomistic simulations. Backbone directionality coupled with purine-pyrimidine asymmetry yields four distinct angle types. Difference between ssDNA and dsDNA is most significant in this set.

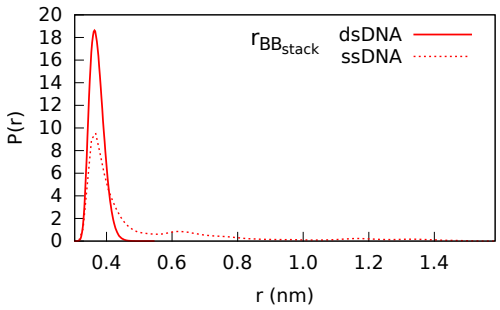


Figure 1.11: Histogram for base-base stacking bond calculated from atomistic simulations. Stacking interaction is localized for dsDNA while it has non-zero probability of disassociation for ssDNA. Its maximum value for ssDNA is determined by the bonds and angles connecting consecutive bases through the backbone.

bond	r_0 (nm)	k $\left(\frac{nm.kJ}{mol}\right)$	angle	θ_0 $(degrees)$	k $\left(\frac{degrees.kJ}{mol}\right)$
SP	0.36	35000	SPS	85	500
PS	0.45	11000	PSP	110	90
SB_{pur}	0.36	100000	PSB_{pur}	126	250
SB_{pyr}	0.41	120000	PSB_{pyr}	106	206
			BSP_{pur}	121	264
			BSP_{pyr}	130	408

Table 1.1: Parameters for harmonic bonds and angles extracted by least squares fitting of a harmonic function (Eq. 1.4) to data from dsDNA.

Inter-strand interactions between the complementary bases (Fig. 1.12) were very rigid, yielding the highest stiffness we calculated for any bond in the system. We also measured the distances to neighbouring complementary bases which were roughly as constrained as backbone bonds. We measured inter-strand S-S distance in addition to inter-strand B-B distances and found it to be constrained near the order of intra-strand bonds. We approximated the potentials for inter-strand interactions using harmonic cosine functions (Eq. 1.5) to represent breakable bonds, yielding the parameters summarized in Table 1.2. It should be noted that the ϵ parameter of the base-base stacking interaction can be extracted from histograms but the ϵ parameter of the inter-strand bonds had to be guessed and tuned by iteration, as we do not have DNA melting available

to us through atomistic simulations.

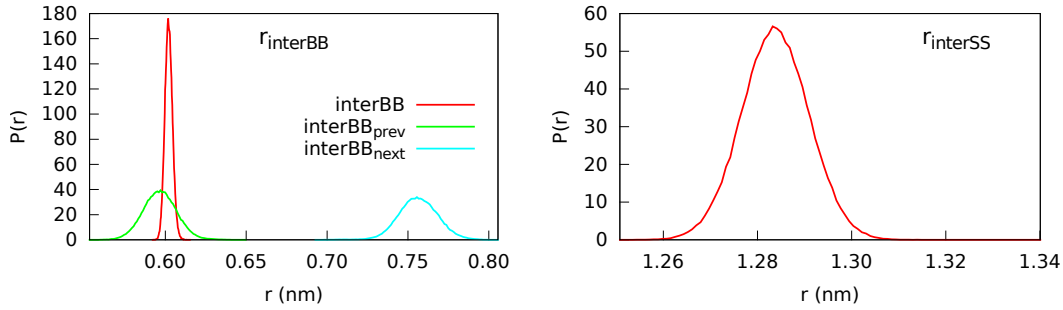


Figure 1.12: Inter-strand bond histograms from atomistic simulations of double stranded systems. We measured distances to the neighbouring (previous and next) bases in the complementary strand in addition to the exact complementary base. Complementary base distance had a very narrow distribution due to multiple hydrogen bonds whereas neighbouring complimentary base distances were distributed roughly as broadly as backbone bonds.

bond	r_0 (nm)	r_{cutoff} (nm)	k $\left(\frac{nm.kJ}{mol}\right)$	ϵ (kJ/mol)
$BB_{stacking}$	0.39	0.55	5000	10
BB_{inter}	0.602	0.618	500000	25
SS_{inter}	1.283	1.33	50000	25

Table 1.2: Parameters for harmonic cosine bonds extracted by least squares fitting of a harmonic cosine function (Eq. 1.5). Data from ssDNA was used for base-base stacking and data from dsDNA was used for inter-strand bonds. The ϵ parameter for inter-strand bonds could not be reliably extracted from dsDNA data, so they are considered tunable.

1.3.1.2 Dihedrals

We calculated histograms for backbone dihedrals from both double and single stranded DNA simulations (Fig. 1.13). We approximated the potentials for backbone dihedrals using periodic functions (Eq. 1.6), removing the finer features observed in single stranded systems, yielding the parameters summarized in Table 1.3. Unlike bonds and angles, dihedral histograms showed significant difference in stiffness between single and

double stranded systems. Despite having similar equilibrium values, dihedrals of ds-DNA were roughly 5 times as stiff as the dihedrals of ssDNA. This is a fundamental feature of DNA, presumably contributing significantly to global properties. To assess the contribution of sugar-phosphate backbone to dihedral stiffness, we simulated a bare sugar-phosphate polymer and calculated histograms for backbone dihedrals from this simulation. The bare sugar-phosphate polymer did not significantly prefer the equilibrium values of dsDNA and ssDNA backbone dihedrals, suggesting a high contribution from the interactions involving bases.

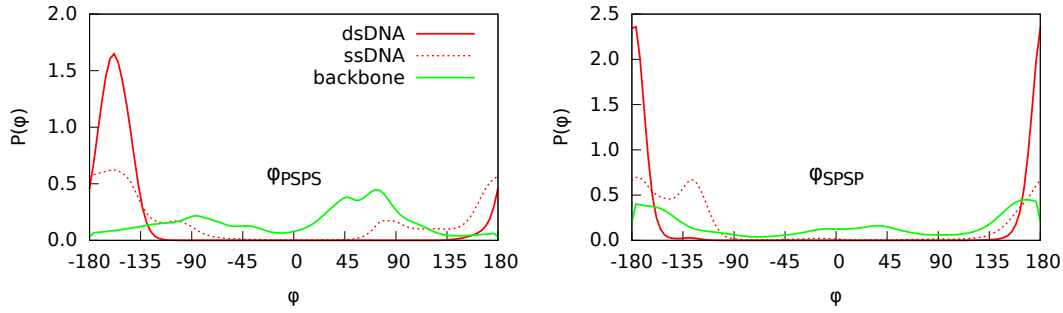


Figure 1.13: Histograms for backbone dihedrals calculated from atomistic simulations of dsDNA, ssDNA and bare sugar-phosphate polymer systems. There was significant difference in both stiffness and distribution between ssDNA and dsDNA. The bare sugar-phosphate backbone did not significantly prefer the equilibrium values of dsDNA and ssDNA backbone dihedrals, suggesting a high contribution from the interactions involving bases.

dihedral	ϕ_s $(degrees)$	k $\left(\frac{degrees.kJ}{mol}\right)$
PSPS_{single}	12.9	37.4
PSPS_{double}	2.5	7.0
SPSP_{single}	12.6	30.8
SPSP_{double}	17.4	7.5

Table 1.3: Parameters extracted for backbone dihedrals via least squares fitting of periodical functions (Eq. 1.6). Despite having similar equilibrium values, backbone dihedrals of dsDNA were roughly 5 times as stiff as the backbone dihedrals of ssDNA.

We also defined an improper dihedral (SPPB) to measure the base orientation with respect to the backbone. We observed this improper dihedral to be distributed around 0 for single strand and constrained to only positive values for double strand systems (Fig. 1.14), effectively describing the right-handedness of the helix.

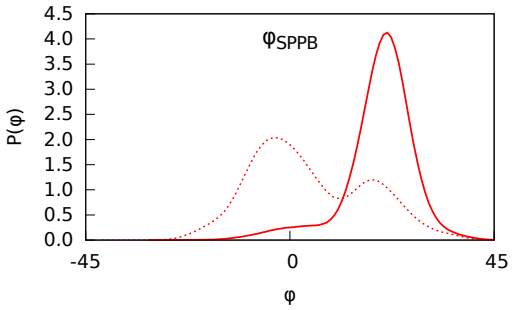
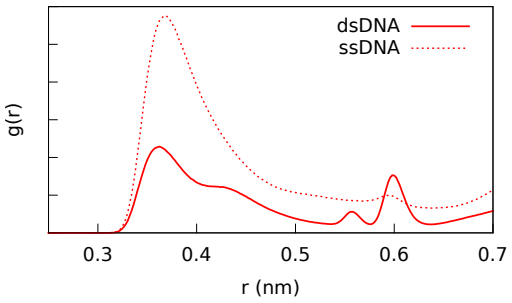


Figure 1.14: Improper dihedral histogram calculated from atomistic stranded simulations

1.3.1.3 Non-bonded interactions

We calculated radial distribution functions separately for all atoms and for P atoms only. We were mainly interested in the left tail of the distribution to determine cutoff parameters for non-bonded interactions.

(a) RDF for all beads



(b) RDF for P beads

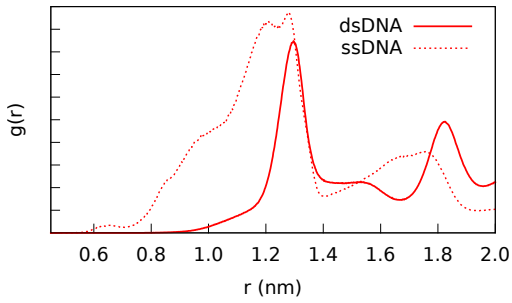


Figure 1.15: Radial distribution functions (RDF) calculated from atomistic simulations, arbitrarily normalized for comparison between systems. Exclusion rules used for CG systems were taken into account while calculating RDFs.

We chose a cutoff radius of 0.35 nm for the generic hard core repulsion and set the

ϵ value to a trivial 1 kJ/mol considering the high order of the hard core repulsion. We investigated cutoff values in the range of 1.0 - 1.5 nm and ϵ values in the range of 1 - 100 kJ/mol for the soft core P-P repulsion.

1.3.1.4 Global properties

We calculated bond correlations from both single and double stranded atomistic simulations. We then extracted pitch (λ) and persistence length (l_p) via least squares fitting of an analytical form (Eq. 1.10) to the correlation data.

Double stranded system yielded a l_p of 125 bases (40 nm) and a pitch of 11.4 bases. Pitch of our atomistic system was significantly higher than the experimentally observed value of 10.5, but the persistence length was in agreement with recent experimental observations of 35 nm [21]. The incorrect pitch is presumably an artifact introduced by the AMBER99 forcefield.

Single stranded system yielded a persistence length of 3 bases (1 nm) and a pitch of 10.1 bases. We took these results with a grain of salt as our single stranded system, being less than one pitch long, was not big enough for such a calculation. A persistence length of 1 nm is however not far from recent experimental observations of 2.2 nm [22]; it roughly agrees with the order of magnitude difference observed between single and double stranded DNA.

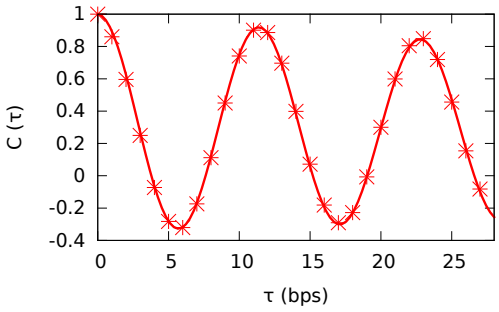


Figure 1.16: Pitch ($\lambda = 11.4$ *bps*) and persistence length ($l_p = 125$ *bps*) for dsDNA extracted from bond correlations along the backbone via least squares fitting of an analytical form (Eq. 1.10).

1.3.2 Coarse grained systems

The structural modeling approach we have taken implied an enormous parameter space to be explored. As systematic exploration of such a space was infeasible, we followed an intuitive approach based on our current understanding of the molecular structure of DNA. Iterations were therefore based on trial and error rather than a predefined algorithm. We quantified the parameter optimization process by using analytical potentials instead of arbitrary forms that resulted from the Boltzmann inversion method. This allowed us to systematically span small volumes of the parameter space when we found it necessary.

There might exist significant interactions that we have overlooked through iterations. There were also a number of interactions that we have omitted after observing that they were not significant contributors to overall behavior of the system. The final set of interactions that we decided to include in our CG model are summarized in Table 1.4.

				name	type
dsDNA	ssDNA	bonded	backbone	bond PS	harmonic (Eq. 1.4)
				bond SP	harmonic (Eq. 1.4)
				angle PSP	harmonic (Eq. 1.4)
				angle SPS	harmonic (Eq. 1.4)
				dihedral PSPS	periodic (Eq. 1.6)
				dihedral SPSP	periodic (Eq. 1.6)
			bases	bond SB[1]	harmonic (Eq. 1.4)
				angle BSP[1]	harmonic (Eq. 1.4)
				angle PSB[1]	harmonic (Eq. 1.4)
				stack BB[2]	harmonic cosine (Eq. 1.5)
				improper SPPB	harmonic (Eq. 1.4)
		non-bonded		excluded volume	hardcore (Eq. 1.7)
				PP repulsion	softcore (Eq. 1.8)
	inter-strand			inter BB[2]	harmonic cosine (Eq. 1.5)
				inter SS[2]	harmonic cosine (Eq. 1.5)

Table 1.4: Summary of interactions included in the CG model ([1]: purine-pyrimidine asymmetric, [2]: non exclusion generating).

1.3.2.1 Bonds and angles

We used potentials extracted from the atomistic dsDNA system in parameterization of the CG model to ensure compatibility with the double helical structure. Intra-strand bond histograms of the atomistic reference were readily reproduced by CG simulations as predicted by the Boltzmann relation. We observed that intra-strand bond stiffness does not contribute significantly to the global behaviour of the system as increasing all intra-strand bond stiffness parameters by an order of magnitude only results in a %1 increase in persistence length. Angle histograms of the atomistic reference were also readily reproduced by CG simulations. We observed that an order of magnitude increase in stiffness parameters of all angle potentials yield a %100 increase in persistence length of dsDNA. We observed that neither bond nor angle constraining occur upon hybridization in the CG model; ssDNA and dsDNA simulations yield nearly identical histograms.

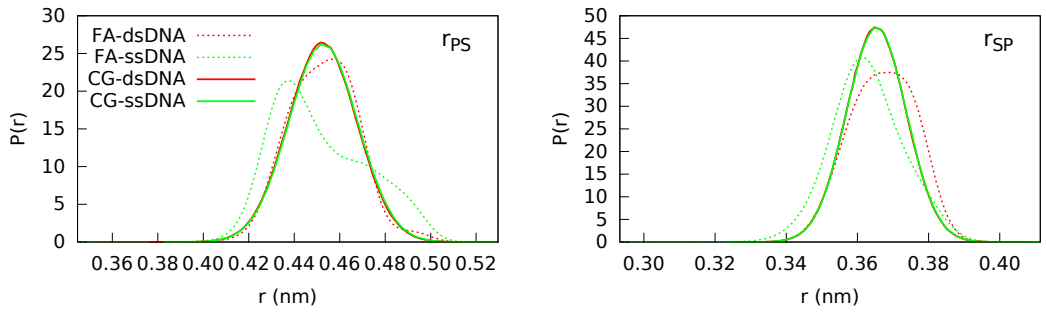


Figure 1.17: Backbone bond histograms calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

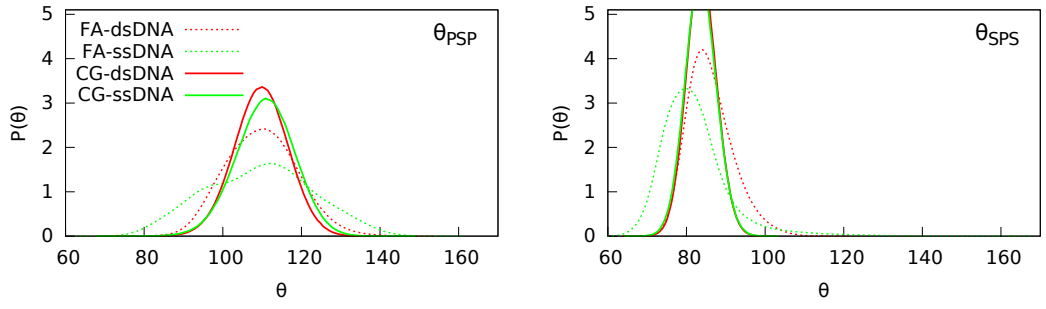


Figure 1.18: Backbone angle histograms calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

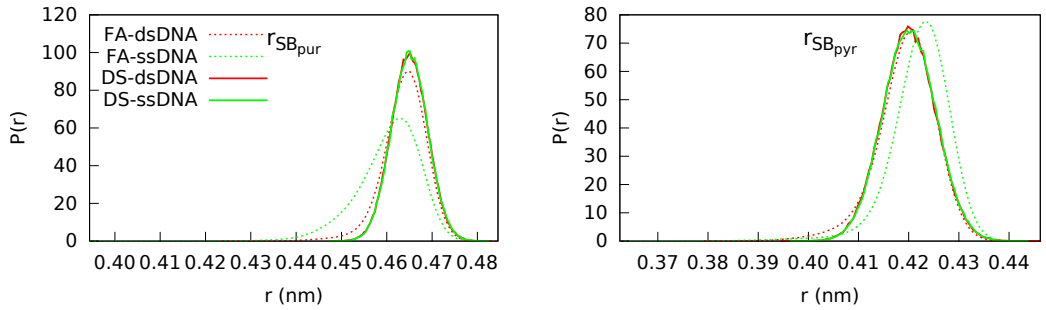


Figure 1.19: Histograms for base-backbone bonds calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

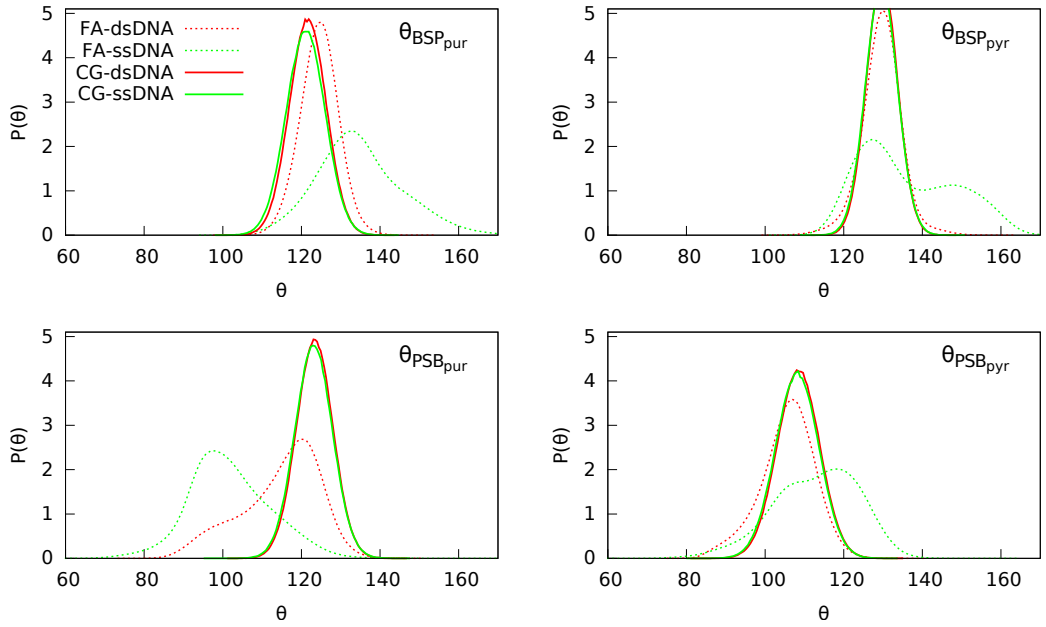


Figure 1.20: Histograms for base-backbone angles calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

For the initial process of parameterizing the backbone, we used effectively unbreakable inter-strand bonds. For this reason, we introduced inter BB potentials with ϵ values of 25 kJ/mol ($10k_BT$) to ensure rigidity of inter-strand interactions. Our plan of action was to later reduce the ϵ values iteratively to tune for desired melting behaviour. We observed that even though complementary bases were constrained as intended, the distances to neighboring complimentary bases yielded very broad distributions (Fig. 1.21). Introducing harmonic cosine bonds with neighbouring complimentary bases readily re-

produced atomistic distributions, but this topology proved infeasible for our requirements. Preliminary experiments with ϵ values as low as 10 kJ/mol at temperatures as high as 400 K showed that such systems, having three inter-strand bonds per base, demonstrate unphysical melting behaviour as backbones completely deform before a significant proportion of bases dissociate from their compliments.

We also observed that inter-strand sugar-sugar (interSS) distances were distributed very broadly in the CG systems compared to the atomistic reference. Through many iterations, we observed that despite reproducing atomistic bond, angle and dihedral distributions with reasonable agreement, conformations our model preferred were predominantly not helical. Introducing a harmonic cosine interSS bond significantly improved faithfulness of our model by constraining it around helical conformations.

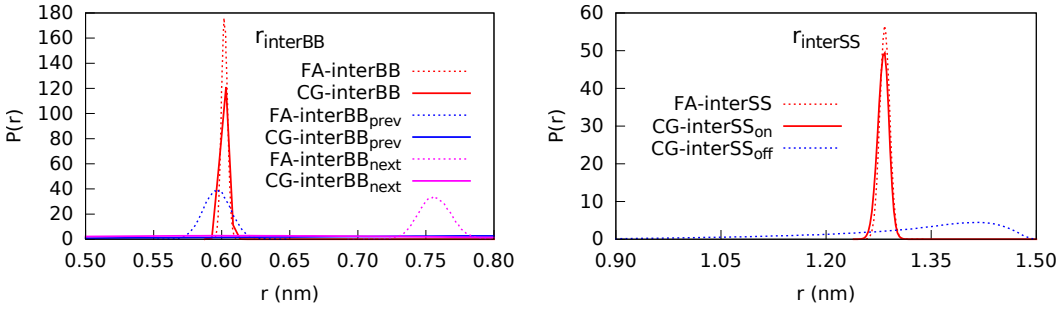


Figure 1.21: Histograms for inter-strand bonds calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

1.3.2.2 Dihedrals

Behaviour of the dihedrals in the atomistic systems were significantly different for ssDNA and dsDNA. This proved difficult to reproduce with our coarse grained model.

Backbone dihedrals were the longest range bonded interactions in our model and expected to contribute significantly to the overall behavior of the system. We used the dihedral parameters extracted from ssDNA atomistic reference and expected backbone dihedrals to stiffen upon hybridization as the atomistic reference does. Using dsDNA atomistic reference dihedrals instead yielded a %100 increase in persistence length, confirming our intuition about importance of backbone dihedrals. The best parameter sets we have been able to find however poorly reproduced stiffening behaviour of the atomistic reference.

Dihedrals from CG ssDNA simulations were roughly twice as stiff as the atomistic

reference. This was expected to some extent as we excluded some of the non-zero probability angles while approximating the dihedral potentials. The ssDNA systems were further constrained by an improper dihedral we introduced into the model after some iterations. Dihedrals from CG dsDNA simulations, on the other hand, had roughly two thirds of the stiffness of the atomistic reference. This suggested that inter-strand interactions were not constraining the system sufficiently upon hybridization.

Parameters could be tuned to yield stiffer or looser dihedrals to represent either ssDNA or dsDNA more faithfully, but the amount of constraint introduced by hybridization did not improve. The stiffness ratio between the ssDNA and dsDNA dihedrals in CG simulations were roughly 1.5 with our best parameters whereas this ratio is around 5 in the atomistic reference.

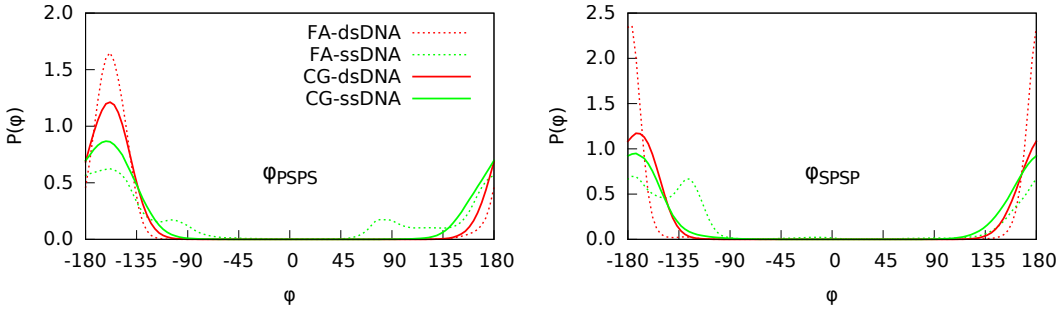


Figure 1.22: Histograms for backbone dihedrals calculated from CG simulations. Corresponding histograms from the atomistic reference are included for comparison.

We measured an improper dihedral (SPPB) that describes the direction of the bases with respect to the backbone (SPP) plane, effectively representing the handedness of the helix. Initially, we did not include this interaction in our coarse grained model. The bimodal distribution observed in atomistic ssDNA was reproduced in the CG model without explicit representation. The restriction of this dihedral to the positive peak, yielding a strictly right handed helix, was observed in the atomistic dsDNA simulation. This behaviour, however was not reproduced in the CG model.

Constraining the improper dihedral to the positive peak yielded a strictly right handed dsDNA in the coarse grained model, but it also constrained the CG-ssDNA, effectively increasing its persistence length. The increase was varied for different parameter sets, but it was around %10 for the parameter set we eventually chose to work with.

(a) improper disabled (b) improper enabled

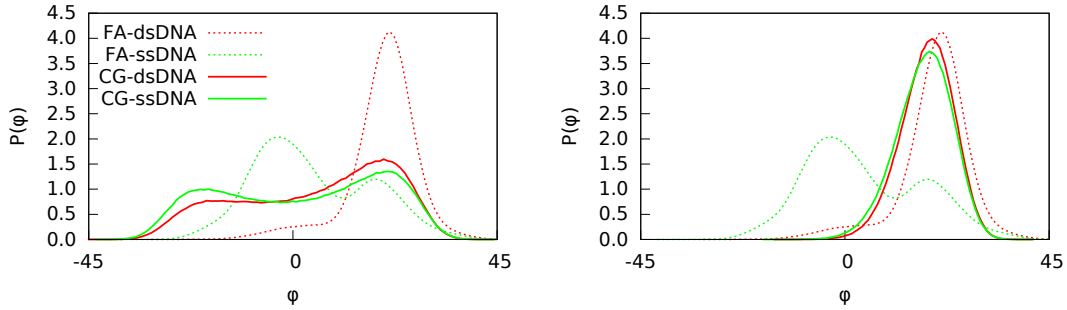


Figure 1.23: Improper SPPB dihedral histograms calculated from CG simulations without and with improper dihedral interaction enabled. Corresponding histograms from the atomistic reference are included for comparison.

1.3.2.3 Non-bonded interactions

The generic hard core repulsion with a 3.5 nanometer cutoff readily produced radial distribution functions compatible with the atomistic reference.

The soft core P-P repulsion had more complicated behavior and we experimented with cutoff values ranging from 1 to 1.5 nm and ϵ values ranging from 10 to 100 kJ/mol. We found P-P repulsion to strongly affect both pitch and persistence length of the systems. This was in agreement with the idea that stiffness of DNA being related to salt concentration through shielding of electrostatic interactions [21, 22]. We did not expect that the actual relation between shielding effects and mechanical properties of DNA could be studied with such a simple representation, so we only focused on finding appropriate parameters to suit our needs. We have found that a cutoff of 1.3 nm and ϵ of 50 kJ/mol yielding a 11.1 base pitch were the optimal parameters for our interaction set. It should be noted that a 1.4 nm cutoff with ϵ of 10 kJ/mol or a 1.5 nm cutoff value with ϵ of 100 kJ/mol yields very similar behavior.

$k \backslash r_c$	1.0	1.1	1.2	1.3	1.4	1.5
10	11.7	11.7	11.7	11.6	**11.4**	11.1
50	11.7	11.7	11.5	**11.1**	10.8	10.6
100	11.7	11.6	**11.4**	10.9	10.6	10.4

Table 1.5: Effect of PP repulsion on pitch (λ) of dsDNA

$k\backslash r_c$	1.0	1.1	1.2	1.3	1.4	1.5
10	8.3	8.2	9.0	9.2	**9.9**	11.0
50	8.6	9.3	9.6	**11.1**	12.1	14.9
100	9.0	9.4	**9.8**	12.3	16.0	18.5

(a) l_p for ssDNA

$k\backslash r_c$	1.0	1.1	1.2	1.3	1.4	1.5
10	42.8	43.9	47.7	49.0	**49.7**	54.3
50	47.0	51.6	49.5	**54.1**	66.7	80.6
100	49.5	49.1	**52.5**	60.5	78.7	101.0

(b) l_p for dsDNA

Table 1.6: Effect of PP repulsion on persistence length (l_p)

1.3.2.4 Global properties

Tuning the pitch of CG dsDNA systems was straightforward using P-P repulsion parameters. Ideally our CG model would yield a pitch near 11.4 bases in agreement with the atomistic reference, rather than 10.5 bases, which is the experimentally observed true pitch of dsDNA. We however biased our pitch preference towards 11.2 bases as this was the pitch yielded by a previous CG model developed by our group [9] which will be discussed later on. We wanted our new model to coincide with the old model in a fundamental structural feature such as pitch.

We found the persistence length (l_p) to be dependant on several parameters, most significants being P-P repulsion and backbone dihedrals. As we used P-P repulsion to tune the pitch, only free parameters remaining were backbone dihedrals in relation with l_p. As we could tune l_p of CG systems using backbone stiffness, the ratio between the l_p of ssDNA and dsDNA systems had an upper limit around 5 for any interaction set; this ratio is closer to 10 in the atomistic reference. In conclusion, we could only reproduce roughly half of global stiffening of DNA associated with hybridization with our CG model.

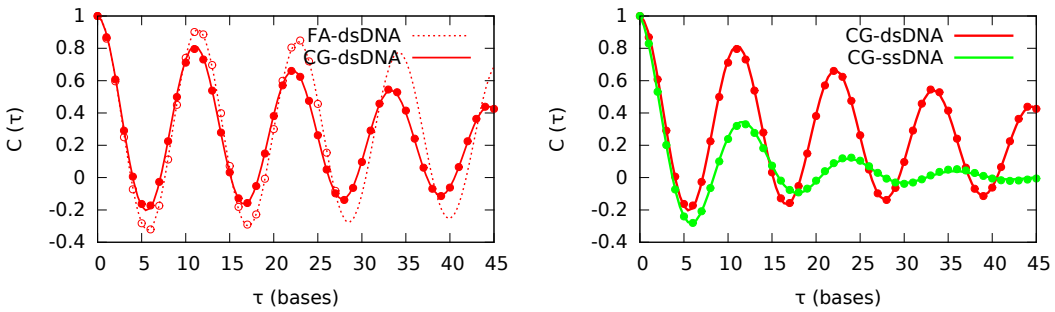


Figure 1.24: Representative bond correlations from atomistic and CG simulations.

1.4 Discussion

1.4.1 Atomistic systems

As straightforward as they are to set up, atomistic MD simulations are not without problems. The AMBER99 forcefield that we used, despite being well established, was far from perfect. The imperfections were especially well pronounced for nucleic acids.

The double stranded atomistic simulations yielded an incorrect pitch. Tuning AMBER99 forcefield parameters was beyond the scope of this work so we accepted the incorrect pitch as reference. When it came to single stranded systems, we avoided simulating longer systems as we know from experience that this forcefield yields unphysical behaviour for such systems. It was also reported that AMBER99 forcefield introduces destabilizing artifacts into dsDNA systems in simulations longer than 10 nanoseconds [23], although we did not directly suffer from this issue.

A refined version of the AMBER99 forcefield for nucleic acids was recently developed to address a number of issues with reported success [23]. We ran a dsDNA simulation using this forcefield with initial conditions identical to our atomistic reference for benchmarking purposes. Resulting data yielded the exact pitch and local histograms as the original AMBER99 forcefield. Convinced that the new forcefield, namely parmbsc0, was not refined in a way that we could benefit from, we did not simulate single stranded systems with this forcefield. No better performing atomistic forcefield for nucleic acids exists to our knowledge, however we recently discovered that AMBER99 forcefield might have performed better for higher salt concentrations.

Atomistic MD simulations typically have very high computational cost, making reruns infeasible in most cases. They also have the potential to yield huge amounts of data well beyond our data storage capabilities. Recording everything calculated in a MD simulation is therefore infeasible. We suffered from a combination of these properties, although not in a major way.

By default, GROMACS records only total energies. An operational mistake on our behalf was not to tune this behaviour to separately record DNA-DNA and DNA-solvent interaction energies. Had we done that, the resulting data could be useful in comparing the magnitude of solvent-DNA interactions in single and double stranded systems.

1.4.2 Coarse grained systems

We have confirmed that a 3-bead CG model of DNA solely based on structural modeling principles can represent either single or double stranded DNA molecules with reasonable accuracy. We also concluded that the stiffening associated with hybridization might not be possible to reproduce using such a simple model. We expected driving forces resulting from hydrogen bonding, electrostatic repulsion and steric hindrance in addition to molecular topology were to contribute to stiffening upon hybridization. We therefore explicitly represented these forces in our model using by harmonic cosine inter BB bonds, soft core repulsive PP repulsion and a hard core generic repulsion in addition to bonded interactions. The results have been unsatisfactory in a functional sense; we have not been able to build a unified model for single and double stranded DNA. The insights gained from iterations however have been very valuable. We had the opportunity to perturb the model in various ways to assess the relative contribution of individual interactions to the global behaviour.

We have concluded that while equilibrium values for intra-strand bonds and angles were important for compliance with the double helical form, our model was not sensitive to changes in stiffness of these interactions. We have also observed that backbone dihedrals were mainly responsible for the form and stiffness of double stranded DNA and they were not readily reproduced by simpler constraints included in our model. On a side note, we ran atomistic simulations on a bare sugar-phosphate backbone and observed that it did not have a significant helical tendency as even ssDNA does. The behaviour of backbone dihedrals must therefore result from a combination of several driving forces, presumably involving major contribution from bases, some of which we were unable to represent fully in our model. Some of the shortcomings of our model that we became aware of are discussed in further detail.

Geometry of bases In our model, every particle had an associated spherical excluded volume described by the hard core repulsive non-bonded interaction. **S**ugar and **P**hosphate groups were faithfully represented by this scheme. **B**ases, on the other hand, being planar groups, were ill represented by spherical volumes. We would expect such an approximation to result in under-constraining of double stranded DNA through both BB stacking and inter BB interactions. Deficiency of stiffening observed in our CG ds-DNA systems was consistent with expectation. There exists a coarse grained model of DNA with elliptical bases [11] but our results are not directly comparable to assess the

contribution of this effect.

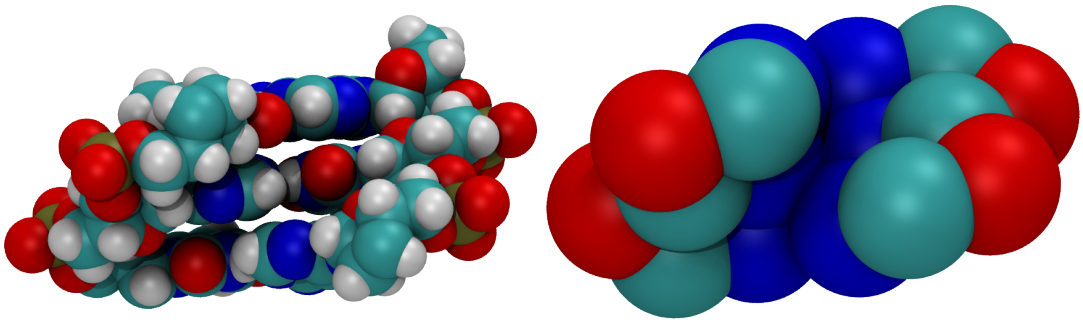


Figure 1.25: Atomistic vs. CG representation of bases, side view. Base-base stacking resolution is significantly reduced in the CG representation.

Hydrogen bonds Hydrogen bonds are highly angle specific and this effect is even more pronounced in the case of DNA where multiple hydrogen bonds exist between complementary bases. We would expect multiple hydrogen bonding to result in a complex energy barrier which yields very stiff but easily breakable inter-strand bonds. We however could not represent such a barrier with degrees of freedom available in our model, so we approximated hydrogen bonds using simple harmonic-cosine potentials. Lack of angle specificity in our inter-strand bonds is consistent with reduced stiffness in our double CG dsDNA systems. Through iterations, we attempted to improve our model by introducing a second bead for purine bases and associated ($B_1^{pur}\widehat{B_2^{pur}}B^{pyr}$) angle-specific inter BB bonds without significant results.

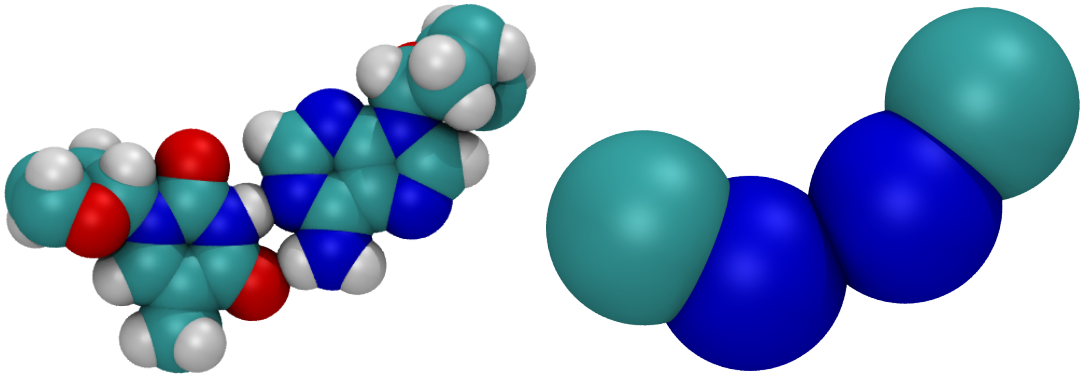


Figure 1.26: Atomistic vs. CG representation of bases, top view. Hydrogen bond resolution is significantly reduced in the CG representation.

Implicit solvent We have omitted solvent particles in our model based on the assumption that effective potentials calculated from full atomistic simulations implicitly included solvent effects. This approach appeared to be valid for both single and double stranded DNA as we can represent either by tweaking our dihedral parameters. It might however not have been as valid for a unified model such as the one we attempted to build.

We expected steric effects coupled with topological constraints introduced by inter-strand bonds to yield stiffening upon hybridization. They however turned out to be responsible for only a fraction of the stiffening observed in the atomistic reference, suggesting that we were missing a major contributor in our model.

Nucleic bases are hydrophobic, so it is reasonable to suggest that DNA-solvent interactions play a significant role in stiffening of DNA upon hybridization and the double helical structure. Such an effect would be reproducible by atomistic MD simulation, but not by a coarse grained model without explicit solvent particles; this is consistent with our observations.

Most of the performance gain of our CG model over atomistic simulations come from implicit solvent representation so it was infeasible for us to try to include solvent particles in our model. There exists a recent coarse grained DNA model with explicit solvent representation [14] with two orders of magnitude better performance than atomistic simulation. This is however still well below the performance required to simulate biologically and technologically relevant time and length scales.

Conclusion Not having been able to yield a functional parameter set after so many iterations, we eventually concluded that we had an understanding of the behaviour and limitations of this particular model. We cherished the insights gained from the process and stopped working on this model in favor of another exciting project. We concluded that a purely structural modeling approach might not be the best option to building a coarse grained model of DNA. We feel that it would be reasonable to reduce the resolution and introduce some phenomenological interactions such as dihedral potentials coupled with inter-strand bonds to yield a high performance unified coarse grained model of DNA. Such a model might be able to feasibly simulate biologically and technologically relevant time and length scales without explicit solvent representation.

Chapter 2

SUPERCOILING DYNAMICS

2.1 Introduction

DNA is a twist-storing polymer in its double stranded form, meaning that its monomers are joined by effectively non rotatable bonds. Sugar-phosphate bonds on each strand coupled by hydrogen bonds between bases provide torsional stiffness to the polymer, inhibiting the release of torsional stress by rotation. Supercoiling is one of the mechanisms through which DNA can express the excess torsional stress that can neither be released nor locally absorbed.

Supercoiling of DNA is of biological interest because it is frequently observed in genomic DNA. Most organisms keep their genome in a negatively supercoiled state, which in general serves to promote transcription. Distribution of supercoiled domains is however complicated and the resulting conformation presumably affects gene expression patterns. Torsional stress is also introduced locally by processes involving separation of DNA strands, such as transcription and replication. While supercoils formed through such processes have shorter lifetimes, they presumably affect transcription dynamics. There are a family of enzymes called topoisomerases that manage the DNA topology which includes formation and relaxation of supercoils. Even though the phenomenon is readily observed in experiments, the dynamics of supercoils and the exact mechanism through which they regulate gene expression are currently unknown.

With the advent of single molecule experiments, single linear DNA molecules under tension were observed to go through a buckling transition beyond a threshold torsional stress [24]. A more recent experiment suggested that over wound linear DNA, beyond the buckling threshold, localizes excessive torsional stress into supercoiled regions, also called plectonemes. [7]. Local density peaks, interpreted as plectonemic domains that were induced by over winding DNA were observed using fluorescence microscopy with millisecond time resolution, yielding new information on supercoiling dynamics. Observed plectonemes demonstrated sudden displacements, termed hopping, in addition to regular diffusive displacements. Although it provided unique insights, spatial resolution

of the experiment was limited by optical diffraction; as a result, conformations associated with supercoiled regions and the mechanism behind hopping remained unknown.

A DNA strand as a twist-storing polymer can be analytically represented by a ribbon. A closed ribbon that is twisted about its central axis has a conserved quantity called linking number. Linking number can be decomposed into two components, namely twist and writhe, yielding the famous relation $Lk = Tw + Wr$ [25, 26]. Twist here represents the torsional stress stored locally and writhe represents the torsional stress expressed by the global conformation of the molecule, supercoils being a common form of this expression. From this analytical point of view, DNA distributes its linking number between twist and writhe through conformational changes. Unfortunately, some of the assumptions made in this treatment result in difficulties transforming it to supercoiled linear DNA [27].

In this work, we reproduced a recent single molecule experiment [7] in silico using a coarse grained model of double stranded DNA. We produced conformations associated with overwound supercoiled DNA and reproduced the experimentally observed hopping behavior in near atomic resolution.

2.2 Methods

2.2.1 Supercoiling nomenclature and analysis

DNA as a twist storing polymer can be analytically represented by a ribbon. A ribbon is described in Cartesian coordinates by axis vectors $\vec{r}(t)$ and a set of normal vectors $\vec{U}(t)$ where t is an internal parameter spanning the length of the ribbon. Unit tangent vectors $\vec{T}(t)$ are defined over $\vec{r}(t)$ given that $\dot{r}(t) > 0$ for all t. Vectors $\vec{U}(t)$ pointing to the edges of the ribbon are then chosen to be orthonormal with respect to $\vec{T}(t)$.

In our implementation we started with two sets of n vectors $\vec{R}_i$ and $\vec{R}'_i$ describing the complementary backbones. We defined $\vec{r}_i$ as the centerline of the DNA (Eq. 2.1), and calculated $\vec{T}_i$ by normalizing the numerical derivative of $\vec{r}_i$ with appropriate boundary treatment (Eq. 2.2). We then calculated $\vec{U}_i$ by orthonormalizing backbone-to-backbone vectors $\vec{u}_i$ with respect to $\vec{T}_i$ (Eq. 2.3) and finally calculated $d\vec{U}_i$ by numerical derivation of U_i with appropriate boundary treatment (Eq. 2.4).

$$\vec{r}_i = \frac{\vec{R}_i + \vec{R}'_i}{2} \quad \text{where } \vec{R} \text{ and } \vec{R}' \text{ are complementary backbones} \tag{2.1}$$

$$\vec{T_i} = \begin{cases} \frac{\vec{r}_{(i+1)} - \vec{r}_i}{|\vec{r}_{(i+1)} - \vec{r}_i|} & i = 0 \\ \frac{\vec{r}_{(i+1)} - \vec{r}_{(i-1)}}{|\vec{r}_{(i+1)} - \vec{r}_{(i-1)}|} & 0 < i < n \\ \frac{\vec{r}_i - \vec{r}_{(i-1)}}{|\vec{r}_i - \vec{r}_{(i-1)}|} & i = n \end{cases} \tag{2.2}$$

$$\vec{U_i} = \frac{\vec{u_i} - (\vec{u_i} \cdot \vec{T_i})\vec{u_i}}{|\vec{u_i} - (\vec{u_i} \cdot \vec{T_i})\vec{u_i}|} \text{ where } \vec{u_i} = \vec{R_i} - \vec{R_i'} \tag{2.3}$$

$$d\vec{U}_i = \begin{cases} \vec{U}_{(i+1)} - \vec{U}_i & i = 0 \\ \frac{\vec{U}_{(i+1)} - \vec{U}_{(i-1)}}{2} & 0 < i < n \\ \vec{U}_i - \vec{U}_{(i-1)} & i = n \end{cases} \tag{2.4}$$

The linking number (Ln) of a ribbon is the number of pitches (λ) the ribbon accomodates along its length. Linking number is a topological constant for a closed ribbon that is conserved for all conformations given that the ribbon cannot pass through itself. It is also conserved for an open ribbon with fixed ends with the additional condition that it cannot loop over the ends. Linking number can be decomposed into two distinct quantities twist (Tw) and writhe (Wr) (Eq. 2.5) [25, 26].

$$Ln = \frac{n}{\lambda} = Tw + Wr \quad \text{where} \quad \lambda > 0 \tag{2.5}$$

The excess or deficient link on the ribbon is then quantified by σ which we calculated by comparing the pitch (λ) of the ribbon with a reference (neutral) pitch (λ_0) (Eq. 2.6). It should be noted here that double stranded DNA assumes a right handed helical structure under physiological conditions. The convention therefore associates positive σ values with overwound DNA, implying that the torsional stress is in the right handed direction with respect to the axis of the polymer and vice versa.

$$\sigma = \frac{Ln}{Ln_0} - 1 = \frac{\lambda_0}{\lambda} - 1 \tag{2.6}$$

Total twist (Tw) of the ribbon is defined by the integral of angular twist rate (ω) over the axis (Eq. 2.7). Angular twist rate is a local property, therefore the definition of total twist is valid for both open and closed ribbons. In our implementation we calculated angular twist rate for each base step, discretely approximating total twist (Eq. 2.8). We also intorduced δTw to quantify the relation of Tw to the σ of the system (Eq. 2.9).

$$Tw = \frac{1}{2\pi} \int \omega \, \mathrm{d}t \quad \text{where} \quad \omega(t) = \frac{\mathrm{d}\vec{U}}{\mathrm{d}t} \times \vec{U} \cdot \vec{T} \tag{2.7}$$

$$Tw \approx \frac{1}{2\pi}\sum^{n}\omega_i \quad \text{where} \quad \omega_i = d\vec{U}_i \times \vec{U}_i \cdot \vec{T}_i \tag{2.8}$$

$$\delta Tw = \frac{\Delta Tw}{Ln_0} = \frac{Tw}{Ln_0} - 1 = \frac{Tw\lambda_0}{n} - 1 \tag{2.9}$$

Writhe (Wr), on the other hand, is a global property of the ribbon that is a function of the axis only. The original definition of total writhe consists of a double integration of a non physical quantity (Eq. 2.10) and is only valid for closed ribbons [25]. A method to calculate total writhe incrementally using a single integral was later reported (Eq. 2.11) but this method is not applicable to supercoiled open ribbons due to assumptions involved in the derivation [26, 27]. As a result, even though we could calculate total writhe by subtracting the total twist from the linking number ($Wr = Ln - Tw$), we could not calculate the local writhe distribution for a linear supercoiled ribbon, which was the main object of our work.

$$Wr = \frac{1}{4\pi}\oint\oint \vec{T}(t) \times \vec{T}(t') \cdot \frac{\vec{r}(t) - \vec{r}(t')}{|\vec{r}(t) - \vec{r}(t')|^3}\ \mathrm{d}t\ \mathrm{d}t' \tag{2.10}$$

$$Wr = Wr_0 + \frac{1}{2\pi}\int \frac{\vec{T}_0(t) \times \vec{T}(t)}{1 + \vec{T}_0(t) \cdot \vec{T}(t)} \cdot \frac{\mathrm{d}}{\mathrm{d}t}[\vec{T}_0(t) + \vec{T}(t)]\ \mathrm{d}t \tag{2.11}$$

where Wr_0 is the writhe for a reference ribbon with unit tangents $\vec{T}_0$

Curvature (κ) of any curve is defined as the magnitude of the rate of change of the tangent vectors. We used κ as an alternative local quantity for characterization of supercoiled domains. As local curvature fluctuations of DNA were significant due to the off-centered helical structure, we further coarse grained the axis using a grain size (g) of 11 bases, effectively calculating κ for each pitch.

$$\kappa(t) = \left|\frac{\mathrm{d}\vec{T}}{\mathrm{d}t}\right| \approx \kappa_j = \frac{\left|\vec{T}'_{(j+1)} - \vec{T}'_{(j-1)}\right|}{2} \tag{2.12}$$

where $\vec{T}'_j$ is the unit tangent for coarse grained axis $\vec{r'}_j = \frac{1}{g}\sum_{i=gj}^{gj+g}\vec{r}_i$

End-to-end distance (R_{e2e}) is one of the few variables that is directly observable in experiments. We therefore used the size normalized R_{e2e} (Eq. 2.13) for the systems to

directly compare our results to experiments, $< R_{e2e} >$ being the time average of R_{e2e} calculated from complete trajectories.

$$R_{e2e}(t) = \frac{|\vec{r}_n(t) - \vec{r}_1(t)|}{n} \quad \text{where } n \text{ is chain length} \tag{2.13}$$

We implemented all calculations with Python, making extensive use of NumPy and SciPy scientific computing libraries [28] in addition to MDAnalysis library [29] which exposes binary GROMACS trajectories to Python as NumPy arrays. We also made extensive use of the iPython interactive python shell [30] and automated parts of analysis using Bash scripting which is native to the Linux environment we were using for the entire process.

2.2.2 Coarse grained model

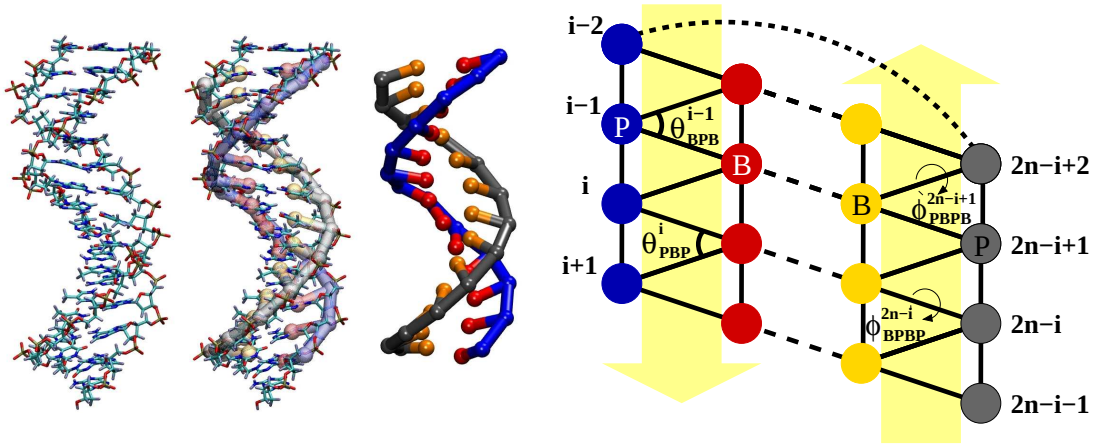


Figure 2.1: Illustration for the 2 bead CG model. A single bead was used to represent the sugar-phosphate backbone, instead of two beads.

We used a 2 bead coarse grained model that was previously developed in our group [9]. This model used a single bead to represent the backbone monomer, rather than distinct sugar and phosphate groups we used in the model described in the previous chapter. It also had all harmonic bonds, yielding a strictly double stranded model. This model was originally used to simulate supercoiled DNA minicircles, validating it for representation of supercoiled DNA structures [9].

We ported the model to GROMACS MD simulation package for higher performance and scalability as it was originally designed for ESPRESSO MD package. We simulated

a 100 bp long free ended DNA structure for 20 microseconds (μs) at 298 K for benchmarking purposes. We used a stochastic dynamics (SD) integrator that is reportedly appropriate for simulating dynamics of implicit solvent systems [20] with a timestep of 25 femtoseconds. We characterized this system to serve as a reference for the later simulations.

We then reproduced the experimental conditions using the following scheme : By setting the pitch for a dsDNA chain while generating the initial linear conformation, we defined its linking number and associated σ value. We then constrained the end points of this chain in X and Y axes; introducing grids of impermeable wall particles in the XY planes at either end, we represented a chain with constant liking number. After coupling the end points with the walls via harmonic bonds, we applied negative pressure coupling in Z axis using a Berendsen barostat [18], introducing tension in the chain axis. We optimized the relaxation time of the barostat by iteration and chose a value of 10 picoseconds. We calculated the required pressure by dividing the desired pulling force by the area of the wall in the XY plane (Eq. 2.14). We applied periodic boundary conditions in X and Y directions. Such a system effectively represented a dsDNA strand with constant linking number under tension, as were the conditions of the experiment that we reproduced [7].

Using the described scheme, we created 222, 444, 666 and 1000 bp long dsDNA chains with σ values ranging from 0 to 0.12, and pulling forces ranging from 1.5 to 9 pN. We then simulated these systems for up to 100 μs at 298 K using the same SD integrator and 25 fs timestep used for the reference system.

$$P = \frac{-F_{pull}}{A_{xy}} \tag{2.14}$$

2.3 Results

2.3.1 Equilibrium properties

We characterized the model using a 100 bp long chain we simulated for $20\mu s$ as the reference system. We calculated the reference pitch λ_0 from this system by fitting an analytical form to the bond correlations along the backbone (Eq. 1.10). The least squares fit yielded a λ_0 of 11.2 bp and a l_p of 89 bp.

We then proceeded to calculate the excess twist (δTw) on the chain, expecting fluctuations around 0. We however observed that δTw fluctuated around -0.132, implying

that the model was biased toward writhe against twist. This suggested that our model tended to constantly express around %13 of its total link as writhe, rather than evenly distributing it between twist and writhe. Upon consistently observing the same bias of %13 for all systems, we corrected all further δTw measurements with respect to this bias for simplification of comparison and interpretation of data (Eq. 2.15).

$$\delta Tw' = \delta Tw + 0.132 \tag{2.15}$$

(a) Bond correlations (b) δTw vs. time

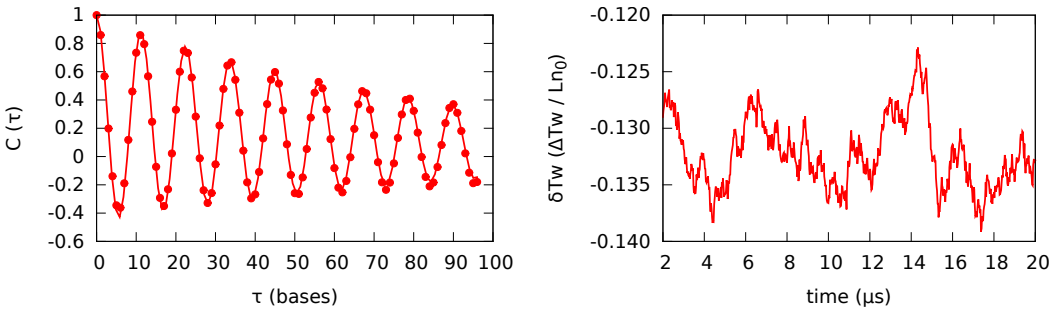


Figure 2.2: Pitch (λ), persistence length (l_p) and δTw calculated for a 100 base-long free ended chain. A pitch of 11.2 bases and l_p of 89 bases were calculated via least squares fitting of an analytical form to the bond correlation data (Eq. 1.10). δTw fluctuated around -0.132.

Average end-to-end distance ($< R_{e2e} >$), being the only variable that was directly observable in earlier experiments, was straightforward to calculate from simulations. We therefore calculated ($< R_{e2e} >$) for a each simulation and reproduced experimentally observed buckling transition upon increasing σ (Fig. 2.3) [24, 7].

We also calculated the average twist ($< \delta Tw' >$) on the chain for a given linking number and force (Fig. 2.4). The buckling transition was evident in $< \delta Tw' >$ data as an overshoot followed by a plateau with increasing σ. Tw, therefore δTw is not measurable in experiments, but overshooting behaviour was consistent with torque overshoots predicted by statistical models [8].

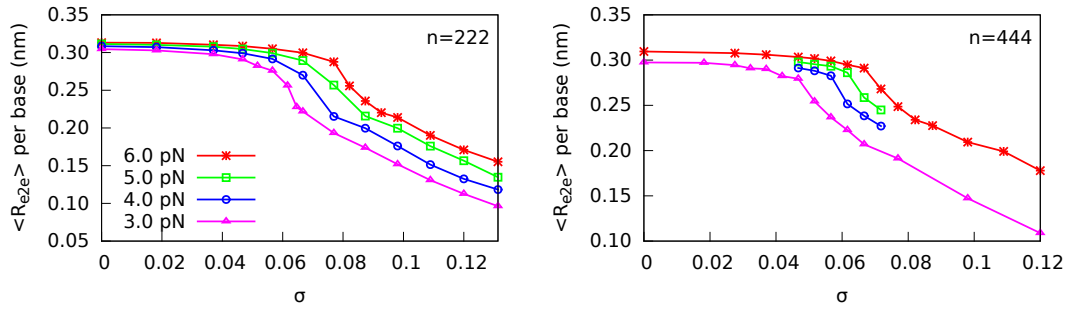


Figure 2.3: Average end-to-end distance ($< R_{e2e} >$) vs. σ calculated over 50 μs for 222 and 444 bp long systems. Threshold σ value where buckling occurs increases with increasing pulling force.

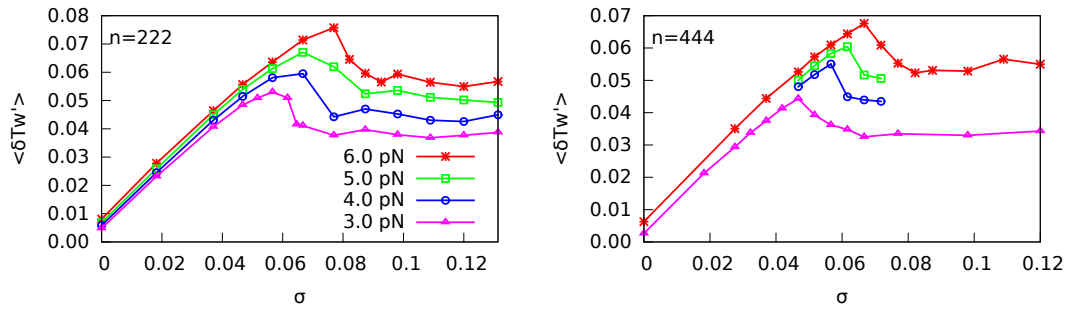


Figure 2.4: Average total twist ($< \delta Tw' >$) vs. σ calculated over 50 μs for 222 and 444 bp long systems. Overshooting behaviour is consistent with statistical models and the magnitude represents the excess twist chain can absorb before buckling.

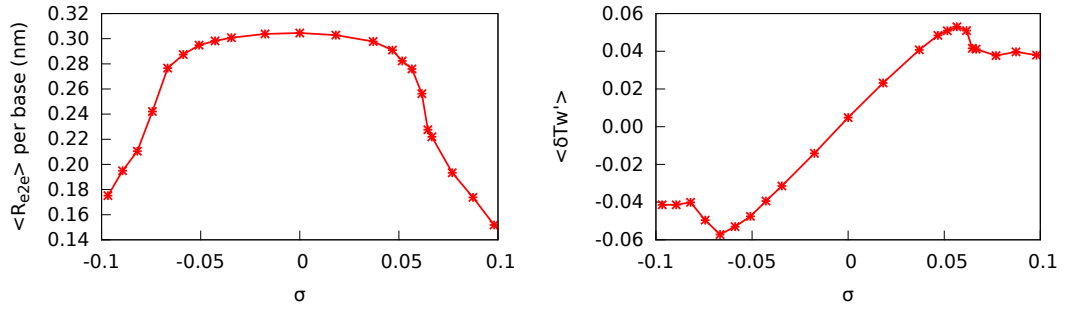


Figure 2.5: $< R_{e2e} >$ and $< \delta Tw' >$ vs. σ for 222 base long chain under 3 pN pulling force, including underwound ($\sigma < 0$) systems. The nearly symmetrical behaviour around $\sigma = 0$ is presumably an artifact due to the strictly double stranded model.

2.3.2 Effect of length

We observed finite size effects when plectoneme sizes were comparable with total chain length. These effects were especially well pronounced in shorter chains with higher σ values. Although with large enough σ values, any chain will be dominated with plectonemic domains, we were not interested in such systems. We were interested in systems near buckling transition, which demonstrated dynamical behaviour in terms of plectoneme formation, deformation and hopping events, mostly dealing with single looped plectonemes. We showed that the finite size effects decay rapidly for chains longer than 444 bps (Fig. 2.6) for the σ.

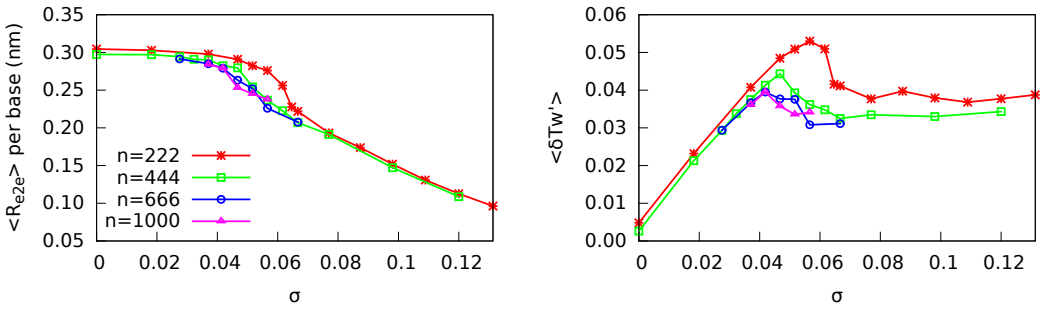


Figure 2.6: $< R_{e2e} >$ and $< Tw >$ vs. σ compared for 222, 444, 666 and 1000 bp long chains under 3 pN pulling force. The finite size effects result in significant differences between 222 and 444 base long chains but decay rapidly for 666 and 1000 base long chains.

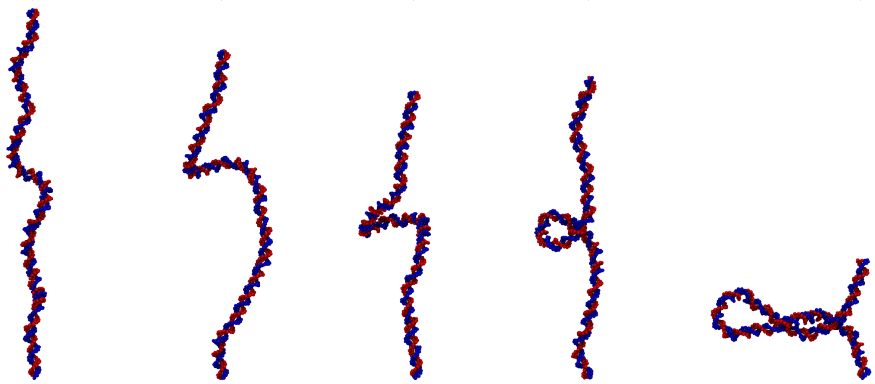


Figure 2.7: Conformations for a 222 bp long system demonstrate that the system is quickly dominated by plectonemic domains upon buckling.

We also observed that some σ - pulling force combinations yielded bimodal states near buckling transition whereas others buckled smoothly. Although we do not have enough data to demonstrate the exact relation, statistical models suggest that bimodality should only be expected for a narrow force window for each chain length [8].

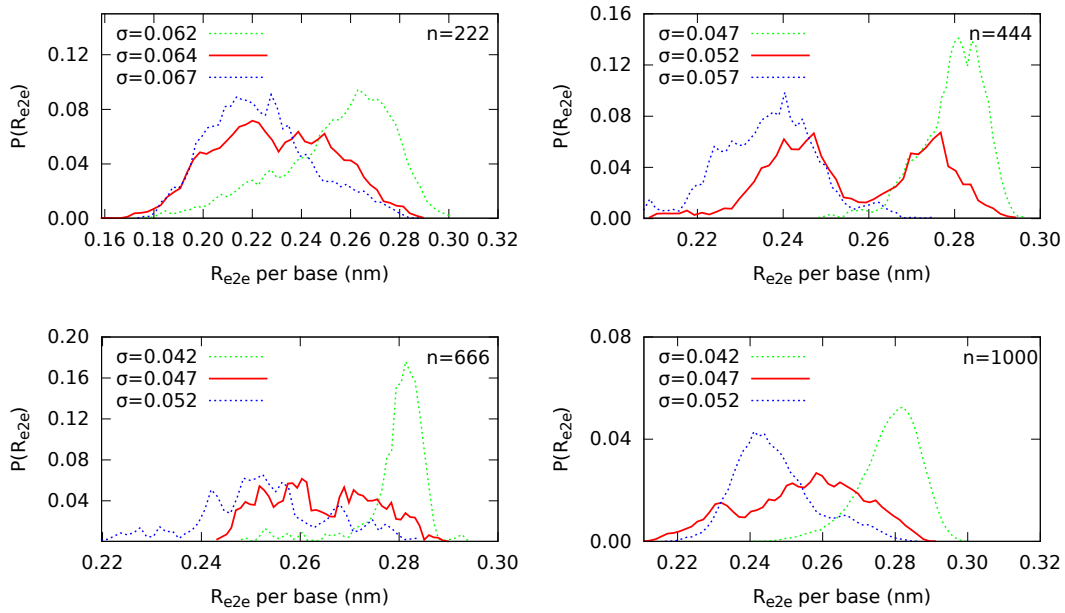


Figure 2.8: R_{e2e} histograms for 222, 444, 666 and 1000 bp long chains under 3 pN pulling force. 222 and 1000 base long systems buckle smoothly whereas the 444 and 666 base long systems buckle through a bimodal transition state. This behaviour is dependant on chain length and pulling force.

2.3.3 Dynamical properties

We made timeline measurements of R_{e2e} and δTw for each system (Fig. 2.9). Small σ systems showed small and fast fluctuations indicating a well equilibrated system whereas systems with higher σ showed lower frequency fluctuations with higher magnitude, suggesting bistability. We found a strong correlation between flucutations in Tw and R_{e2e}.

Rare higher magnitude jumps in R_{e2e} were associated with formation of plectonomes, and were abundant near the buckling regime. Smaller magnitude jumps abundant in all systems were associated with coil formations rather than plectonomes.

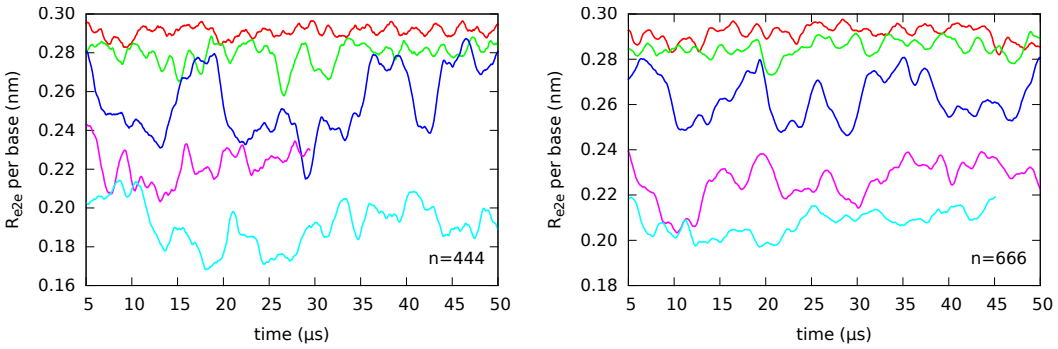


Figure 2.9: Timeline data for R_{e2e} for 444 and 666 bp long chains under 3 pN pulling force. Data is smoothed via moving averaging with a window size of 1 μs. Color key is identical with Figure 2.10 for respective chain lengths

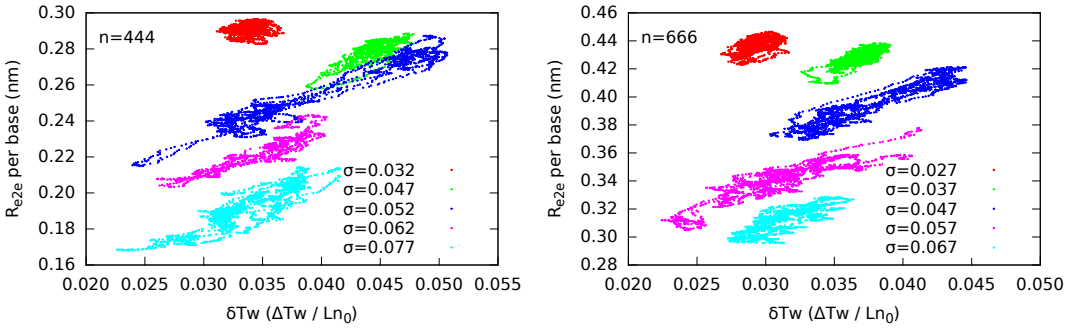


Figure 2.10: R_{e2e} vs. Tw for 444 and 666 bp long chains under 3 pN pulling force. A strong linear correlation is observed between R_{e2e} and Tw. Bimodality is observed in the transition systems.

2.3.4 Conformations

Even though experimental observations such as the buckling transition in $< R_{e2e} >$ and density peaks observed via fluorescence microscopy were interpreted as plectoneme formations, no experiment offered the resolution to sample actual conformations of single molecules. Visual representation of trajectories provided a qualitative description of the system behaviour. We were able to observe the conformational preferences of unbuckled systems (Fig. 2.11). Despite our inability to sample them to statistical significance, we were able to identify events such as plectoneme formation (Fig. 2.12) and hopping (Fig. 2.13). We were also able to qualitatively observe the nature of finite size effects (Fig. 2.7) and plectoneme dominated systems (Fig. 2.14)

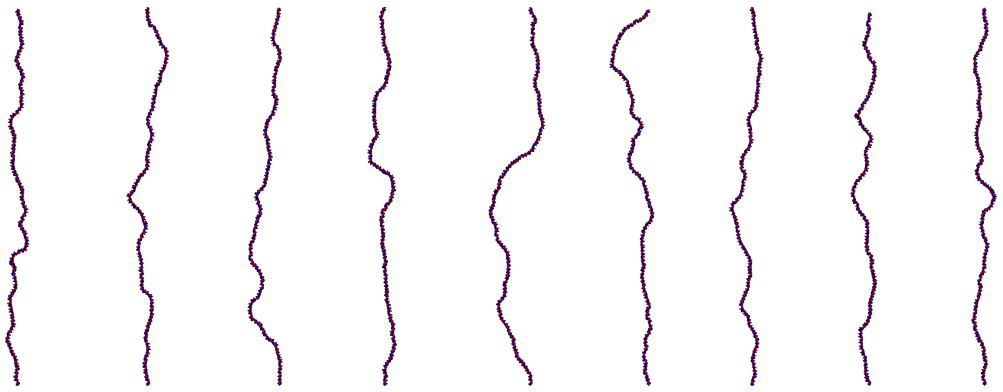


Figure 2.11: Conformations for a small σ system. Snapshots from 666 bp long chain with $\sigma = 0.037$ under 3pN pulling force. This system is below the threshold σ value. Loose coil formations were preferred conformations for all such systems, presumably reflecting the writhe bias of the model.

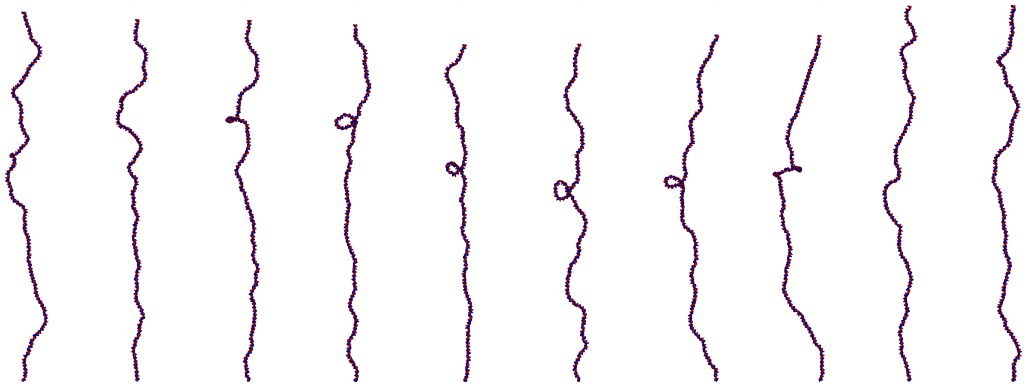


Figure 2.12: Conformations through a plectoneme formation and dissipation event. Snapshots from 666 bp long chain with $\sigma = 0.042$ under 3pN pulling force. This is a true plectoneme formation and dissipation process, with no stable plectoneme formation before or after. Such events were wery rare.

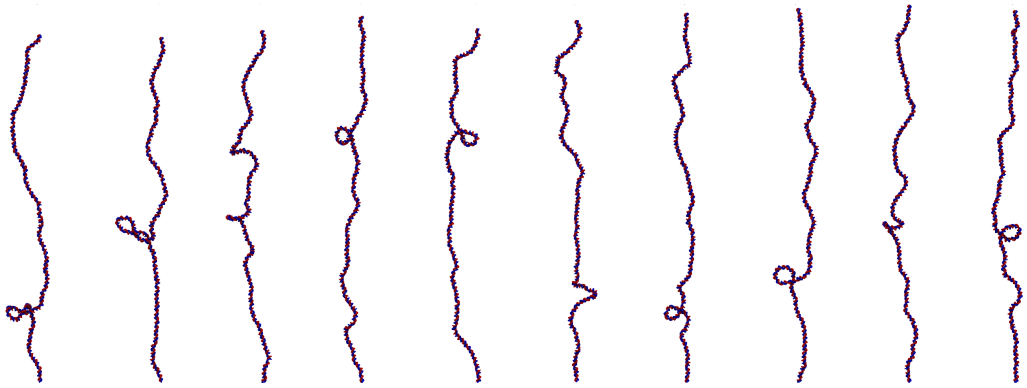


Figure 2.13: Conformations through a hopping event. Snapshots from 666 bp long chain with $\sigma = 0.052$ under 3pN pulling force. This is a system near buckling transition that demonstrates dynamical behaviour. Hopping events such as this one displace the plectonemes faster than diffusive motion typically does.

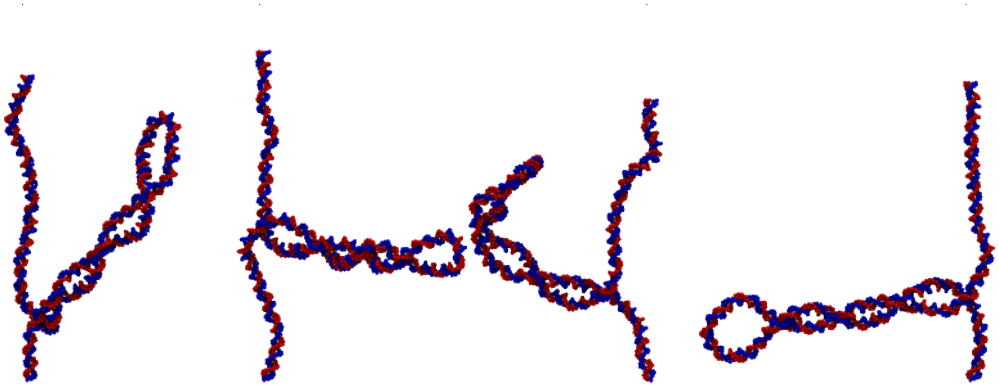


Figure 2.14: Conformations for a plectoneme dominated system. Snapshots from 444 bp long chain with $\sigma = 0.12$ under 3pN pulling force. As lifetime increases and diffusive displacement slows with increasing plectoneme size, the plectonemic domain begins to explore its own confirmational space.

2.3.5 Kymographs

Drawing kymographs is an efficient method for representing the time evolution of a distribution. We prepared kymographs for the time evolution of density and curvature distributions, resulting in 2 dimensional images where $f(x,t)$ is color coded and t, x variables are in the x and y axes respectively.

We defined $\rho(z,t)$ as the time dependent particle density along the Z axis, and produced kymographs visualizing the time evolution of particle density. Resulting images are analogous to the density kymographs produced by Dekker et al. in their single molecule experiments. Plectonemes showed up as density peaks in these kymographs which we confirmed via visual representations of trajectories. We observed formation and diffusion of plectonemes with increasing linking number. Hopping events, on the other hand, were rare and confined to a narrow region near buckling regime.

We defined $\kappa_j(t)$ as the time dependent coarse grained curvature of pitch j (Eq. 2.12) and produced kymographs visualizing the time evolution of curvature. The $\kappa_j(t)$ kymographs, being dependent on the internal parameter j instead of an external reference frame, yielded different signals corresponding to plectonemic domains. Plectonemic domain was characterized by three distinct peaks in curvature, signifying deviations from the linear conformation and the looping middle of the plectoneme.

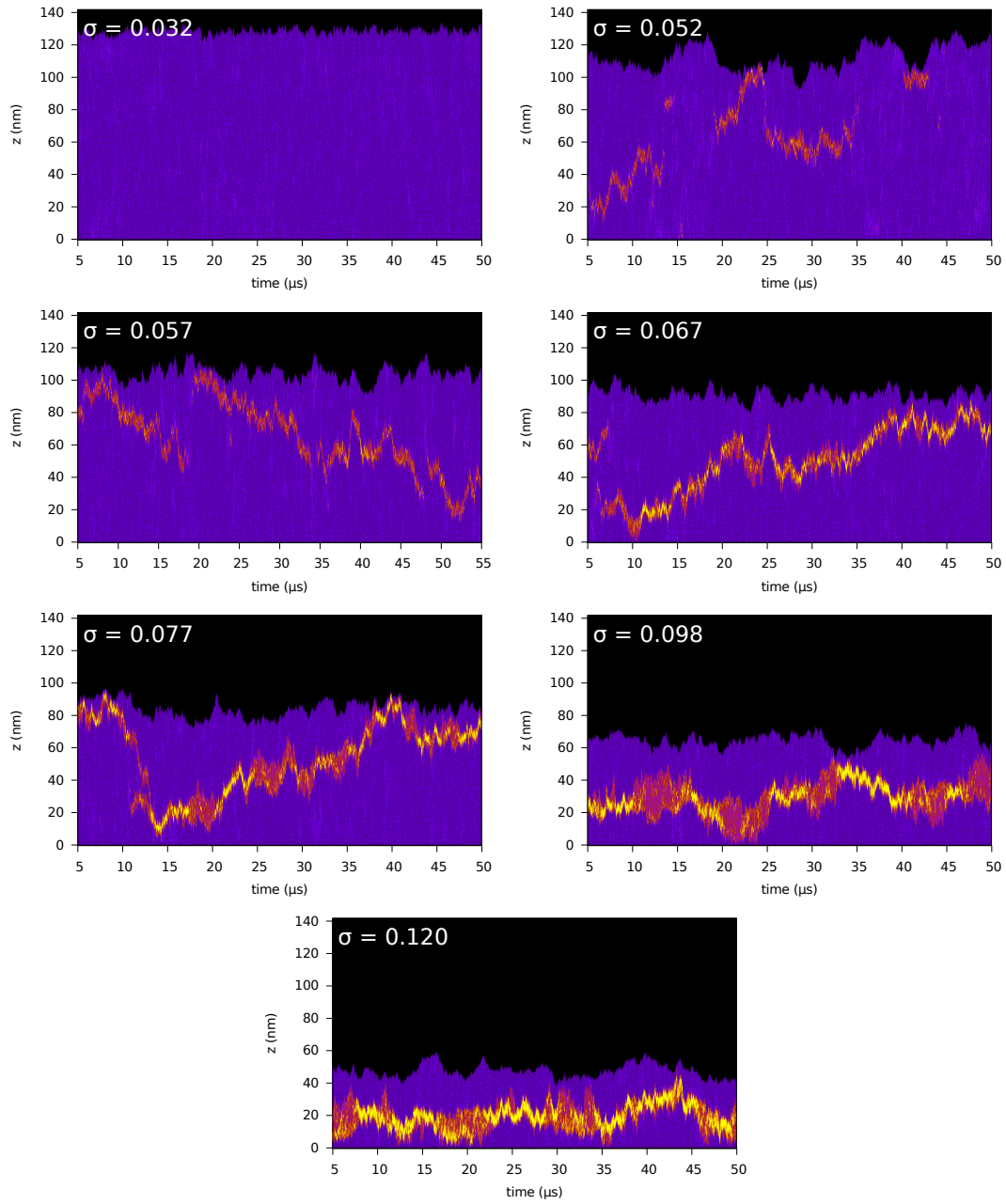


Figure 2.15: Density kymographs for 444 bpl chain under 3 pN pulling force with various σ values. This data is directly comparable to experimental observations [7]. We observed alternating formation and dissipation of plectonemes at $\sigma = 0.05$ which approximately represents the buckling threshold for this system. Plectonemes did not form for smaller σ values and were stable once formed for higher σ values. We observed a hopping event in addition to diffusion at $\sigma = 0.08$, which is just above the threshold σ.

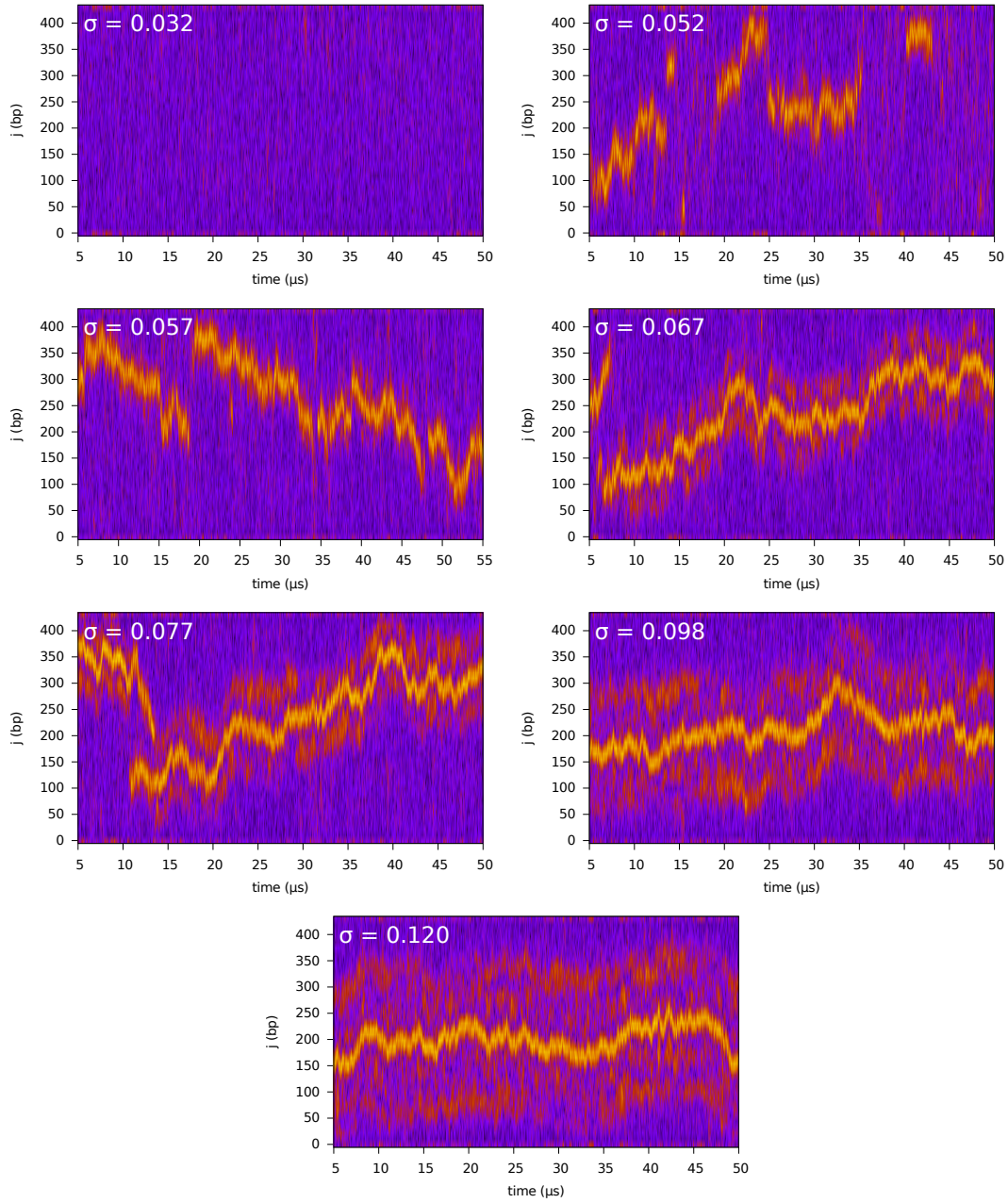


Figure 2.16: Curvature (κ_j) kymographs for 444 bpl chain under 3 pN pullling force with various σ values. Coarse grained with approximately one pitch (11 base) grain size.

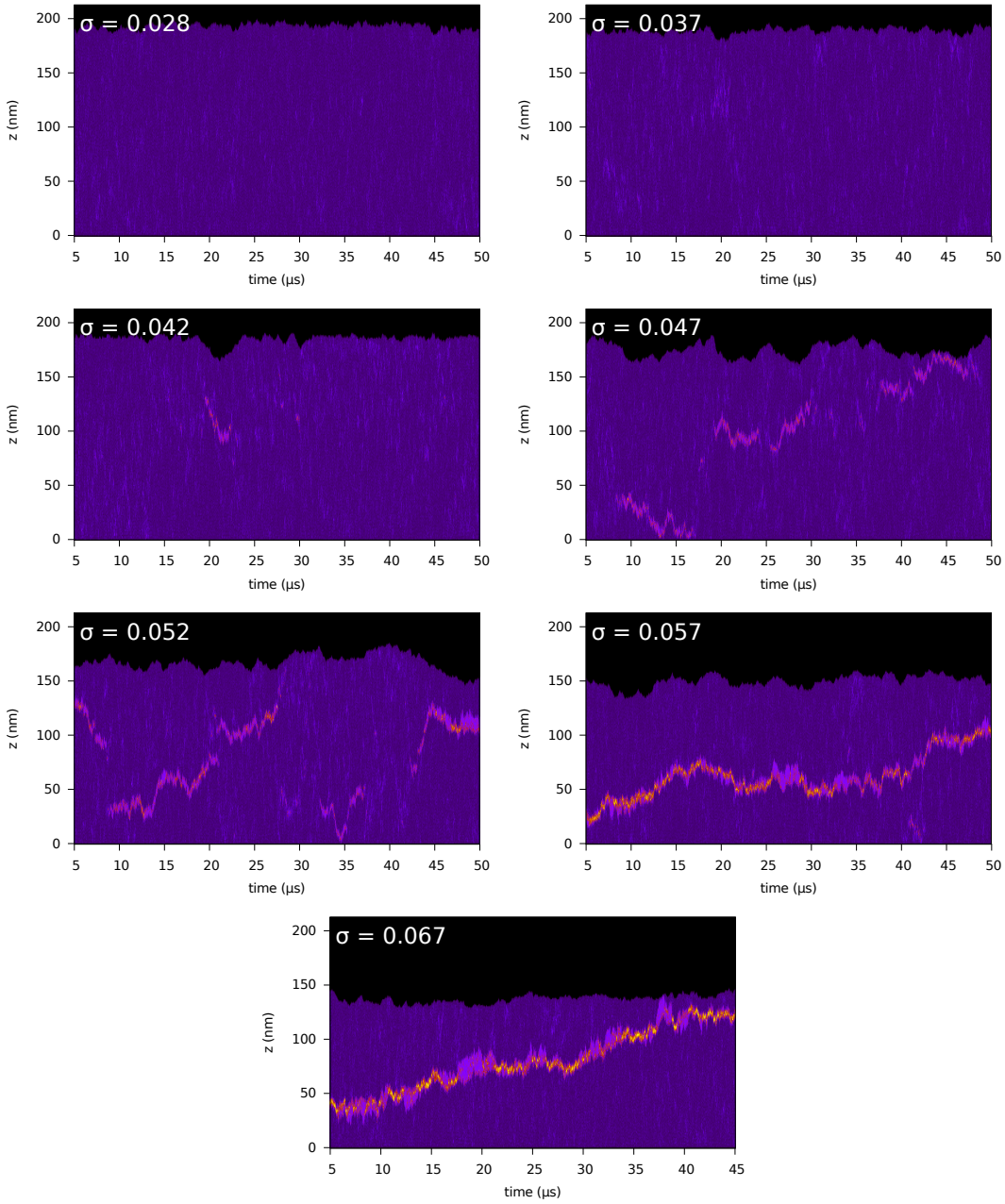


Figure 2.17: Density kymographs for 666 bpl chain under 3 pN pulling force with various σ values. This data is directly comparable to experimental observations [7].

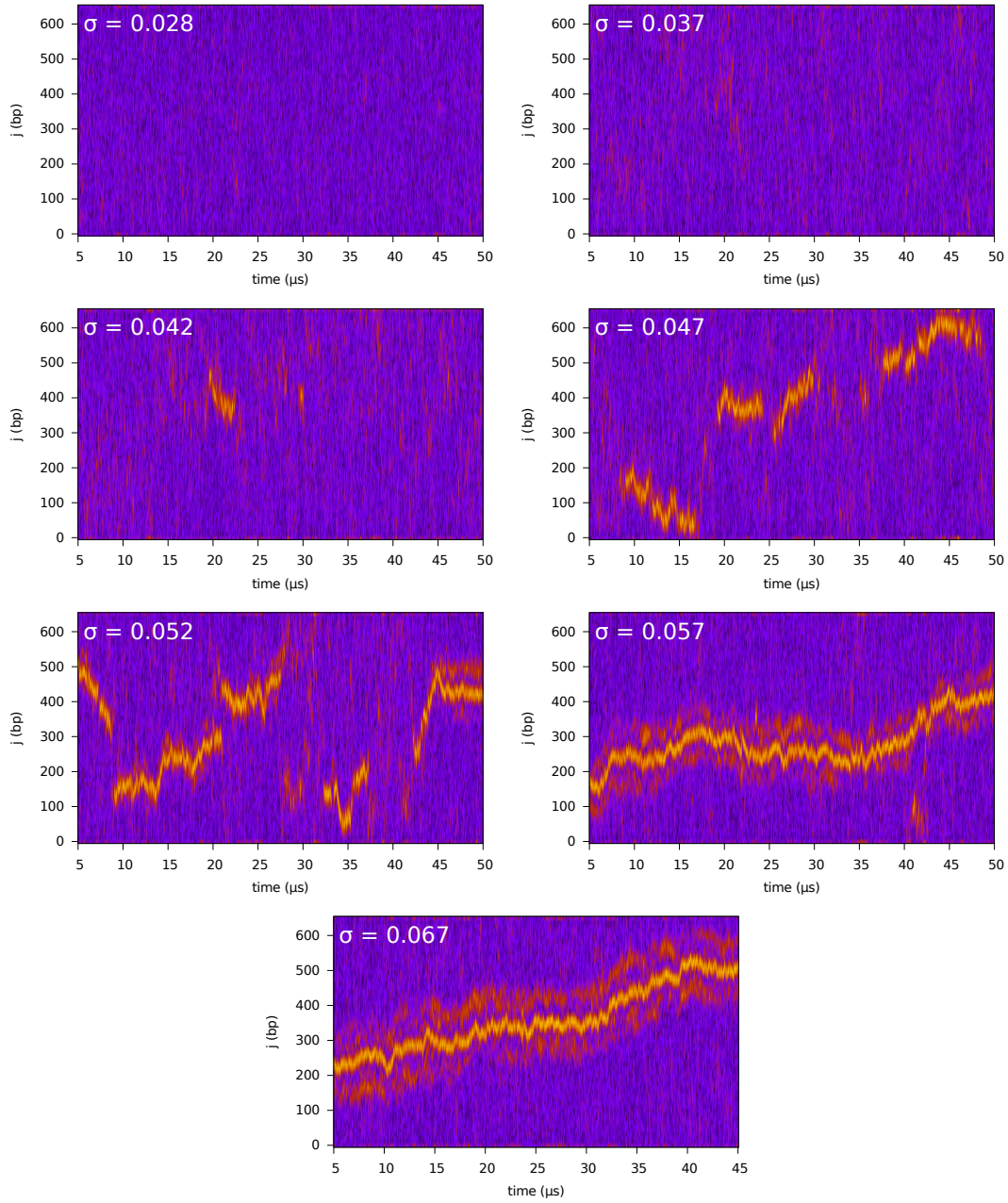


Figure 2.18: Curvature (κ_j) kymographs for 666 bpl chain under 3 pN pullling force with various σ values. Coarse grained with approximately one pitch (11 base) grain size.

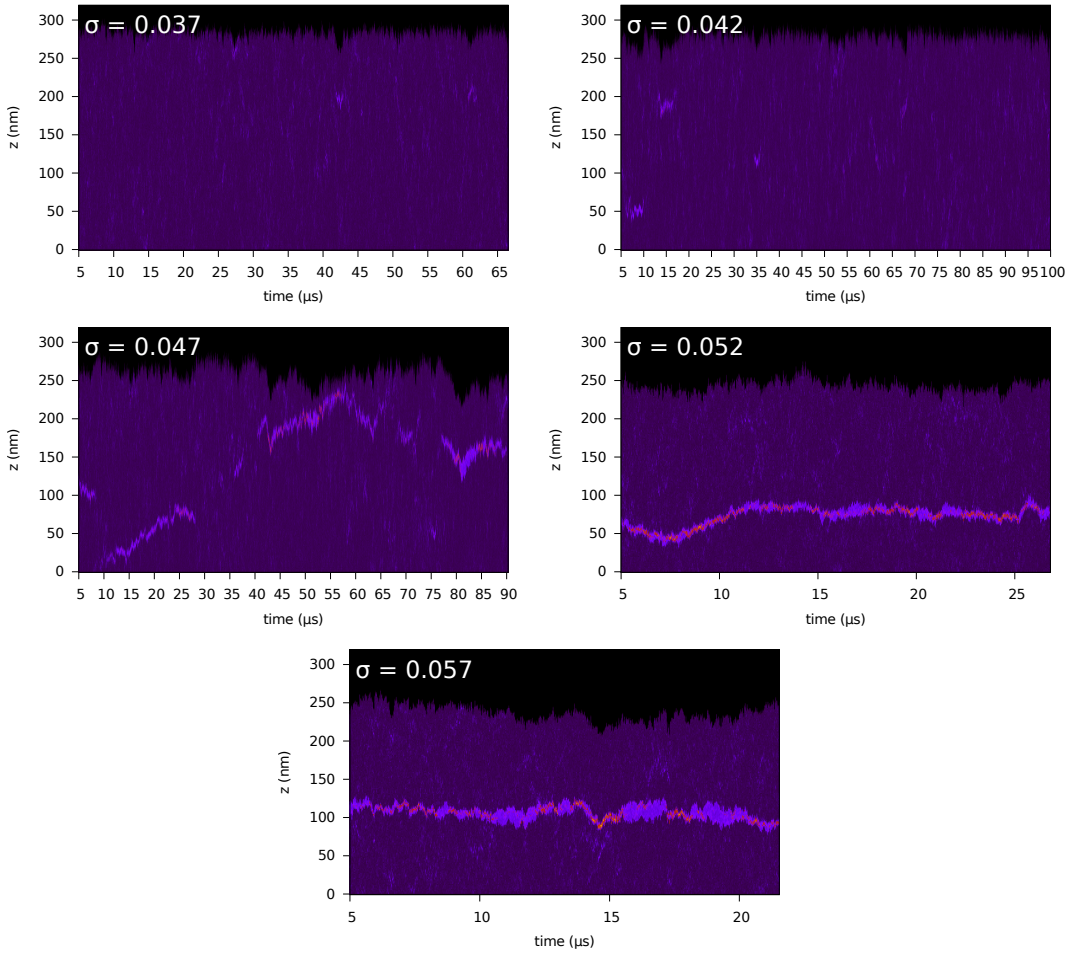


Figure 2.19: Density kymographs for 1000 bpl chain under 3 pN pulling force with various σ values. This data is directly comparable to experimental observations [7].

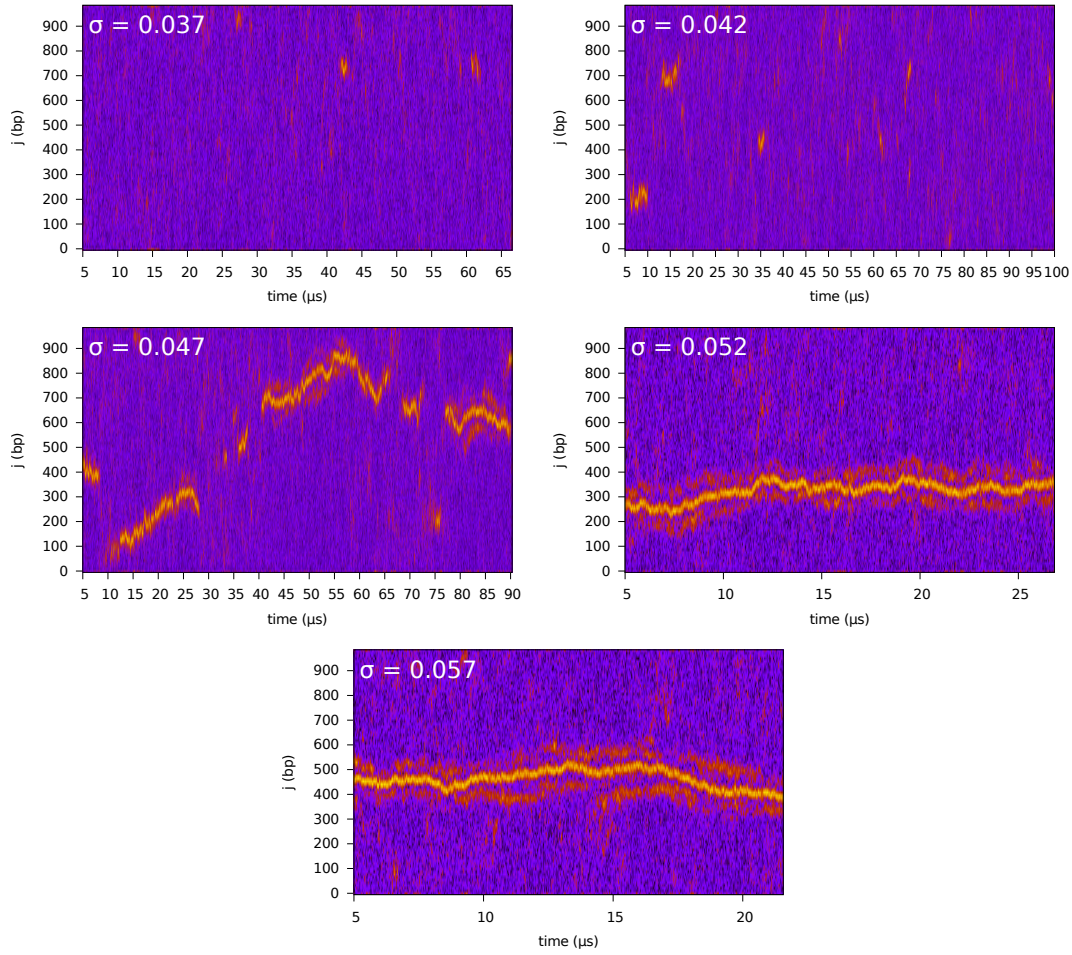


Figure 2.20: Curvature (κ_j) kymographs for 1000 bpl chain under 3 pN pullling force with various σ values. Coarse grained with approximately one pitch (11 base) grain size.

We also prepared kymographs for the time dependent angular twist rate distribution $\omega(i,t)$, but these yielded no information as ω was uniformly distributed on the chain for all systems. There was some weak time correlation in plectoinemic domains, suggesting that twist fluctuations were frozen in such domains.

2.4 Discussion and Conclusions

Performance Porting the previously described model to GROMACS significantly improved both performance and scalability. Microsecond timescales on kilobase (kb) long chains were accessible using locally available computers. We however chose 444 and 666 base long chains as our primary systems as reduced computational cost allowed us to span a larger space with respect to both σ and pulling force. 50 μs simulations of these chains yielded well equilibrated trajectories and took about a week on 8 (2 to 3 GHz) processors. We spatially decomposed the systems in the Z axis for parallel computation, which introduced some load imbalance as particle density was localized as plectonemes formed in supercoiled systems. The scalability for supercoiled systems were therefore lower than linear systems.

The shortest chain lengths available for single molecule experiments are in the order of few kilobases while chain lengths around tens of kilobases are more commonly used. In this respect, we were working roughly an order of magnitude below the experimentally relevant length scales. We did however observe that finite size effects decay rapidly for chain sizes above 222 bases.

Our time scales, on the other hand, were not directly comparable with experiments as our model involves implicit solvent representation. Even though the SD integration scheme we used was reported to successfully reproduce dynamics of implicit solvent systems [20], we did not tune the friction coefficients associated with solvent interactions for any particular reference. The preliminary calculations we made on diffusion coefficients of plectonemes were therefore off by several orders of magnitude.

Supercoiling dynamics Buckling transition and associated supercoil formation in overwound DNA was observed indirectly in single molecule experiments first as an abrupt decrease in end-to-end distance [24] and then recently as a local density peaks along the length of the molecule [7]. The characteristics of the associated plectonemes were further predicted by statistical modeling [8], the time dependent behaviour and the exact conformations were however not available to our knowledge. Our work readily reproduced the behaviour described in recent experiments while being in reasonable agreement with statistical predictions. As a result we have a unprecedented description of time dependent behaviour of overwound DNA strands under tension in molecular resolution.

Upon recreating the experimental conditions in silico, we showed that the density localizations observed experimentally do indeed correspond to plectonemic domains. We found the buckling threshold σ value to be proportional with the amount of tension and inversely proportional with chain length. Buckling transition was smooth for some systems whereas bistable transition states were observed for other systems. This behaviour was dependant on both chain length and pulling force. We presume that bistable states exist in a narrow pulling force window dependant on chain length although we do not have enough data to suggest an exact relation.

We observed hopping events in systems near the buckling transition. This behaviour was in good agreement with experiments. Hopping events were most abundant in 666 base-long systems, they were however very rare in general and proved infeasible to sample to statistical significance. We postulated that hopping events were due to the fast-diffusion of a topological quantity that cannot be measured in experiments, writhe being the first suspect. A valid mathematical definition for local writhe on open curves however does not exist for the particular system we were working with as plectonemic domains violate certain assumptions involved in all derivations of local writhe to our knowledge. [26, 31, 27].

We analysed the data with respect to local curvature which is directly related to bending energy of the conformation. Local curvature fluctuations were correlated with particle density with respect to the chain axis but provided additional information on plectoneme structure. A typical plectoneme was represented by three distinct peaks in curvature; first and last peaks signified the deviation from linear conformation and the middle peak, higher in magnitude, signified the looping section of the plectoneme. Curvature data was immune to artifacts introduced into density data by longer plectonemes that aligned with the polymer axis and therefore reliably estimated plectoneme size.

Time dependent measurements of R_{e2e}, δTw, and density entropy yielded strongly correlated signals for all systems, providing no additional information.

Total twist was uniformly distributed along the chain for all systems, suggesting that it did not contribute significantly to the formation of plectonemes. There was, however, a weak time correlation in the plectonemic domains with respect to angular twist rate, suggesting that twist fluctuations were frozen within such domains.

BIBLIOGRAPHY

[1] F. Crick and J. Watson, "Molecular structure of nucleic acids," *Nature*, 1953.

[2] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. r. Kjems, "Self-assembly of a nanoscale DNA box with a controllable lid.," *Nature*, vol. 459, pp. 73–6, May 2009.

[3] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns.," *Nature*, vol. 440, pp. 297–302, Mar. 2006.

[4] B. Wei, M. Dai, and P. Yin, "Complex shapes self-assembled from single-stranded DNA tiles.," *Nature*, vol. 485, pp. 623–6, May 2012.

[5] L. Adleman, "Molecular computation of solutions to combinatorial problems," *SCIENCE-NEW YORK THEN WASHINGTON-*, vol. 226, no. 5187, 1994.

[6] C. Bustamante, Z. Bryant, and S. Smith, "Ten years of tension: single-molecule DNA mechanics," *Nature*, vol. 421, no. January, 2003.

[7] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker, "Dynamics of DNA supercoils.," *Science (New York, N.Y.)*, vol. 338, pp. 94–7, Oct. 2012.

[8] J. F. Marko and S. Neukirch, "Competition between curls and plectonemes near the buckling transition of stretched supercoiled DNA," *Physical Review E*, vol. 85, p. 011908, Jan. 2012.

[9] M. Sayar and B. Avarolu, "Twist-writhe partitioning in a coarse-grained DNA minicircle model," *Physical Review E*, vol. 81, pp. 1–9, Apr. 2010.

[10] P. D. Dans, A. Zeida, M. R. Machado, and S. Pantano, "A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics," *Journal of Chemical Theory and Computation*, vol. 6, pp. 1711–1725, May 2010.

[11] A. Morriss-Andrews, J. Rottler, and S. S. Plotkin, "A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality.," *The Journal of chemical physics*, vol. 132, p. 035105, Jan. 2010.

[12] T. E. Ouldridge, A. a. Louis, and J. P. K. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model.," *The Journal of chemical physics*, vol. 134, p. 085101, Feb. 2011.

[13] T. a. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, "A coarse grain model for DNA.," *The Journal of chemical physics*, vol. 126, p. 084901, Feb. 2007.

[14] R. C. DeMille, T. E. Cheatham, and V. Molinero, "A coarse-grained model of DNA with explicit solvation by water and ions.," *The journal of physical chemistry. B*, vol. 115, pp. 132–42, Jan. 2011.

[15] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free.," *Journal of computational chemistry*, vol. 26, pp. 1701–18, Dec. 2005.

[16] J. Wang, P. Cieplak, and P. a. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?," *Journal of Computational Chemistry*, vol. 21, pp. 1049–1074, Sept. 2000.

[17] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling.," *The Journal of chemical physics*, vol. 126, p. 014101, Jan. 2007.

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

[19] V. Ruhle and C. Junghans, "Versatile object-oriented toolkit for coarse-graining applications," *Journal of Chemical ...*, pp. 3211–3223, 2009.

[20] P. Hünenberger, "Thermostat algorithms for molecular dynamics simulations," *Advanced Computer Simulation*, pp. 105–149, 2005.

[21] S. Brinkers and H. Dietrich, "The persistence length of double stranded DNA determined using dark field tethered particle motion," *The Journal of chemical ...*, vol. 130, p. 215105, June 2009.

[22] Q. Chi, G. Wang, and J. Jiang, "The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measure-

ments using mean field theory," *Physica A: Statistical Mechanics and its Applications*, vol. 392, pp. 1072–1079, Mar. 2013.

[23] A. Pérez, I. Marchán, and D. Svozil, "Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of $\alpha/\ \gamma$ Conformers," *Biophysical journal*, vol. 92, pp. 3817–29, June 2007.

[24] S. Forth, C. Deufel, M. Sheinin, B. Daniels, J. Sethna, and M. Wang, "Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules," *Physical Review Letters*, vol. 100, pp. 1–4, Apr. 2008.

[25] F. B. Fuller, "The writhing number of a space curve.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, pp. 815–9, Apr. 1971.

[26] F. B. Fuller, "Decomposition of the linking number of a closed ribbon: A problem from molecular biology.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75, pp. 3557–61, Aug. 1978.

[27] S. Neukirch and E. Starostin, "Writhe formulas and antipodal points in plectonemic DNA configurations," *Physical Review E*, vol. 78, p. 041912, Oct. 2008.

[28] T. E. Oliphant, "Python for scientific computing," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 10–20, 2007.

[29] N. MichaudAgrawal, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," *Journal of Computational Chemistry*, 2011.

[30] F. Perez and B. E. Granger, "Ipython: A system for interactive scientific computing," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 21–29, 2007.

[31] R. Kamien, "Local writhing dynamics," *The European Physical Journal B-Condensed Matter ...*, vol. 4, pp. 1–4, 1998.

VITA

Murat Öztürk was born in İzmir, Turkey, on February 2, 1988. He graduated from İzmir Fen Lisesi in 2005, recieved a BSc Degree in Materials Science and Engineering from Sabancı University in 2010 and a MSc degree in Computational Sciences and Engineering from Koç University in 2013. He rides a motorcycle and likes being around trees and rivers.