MICROMECHANICS OF DNA UNDER SHARP BENDING

CONG PEIWEN

(B. Sci. (Hons.), NUS, 2008)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

COMPUTATION AND SYSTEMS BIOLOGY
SINGAPORE-MIT ALLIANCE

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLEARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety.
I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Cong Peiwen
October 14, 2014

Acknowledgment

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Yan Jie for the continuous support of my Ph.D study and research; for his encouragements, inspirations, criticisms which lead to motivate my curiosities, expose my weaknesses and breakthrough my limits. He also demonstrated what a wise and enthusiastic scientist could accomplish. I am very grateful to him for keeping me under his supervision for eight years, since my undergraduate, through which I gained a lot, way beyond scientific discoveries.

My sincere thanks also goes to Prof. Chen Hu, for offering stimulating discussions and valuable advices, as well as Dr. Dai Liang, for providing instructions on computer simulations and fruitful debates. I will also take the opportunity to thank Prof. Patrick Doyle for hosting my visit to MIT, during which I absorbed lots of fundamental knowledge from multiple perspectives.

I thank my fellow labmates in Group: Zhou Zhen, Hongxia, Yingjie, Lin Jie, Wenbo, Ding Ying, Mingxi, Yuanyuan, Xiaoying, Yee Teck, Xinghua, Yanan, Yuan Xin, Shimin, Chen Jin, Ricksen, Ci Ji, Li You, Rus, Zibo, Saranya, Sin Yi, Ranjit, Xu Yue, Wei Juan, Carmen, Wenwen, Artem, Huijuan, Yingjian, Xiaodan, Ryo, Durgarao, for the support and companion through the hard times.

Special thanks must go to Dr. Zhang Xinhuai and other High Performance Computing specialists in Computer Centre, NUS to keep such efficient supercomputers customized and well maintained, also Mr. Alan Davis for his help in sustaining computational facilities in MBI and CBIS.

Last but not the least, I thank with love to my wife Jiang Shu, and my sons Jingxin, Chenxin, and my parents for their love, understanding and forgiveness.

October 14, 2014

Contents

Summary

DNA, the most fundamental building block of life, is a long linear polymer that stores the genetic codes for all living organisms. The most common physiological secondary structure of DNA is the right-handed anti-parallel double helical structure, so-called B-form, which is stabilized by Watson-Crick basepair interactions. However, it can dynamically deform to other variations to perform its multiple cellular functions. For example, melted DNA bubble forms during transcription, and the left-handed helical Z-DNA exists *in vivo*, playing a role in transcription regulations.

Structural transitions of DNA can be induced by many factors, including mechanical and topological constraints such as tension, torsion, bending, supercoiling, *etc.* Many DNA binding proteins including histones, nucleoid-associated proteins and various transcription factors introduce sharp DNA bending, while the structural stability of DNA under bending constraint has not been extensively studied. The importance of this study is highlighted by several recent experimental evidences about DNA anomalous elasticity when it is sharply bent. Motivated by the lacking of understanding about such structural stability of sharply bent DNA and its potential physiological importance, my Ph.D research has been mainly devoted to study DNA structural defect formations under sharp bending condition using full-atom molecular dynamics simulations. Using a self-developed novel simulation method, that gradually changes the strain of DNA while recording its resulting stress; I was able to obtain near equilibrium information regarding the micromechanical properties of DNA during bending.

In my studies, a 20 basepair DNA fragment was controlled to bent by external compressional forces. We found that sharp bending could excite flexible defects that consist of $1-3$ disrupted basepairs with an effective persistence length of

$\sim$ 15 nm. Consequently, these flexible defects lead to the formation of large local bends at defected sites and reduce the force to maintain DNA bending. For DNA containing pre-existing nicks that mimics the conditions in some experiments, it bends like normal *B*-DNA under weak constraints, but upon further bending it is easier to unstack at nicked sites in a temperature depednent manner. Overall, our results provide direct insights to the structural stability of DNA under sharp bending condition, and contribute to mechanistic understandings of several recent experiments been debated.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Deoxyribonucleic acid (DNA) is the genetic material adopted by all known living organism and many viruses on earth, which stores and processes genetic information. This macromolecule commonly exists as a right-handed double helix, composed of two sugar-phosphate alternating backbones running opposite with each other and complimentary basepair linkage buried in between. The four distinct bases (`A`, `T`, `G`, `C`) code the genetic instructions, in particular, protein expression using triplet codon through transcription and translation. While, the `A`=`T`, `G`≡`C` pairing ensure DNA complimentary replication and genetic inheritance.

Under most common physiological conditions, double-stranded DNA (*ds*DNA) adapts into this "native" *B*-form, and such homogeneous polymer is known to be quite rigid, which is well described by worm-like chain (WLC) model with an experimentally determined persistence length $A \approx 50$ nm. About its physical dimensions, in bacterial such as *E. coli*, *ds*DNA often has several mega basepairs (bp), corresponding to a contour length of ~ 1 millimeter (mm). It is about $1,000$ folds longer than the linear dimension of the bacteria. Similarly, in eukaryotic cells such as human cells, the DNA has a linear dimension of ~ 2 m, which is $100,000$ folds longer than the cell dimension. The tight compaction of *B*-DNA are primarily gained from structural protein bindings, mainly histones in eukaryotic cells and nucleoid associated proteins (NAPs) in prokaryotic cells. These associations usually create extreme constrains on DNA in very short length

scales. In addition, from functional point of view, DNA is very actively processed and dynamically packed, unpacked all the time. Each gene inside is frequently accessed up to 100s times per hour through transcription.

Evidently, in order to store vast amount of information about complete and complicated instructions on biological functions, a long DNA with numerous basepairs is required. While, to maintain the readability of such sequential genetic information, the secondary helical structure of DNA is necessary. Because, we can easily trace the sequence of *ds*DNA from first to last bp in solution, while hard to do that on strand separated two single-stranded DNA (*ss*DNA), due to low rigidities and random self-interactions. On the other hand, these DNA molecules must be highly packed into the tiny volume of each living cell, at the same time, sustain its frequent accessibilities for other molecular machineries, such as DNA polymerases. Therefore, biophysically, it is very challenging for organisms to preserve such long, rigid structures, while keeping them tightly, dynamically organized into small space.

In order to understand this contradictory more quantitatively, the DNA molecules are modelled as homogeneous thin elastic rod, by well-established polymer theory, 50-nm persistence length WLC model. It has been proven successful in describing medium to large-scale experimental results over and over again, since the remarkable single-molecule force extension fitting by Marko *at el.* in 1995 [1]. But if we assume the homogeneities of DNA molecule here, it will cost huge amount of energy to keep DNA tightly packed. A notable example is DNA wrapped around histones to form the fundamental DNA organization units, nucleosomes, in eukaryotic cells. On each nucleosome, a piece of 147 bp DNA is wrapped by ~ 1.7 turns with a left-handed chirality. This roughly corresponds to the bending of 94 bp (~ 31 nm) DNA into a planar loop, which results in an energy cost of $\sim 31\ k_{\mathrm{B}}T$, based on WLC predictions.

Recent years, increasing evidences have suggested that DNA may adopt certain mechanisms to disrupt helical structures under constrained conditions, through which it can reduce its local rigidities. As a result, it induces homogeneity breakages on DNA to adapt such constraints. So, it is questionable that whether we can directly extrapolate homogeneous WLC model to such small

length scales, under extremely constrained circumstances.

The secondary helical structures of DNA are stabilized by non-covalent interactions, including hydrogen bonding based complimentary base-pairing, and electron correlation originated neighbouring base-stacking. This provides the structural basis for aforementioned homogeneity breakages. The disruptions of these interactions cause structural transitions in DNA molecule, which has been observed and studied long before, but usually in large scales with more server forms, such as DNA melting. The phenomenon of DNA melting, during which the *B*-form DNA duplex transits into two separated *ss*DNA, can be induced by changes in various factors such as increasing temperature, decreasing salt concentration, or increasing forces on DNA. It generally follows nearest-neighbour (NN) model, whose sequence dependent entropy and enthalpy changes per base-pair have been parameterized in various ways and further unified by SantaLucia *et al.* [2]. Biologically, local and transient homogeneity breakages are more relevant to cellular functions, such as the energy cost reductions during DNA packaging, where these interactions are under dynamic disruptions and restorations.

Regarding the external constraints, it has become clear that DNA packaging in cells creates many topological and mechanical constraints to DNA. Packaging itself requires tight DNA folding or wrapping locally, creating high curvature. Local DNA wrapping also often leads to DNA backbone rotation, resulting in torsion stress to DNA. DNA is known anchored to cell walls in bacterial and to the nuclear membrane in eukaryotic cells; therefore, DNA packaging also builds a passive tension along DNA. Various machineries also actively generate tension, during transcription activities by RNA polymerases and DNA replication by DNA polymerases. In addition, recent experiments have suggest that tension produced by actomyosin cytoskeleton contraction is also propagated through the actin network to chromosomes directly. Moreover, the cellular environments is very crowded; up to $\sim$ 400 mg/ml macromolecules collide with each other in nucleus, enhancing the mechanical constraints exerted on DNA. All these evidences highlight the importance to understand DNA structures and stabilities under various physical constraints. Particularly, we focus on sharp bending in this thesis, because the most sensitive stress-strain response for DNA molecules is

bending. Together with the relatively deformable non-covalent interactions, such mechanical constraints may induce local disruptions on DNA secondary structures. In spite of its potential physiological importance, this field has remained largely unexplored, due to the difficulties of applying such constraints to small scale DNA and making accurate measurements of their mechanical responses *in vitro*.

In 2004, ground-breaking results from Cloutier and Widom challenged the canonical DNA polymer model by reporting an anomalous bending elasticity of 94 bp minicircle through DNA looping assays [3]. They found the observed looping probabilities are $1,000$ folds larger than that expected from the 50-nm persistence length WLC model, which corresponds to a reduction in effective bending energy of $\sim 7\ k_{\mathrm{B}}T$. Although the validity of this experimental observation was debated, this study has motivated many laboratories devoting their attentions on DNA behaviours at small length scales under bending constraints. Using different experimental approaches, several experiments have reported results that are consistent with anomalous DNA bendability under high curvature constraint. Due to the experimental challenges, almost all such experiments have been debated on the validity of their experimental designs or interpretations of the resulting data. Up to date, the mechanics of DNA under physiological level of bending (*i.e.*, around 0.2 rad·nm^{-1} based on DNA curvature on a nucleosome) remains an open question.

Several theoretical models have been proposed to account for the anomalous bendability of DNA under sharp bending conditions. Most of them are based on excitation of one or a few flexible hinge or bends inside a sharply bent DNA [4, 5]. Although it costs certain energy to excite such local DNA defects, the presence of such defects may significantly reduce the level of DNA bending, morphologically. Certainly, from a theoretical perspective, such defect excitation is anticipated to occur when the bending is sharp enough. In these theoretical discussions, there have been debates on the possible types of defects that may be excited by sharply bending a DNA. Candidates for flexible hinges have been proposed to be locally melted DNA basepairs and/or a specific type of basepair DNA called *S*-DNA which is 68% longer while 5-folds more flexible than *B*-DNA. A candidate for

intrinsic kinks without basepair breaking has been proposed by Crick and Klug in 1975 [6]. Up to date, no experimental approaches allow people to directly observe the specific type of defects excited in a sharply bent DNA.

Motivated by the above experiments and theoretical discussions, I have been devoted to look into the details of DNA defects excited when it is sharply bent using full-atom molecular dynamics (MD) simulations in explicit water. This approach allows me to obtain dynamic details at the level of atoms inside each bases. Using a spring connecting to the two ends of a short DNA fragments, I was able to introduce sharp DNA bending, and investigate the structural changes as bending increases. I confirmed that DNA homogeneity breakages are indeed generated under sharp bending, mainly in form of basepair disruptions. I also quantified the energy and force responses during bending for DNA with and without defects. I further focused on the impacts of these defects on the overall mechanical properties of the DNA fragments, which could then be related back to the recently proposed theoretical polymer model that permits defect excitation. In addition, effects of pre-existing nicks on DNA elastic responses under bending constraints was also examined. My results show that such nicks can significantly reduce the energy cost to unstack the DNA at the nicked site during DNA bending, which may affect current interpretations of some experimental results.

Overall, besides explicitly pointing out our motivations and briefly summarizing our studies, this overview layouts the contents for Chapter 1, which are necessary backgrounds and core concepts for better and clearer illustrations of my research.

1.2 DNA atomic structures

(a) Double helix (b) Atomic structre of single strand

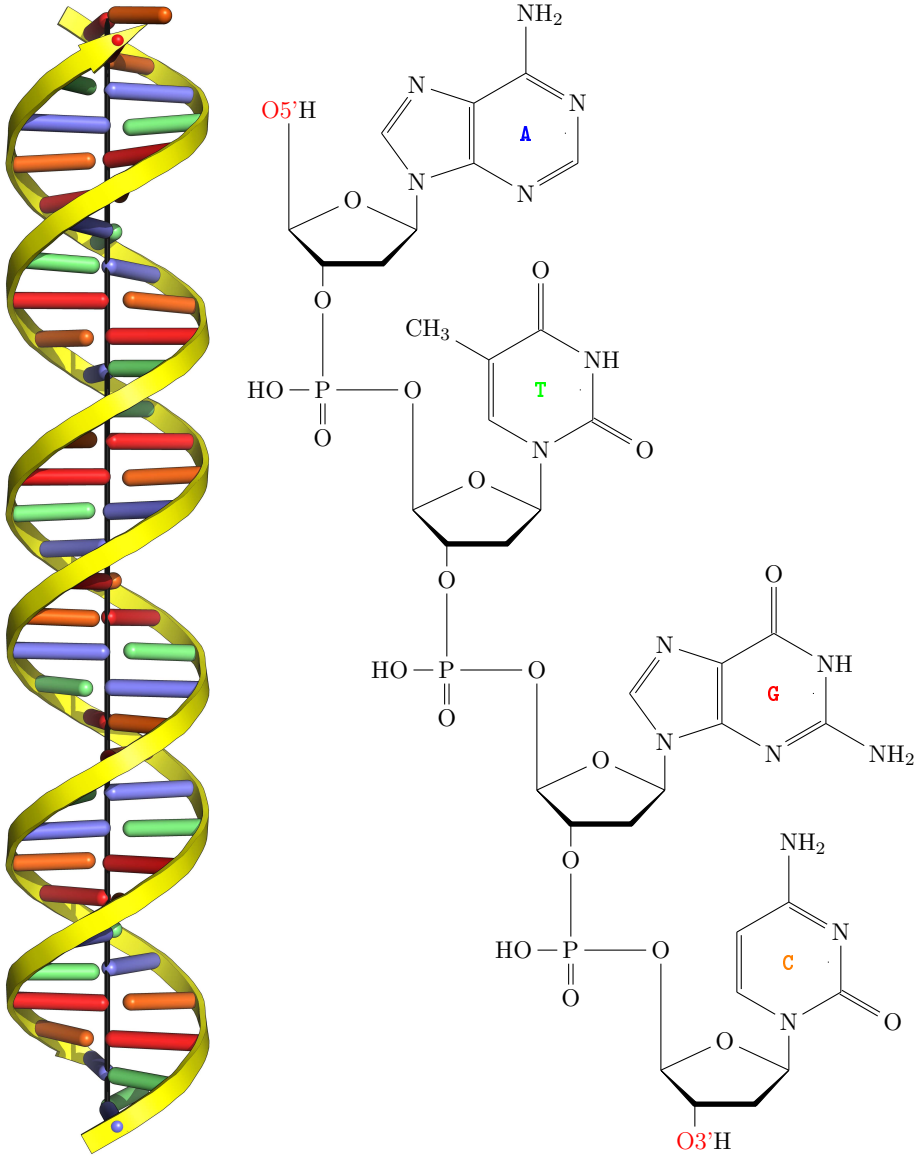


Figure 1.1: DNA atomic structures (a) The diagrammatic representation of *B*-DNA. Its overall shape is a right-handed periodic double helix, with two backbones (coloured in yellow) running in antiparallel directions, and four types of bases (coloured in A, T, G, C) adhering inside. (b) The detailed covalent structures for a single strand. It is composed of building blocks, nucleotides; each contains a nitrogenous base, a five-carbon sugar and a phosphate group. They link one after another by forming the phosphaester bonds between current deoxyribose to preceding phosphate, and finally leaving hydrolyzed O5' and O3' ends, which defines the unique backbone direction, denoted by $5'$ to $3'$.

In 1953, Watson and Crick proposed the detailed structures of DNA, inspired by Franklin and Gosling's fibre diagram obtained using X-ray diffractions [7].

Their diagrammatic presentation of DNA shows double helical chains coiling around a common axis, and interlinked by horizontal rods [Figure 1.1(a)]. This is the most influential milestone of human beings in understanding the secrets of life, because it not only discovered the structures of genetic material adopted by all known living organisms, but also nearly immediately revealed the genetic information coding, copying and passing mechanisms. Since then, DNA draws lots of general attentions and scientific interests, due to its biological importance, as well as unique physical properties.

The most common physiological form of DNA (B-form) is a right-handed double-stranded helical structure, with helical pitch of 34 Å, helical repeat of 10.5 bp and helical diameter of 20 Å. Alternating phosphate group and $5'$ to $3'$ β-D-2-deoxyribose covalently link with each other to form individual chains of DNA through phosphodiester bonds, while the two chains run anti-parallel with each other. The genetic codes, which encrypt the information about life, are buried between chains. There are four types of codes for DNA, whose chemical nature are purine (adenine, `A`, guanine, `G`) and pyrimidine (cytosine, `C`, thymine, `T`) bases [Figure 1.1(b)]. The bases covalently attached to the chain by C1' to N1 linkage for pyrimidine and C1' to N9 linkage for purine.

Besides these durable covalent bonds, dispersed variations of electromagnetic interactions, in the forms of base-pairings and base-stackings as shown in Figure 1.2 , also significantly contribute in DNA molecule formation. More importantly, they are closely relate to DNA elasticities and functionalities, because of their dynamically constructive and destructive nature. The base-pairings (`A`=`T` and `G`≡`C`) cross link DNA two complementary chains. This hydrogen bonding originated specific pairing mechanism has a suitable binding strength for both stabilities and accessibility, and ensures DNA to always have an additional copy of its genetic information as backup. In addition to these horizontal interactions, base-stackings stabilize DNA vertically, and organize its genetic codes in order by adhering one with next. These aromaticity and hydrophobicity based interactions also give rise of the DNA rigidities, and dominate DNA physical morphology.

Taking account of all covalent bondings and non-covalent interactions in DNA

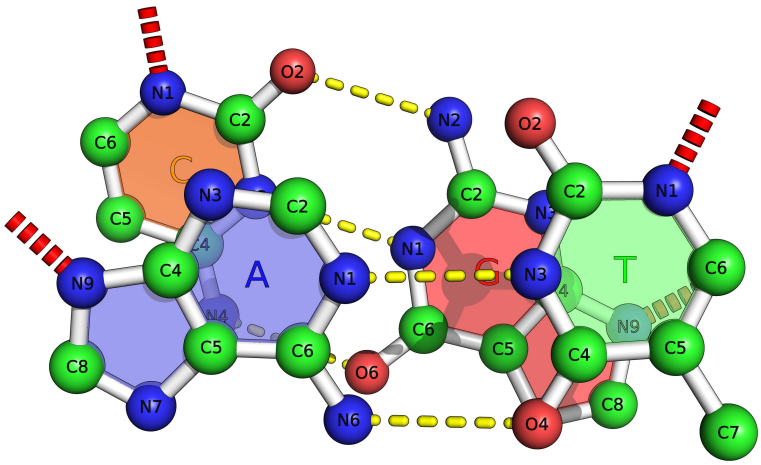


Figure 1.2: Atomic structures of DNA core demonstrated by four bases, which pair and stack with one another in both horizontal and vertical directions through non-covalent interactions. The hydrogen bonding is shown using yellow dotted lines, and basepair stacking is generally strong at overlapping regions. Note that the red dotted lines at peripheral are linkages to the two backbones.

molecule, most of the time DNA resumes a definitive secondary structure in cell, the native B-form as proposed by Watson and Crick [7]. In nature, DNA also adapts itself into other forms under different environmental factors and mechanical constraints. Such as, left-handed Z-form discovered in 1979 by single crystal X-ray diffraction [8], which is transient localized conformation that can be stabilized by negative supercoiling during transcription, and specifically recognized by some protein involved in viral pathogenicity [9]. More compact A-form, which closely resembles common B-form, found in dehydrated conditions, during early DNA crystallographic experiments. More extremely, S-form was identified under DNA overstretching in 2010, which is characterized by non-hysteretic transition and negative entropy [10, 11]. They are all DNA local energy minimal states, which are stabilized under particular chemical conditions and/or physical constraints. As we can see, the secondary structure of DNA is very adaptive to external stresses, and these structure responses directly link to its biological functionalities. Therefore, it is critical to treat DNA as a dynamic molecule, which is covalently stable, but non-covalently deformable under common physiological related physical and chemical alternations.

1.3 DNA polymer models

In this section, we will scratch the surfaces of the well-established and widely accepted polymer model. There are multiple models to describe different polymers, and for DNA, we need a model with significant considerations on polymer bending. The rigidities dominant the DNA spatial arrangements, and twisting, stretching are ignored for simplicity, although they are also important factors in DNA functionalities. In the simplest WLC polymer model that only considers DNA bending energy, the DNA is approximated to be an infinitely thin rod. For a given DNA conformation, its conformational energy is described by,

$$\beta\mathcal{H} = \frac{A}{2}\int_{s=0}^{L}\left(\frac{\partial\hat{t}(s)}{\partial s}\right)^2 ds = \frac{A}{2}\int_{s=0}^{L}\left(\frac{\partial^2\boldsymbol{r}(s)}{\partial s^2}\right)^2 ds \tag{1.1}$$

, where $\hat{t}(s)$ is the unit tangent vector, $\boldsymbol{r}(s)$ is the position vector at the contour location s, $\beta = (k_{\mathrm{B}}T)^{-1}$ rescales the Hamiltonian into unit of $k_{\mathrm{B}}T$, and the characteristic parameter A has a dimension of length and is called the bending persistence length. For the B-form dsDNA, the value of A has been measured by single-DNA stretching experiments to be around $A = 53.4 \pm 2.3$ nm under normal experimental solution conditions [12].

For a short segment of length l bent into an arc conformation, the energy becomes,

$$\beta\mathcal{H} = \frac{A}{2}\frac{l}{R^2} = \frac{A}{2}\frac{\theta^2}{l} \tag{1.2}$$

, where R is the radius of the arc curvature and θ is the bending angle between two DNA ends. This formula indicates that 1 $k_{\mathrm{B}}T$ energy approximately can excite a bending of ~ 1 rad for the DNA with length of A. At a length scale $l \ll A$, the DNA segment can be assumed to be straight due the high energy cost for inducing bending.

Based on this, previous continuous WLC polymer model can be discretized by considering the DNA molecule as a chain of rigid short segments, each with a segment length l and an orientational unit vector $\hat{t}_i$, for $i = 1, 2, \cdots, N$. Then,

$$\beta\mathcal{H} = \frac{1}{2}\frac{A}{l}\sum_{i=1}^{N-1}\left(\hat{t}_{i+1} - \hat{t}_i\right)^2 = -\frac{A}{l}\sum_{i=1}^{N-1}\hat{t}_i \cdot \hat{t}_{i+1}. \tag{1.3}$$

The equivalence between the continuous and discretized model is achieved at the limit $l \to 0$, $N \to \infty$, while $lN = L$. In real applications, discretized model is frequently used given $l \ll A$; and following discussions are based on the discretized model.

The polymer of N segment contains $(N-1)$ vertices, each carrying a bending angle of θ_i and the bends at the vertices are independent from each other. Therefore, by working on the partition function of bending for one vertex $\mathcal{Z}_i$, one can obtain the total partition function of the chain by $\mathcal{Z}_N = \mathcal{Z}_i^{N-1}$, where,

$$\mathcal{Z}_i = \int d\Omega_i \exp\left(a\cos\theta_i\right) = 4\pi\frac{\sinh a}{a}. \tag{1.4}$$

Note that the dimensionless vertex bending rigidity $a = \frac{A}{l}$. So, the nearest neighbour bending correlation is,

$$\left\langle \hat{t}_i \cdot \hat{t}_{i+1} \right\rangle = \left\langle \cos\theta_i \right\rangle = \frac{\partial \ln \mathcal{Z}_i}{\partial a} = \coth a - \frac{1}{a} \tag{1.5}$$

, which is the Langevin function, $\mathcal{L}(a)$. And this results in a nice relation between bending correlation function and vertex bending rigidity, when $a \gg 1$,

$$\left\langle \hat{t}_i \cdot \hat{t}_{i+\Delta} \right\rangle = \left(\mathcal{L}(a)\right)^{\Delta} \approx \exp\left(-\frac{\Delta}{a}\right) \tag{1.6}$$

, where Δ is the discretized separation in unit of segment length. Geometrically, it indicates that only the projection of the current vector in the direction of the preceding one propagates to the next, and the correlations decay exponentially. Then, above discretized correlation function can be extended to the continuous case. For any two positions on the continuous homogeneous rod, their tangent vector correlation function is expressed as,

$$\left\langle \hat{t}(s) \cdot \hat{t}(s') \right\rangle = \exp\left(-\frac{|s'-s|}{A}\right) \tag{1.7}$$

, where $a \times l = A$, contour separation $\Delta \times l = |s'-s|$, and s and s' denote two contour locations along DNA.

Now, in order to check the validities of WLC model, we are going to extrap-

olate the theory to compare against experimental measurable quantities. The macroscopic coil sizes of DNA, represented using its mean-square end-to-end distances, can be expressed as,

$$\begin{aligned}\left\langle \vec{D}^2 \right\rangle &= \left\langle \int_{s=0}^{L} \hat{t}(s)\, ds \cdot \int_{s'=0}^{L} \hat{t}\big(s'\big)\, ds' \right\rangle \\ &= \int_{s=0}^{L} ds \int_{s'=0}^{L} \left\langle \hat{t}(s) \cdot \hat{t}\big(s'\big) \right\rangle ds' \\ &= \int_{s=0}^{L} ds \int_{s'=0}^{L} \exp\left(-\frac{|s'-s|}{A}\right) ds' \\ &= 2AL\left(1-\frac{A}{L}\left(1-\exp\left(-\frac{L}{A}\right)\right)\right)\end{aligned} \tag{1.8}$$

, where $\vec{D}$ is the end-to-end distance. Because the bending persistence length is also the bending correlation length, which implied that for large DNA polymer with $L \gg A$, the polymer conformation is a random coil due to the quick loss of the bending correlations at large length scale. So, at the long chain limit the $\left\langle \vec{D}^2 \right\rangle \approx 2AL$, which is identical to the case of random walk with a step size of twice the persistence length. In other words, due to fact that polymer intrinsic elastic responses is shielded by same averaged total correlations (Appendix A), the polymer size is indistinguishable from that predicted by freely joint chain (FJC) model,

$$\left\langle \vec{D}^2 \right\rangle = b^2 \sum_{i=1}^{n} \left\langle \hat{t}_i^2 \right\rangle = nb^2 = bL \tag{1.9}$$

, where the DNA is re-discretized into n segments, each with length of Kuhn length $b = 2A$. As a result, although based on DNA random coil dimensions, persistence length A can be accurately determined, these large-scale non-constrained measurements (*i.e.*, such as radius of gyration) cannot justify the correctness of WLC model.

Later, both the bending correlation function $\left\langle \hat{t}(s) \cdot \hat{t}(s') \right\rangle$ and $\left\langle \vec{D}^2 \right\rangle$ have been obtained by directly imaging DNA conformations deposited on 2D mica surface using atomic force microscopy (AFM), which have supported that DNA can be described by the WLC model with $A \approx 50 - 100$ nm. However, such imaging experiments require ensemble averages, which are subject to perturbations of sample preparations. In addition, the quantification of A is also under a critical assumption that DNA conformations have reached equilibrium in 2D, which is

difficult to be tested.

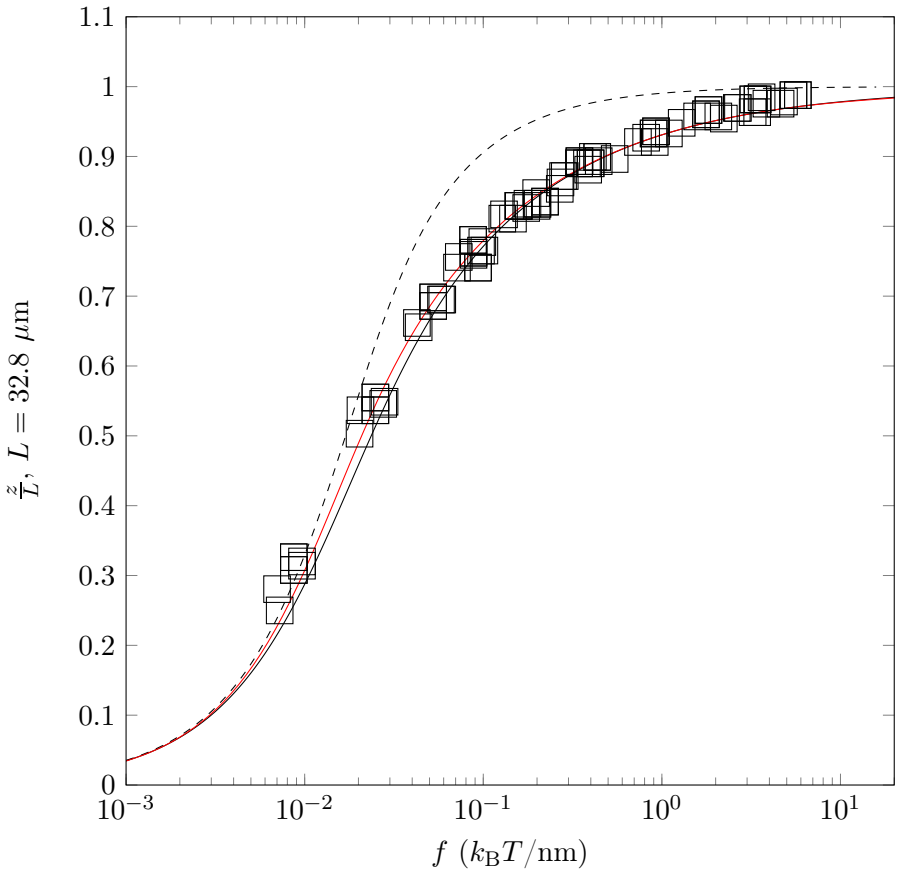


Figure 1.3: The DNA theoretical force extension relations predicted using FJC model (dashed black line), with $b = 2A = 106$ nm against WLC model, based on Equation 1.12 (solid black line), with $A = 53$ nm. It shows that the relative lengths $\frac{z}{L}$ from both models overlap at beginning, while deviate from each other quickly even under small tensions, as we stated in main text. The discrepancy is due to those chain fluctuations, whose wavelength smaller than b, involving in WLC, but lacking in FJC model, shorten the extension under tensions. The square data is the experimental data by Smith *at el.* [13], obtained through stretching λ-DNA dimer under magnetic field and flow in low ionic strength. It follows WLC but differs from FJC predictions in most biologically important force ranges. The "exact' WLC force extension curve is plotted as well (solid red line), which is numerically obtained using transfer matrix approach [4].

A more direct test for whether DNA can be described by the WLC model should be done under constraints, in solution and at a single-DNA level. Recent development of single-molecule manipulation technology has made it possible to stretch a DNA by applying a tension f to its two ends and measuring the resulting extension $z(f)$, which is the mean end-to-end distance projected in the direction of applied force, at a nanometer resolution. This measured force-extension curve $z(f)$ can be compared with the prediction by WLC model, to test the validity of

the WLC model and determine the accurate value of A, as well.

Under high force limit ($f \gg \frac{k_B T}{A}$) where the tangent vector $\hat{t}(s)$ is nearly aligned along the force direction, the predicted $z(f)$ has been worked out by Marko *at el.* in 1995 [1],

$$\frac{z(f)}{L} = 1 - \sqrt{\frac{1}{4\beta A f}} \tag{1.10}$$

, which can also be conveniently expressed by its inverse function, $\beta f A = \frac{1}{4}\left(1 - \frac{z}{L}\right)^{-2}$. At low force limit, $z(f)$ should converge to the force response of a FJC model with a segment length of $2A$, which follows an entropic spring behaviour as,

$$\frac{z}{L} = \frac{2}{3}\beta f A. \tag{1.11}$$

Then, a direct interpolation to connect the extension responses of a WLC polymer at low and high force limit as,

$$\beta f A = \frac{z}{L} + \frac{1}{4\left(1 - \frac{z}{L}\right)^2} - \frac{1}{4} \tag{1.12}$$

, which contains a single free parameter A. This formula was able to fit the experimental force-extension curve of a $97,004$ bp dimeric λ-DNA obtained at 10 mM Na^+ (Figure 1.3), which validates the description of DNA by the WLC polymer model. In addition, it determines the value of A to be around 53.4 ± 2.3 nm under normal experimental solution conditions with great accuracy [12]. This value of A indicates DNA is relative stiff, and only a small force (*i.e.*, $f \approx \frac{k_B T}{A} \sim 0.08$ pN) is required to extend it away from its entropic random coiled conformation. Therefore, the predicted $z(f)$ by the FJC model deviates from the measured $z(f)$ at the large force region (*i.e.*, when $f > \frac{k_B T}{A}$).

It is known from many single-DNA stretching experiments that the inextensible WLC model with a constant contour can describe DNA at forces up to 20 pN. Above which till ~ 60 pN the WLC model requires an modification to take into account of contour stretching rigidity [14]. At forces greater than 65 pN, *ds*DNA becomes unstable and DNA structural transitions may occur [15, 16, 10, 11].

1.4 DNA persistence length determination

In this section we introduce various methods that have been used to verify DNA WLC behaviours through estimating the characteristic parameter A with more details. As soon as it has been confirmed that DNA is well described by WLC model, many efforts have been devoted to determine the persistence length A based on various WLC predictions, under different bending regimes. This semi-flexible model in general describes the distinctive behaviours of DNA under different length scales. When $L \gg A$, it is highly influenced by thermal fluctuations, and randomly coiled. While $L \ll A$, its intrinsic bending rigidity dominates, and DNA behaves more like a stiff mechanical rod. Pioneers often focus on measuring A at intermediate to large length scales, due to relatively low experimental requirements on manipulations and observations. Later, because of the biological relevance at small length scales, more groups start to tackle A at rod regime, using more sophisticated experimental procedures and detection methods.

1.4.1 Measurements of A at medium to large length scales

In order to determine A, experimentally measured DNA extension *vs.* force relationships have been used to fit the predicted force-extension curves based on the WLC model [12, 1] (*e.g.* the Marko-Siggia formula, the Odijk formula [17]). In such experiments, the overall DNA molecule responses from random coiled to extended states can be accurately resolved through applying external stretching forces. The first single-DNA stretching measurements [13] well fits theoretical predictions and the persistence length was determined to be $A = 53.4 \pm 2.3$ nm [12]. Numerous following single-DNA stretching experiments have determined the value of A for *ds*DNA in the range of $42 - 53$ nm, which varies with different environmental factors, such as salt concentration or temperature [18, 19, 20].

The 50 nm bending persistence length was also confirmed by obtaining ensembles of DNA conformations using AFM imaging experiments. With more knowledge about the chain statistics and more carefully designed depositions, the DNA conformations have been equilibrated on mica surface. Then, several quantities, such as tangent vector correlation function or mean-square end-to-

end distances can be measured to estimate the value of A. As a reminder, $\left\langle \hat{t}(s) \cdot \hat{t}(s') \right\rangle_{\mathrm{2D}} = \exp\left(-\frac{l_{\mathrm{s}}}{2A}\right)$ and $\left\langle \vec{D}^2 \right\rangle_{\mathrm{2D}} = 4AL$ have been applied, where $l_{\mathrm{s}} = |s' - s|$ is the counter separation. Note that the correlation length, which has been altered from A in 3D to $2 \times A$ in 2D. Several such AFM experiments obtained the persistence length of $50 - 54$ nm at intermediate length scales $l_{\mathrm{s}} < 200$ nm, in order to limit the volume exclusion effects on 2D surface [21, 22].

DNA bending stiffness can also be experimentally quantified by measuring chances of its two ends meeting each other during thermal fluctuations. In this approach, the DNA bending responses from random coil to tightly bent states were indirectly reflected through obtaining its probability densities of ring closure. This method is sensitive at the length scales comparable to A, where the looping probabilities maximizes, because bending energy per basepair reduces to constant while more degrees of freedom are allowed as DNA counter length increases. Assuming the two ends of DNA meet in parallel to form a circle, aforementioned looping probability densities $\rho^{\mathrm{E}}(\boldsymbol{R} = \mathbf{0})$ approximates theoretical predations $\rho_{\parallel}(\mathbf{0}) \propto \left(\frac{L}{A}\right)^{-6} \exp\left(-\frac{2\pi^2 A}{L} + 0.514 \times \frac{L}{2A}\right)$ [23], where $\boldsymbol{R}$ is the distance vector between the two ends. So, this relationship between DNA contour length and $\rho^{\mathrm{E}}(\mathbf{0})$ at such length scales can be used to evaluate A.

$\rho^{\mathrm{E}}(\mathbf{0})$ has often been measured through ligation based DNA cyclization experiments. In such experiments, a DNA fragment with short complementary *ss*DNA overhangs at the two ends was required. In a solution of such DNA molecules at a concentration $c = \frac{N}{V}$ (N is the number of molecule and V is the volume), a terminus of a molecule can hybridize with a complementary terminus from the same molecule (*i.e.*, looping) or from another molecule (*i.e.*, dimerization), driven by thermal fluctuation. Theoretically, if hybridization between complementary DNA ends can occur without any conformational constraint, then $\frac{K_{\mathrm{loop}}}{K_{\mathrm{dimer}}} = \frac{\rho^{\mathrm{E}}(\mathbf{0})}{c}$; this results in experimental determination of looping probability density as: $\rho^{\mathrm{E}}(\mathbf{0}) = \frac{K_{\mathrm{loop}}}{K^0_{\mathrm{dimer}}}$. Here $K^0_{\mathrm{dimer}} = \frac{K_{\mathrm{dimer}}}{c}$ denotes the dimerization rate per unit concentration of DNA. The ratio $j = \frac{K_{\mathrm{loop}}}{K^0_{\mathrm{dimer}}}$ is often referred as the "j-factor", which has a dimension of concentration [24, 3, 25]. Often, molar concentration is used in experiments, which leads to an additional factor of Avogadro's number: $j = N_{\mathrm{A}} \cdot j_{\mathrm{M}}$.

However, hybridization between two complementary DNA ends actually imposes certain orientational constraints on the two meeting DNA ends, before the subsequent ligation reactions. For the case of dimerization, the hybridized DNA ends are straight, thus in parallel to each other and twisted to match the B-form conformation (hereafter we call this constraint as twist-matching parallel boundary condition, denoted by Ω, see also Figure 5.7). As a result, only a subset of meeting ends forms stable B-form looping, which leads to an additional factor in the looping probability density measurements: $\rho^{\mathrm{E}}(\mathbf{0}) = \frac{j}{4\pi\times 2\pi}$. For the case of looping, in order to achieve correct theoretical estimations for comparison, the knowledge on how two DNA ends meet also matters. The assumption of the Ω boundary constraints has been implied for medium to large-scale DNA (although might not be explicitly mentioned), where the contribution of twisting energy is negligible against bending energy and entropy. Therefore, the modelled DNA ring-closing probability density under parallel boundary condition (*i.e.*, classical WLC model without consideration of DNA twists) relates to j_{M} by equation:

$$\rho_{\parallel}(\mathbf{0}) = 2\pi\rho_{\Omega}(\mathbf{0}) = \frac{N_{\mathrm{A}}\cdot j_{\mathrm{M}}}{4\pi}. \tag{1.13}$$

Based on such j-factor measurements and theoretical interpretation, the DNA persistence length was determined in the range of $450-550$ Å, over a wide contour length range ($>$ 200 bp) in normal solution conditions [26, 27]. The agreement between these measured values and that from single-DNA stretching experiments validates the Ω boundary condition for looped DNA larger than 200 bp.

In summary, at length scales comparable to A or longer, the DNA bending persistence length has been consistently determined to be around 50 nm by different experimental approaches.

1.4.2 Measurements of A at small length scales

Biologically, DNA is highly packed, well organized and tightly bent at small length scales; for instance, in nucleosome, a stretch of 147 bp ($\sim$ 50 nm) DNA tightly wraps around the histone octamer in a left-handed manner by 1.7 turns. In one turn, around 94 bp ($\sim$ 31 nm) DNA is wrapped nearly into a circle, leading

to a sharply bent DNA conformation that is much more severely bent than that probed by previous single-DNA stretching and AFM imaging experiments. Due to its direct biological relevance, it is important to understand the mechanics of sharply bent DNA. In addition, it is only possible, at small length scales, to induce sharp bending into DNA, where we need to treat DNA as stiff rod. Moreover, through local measurements, the sequence effects on A can be directly quantified. Note that $A \approx 50$ nm measured based on intermediate to large scale responses is an overall averaged rigidity. But due to limitations on accuracy under small length scales and experimental difficulties to manipulate short DNA pieces, only several experiments were attempted so far.

Based on j-factor measurements, in 2004, the DNA looping responses for $94-116$ bp DNA minicircles (*i.e.*, with $9, 10, 11$ helical turns, which leads to zero twisting energy) were reported by Cloutier and Widom [3]. The obtained j-factor at such DNA length are several orders of magnitude higher than the predicted value based on the classic DNA WLC polymer model using parallel boundary condition. To fit the data, a significantly smaller apparent A has to be used under such constraints, questioning the 50-nm persistence length based WLC model in sharply bent state. This measurement was challenged by a later publication, which reported similar j-factor measurements by ligation reactions for similar lengths of DNA, while obtained results that support the 50-nm persistence length based WLC predictions [25]. However, we note that the later experiment was conducted at a lower temperature of 21°C than the former one at 30°C [28]. More recently, a few more cyclization experiments were done based on more direct single-molecule Förster resonance energy transfer (smFRET) assays on similar length of DNA [29], which reported anomalously high DNA looping probabilities similar to that reported by Cloutier and Widom.

These sometimes contradictory results reveal a complex nature of both experimental measurements and DNA elastic responses under such sharp bending conditions. It has drawn lots of attentions from theoreticians since then. A popular model to understand the atypical bending elasticity of sharply bent DNA measured from these looping assays is that the bending energy stored in the DNA is relaxed by exciting a flexible or kinked structural defect, thus inducing

the local homogeneity breakage inside the B-DNA. By paying some energy to excite such a defect, the overall bending energy is reduced through absorption of the bend to the defect location. By tuning the defect excitation energy and the flexibility of the defect, these models were found able to fully explain all the available DNA looping data at small, as well as medium to large length scales [30, 4, 5].

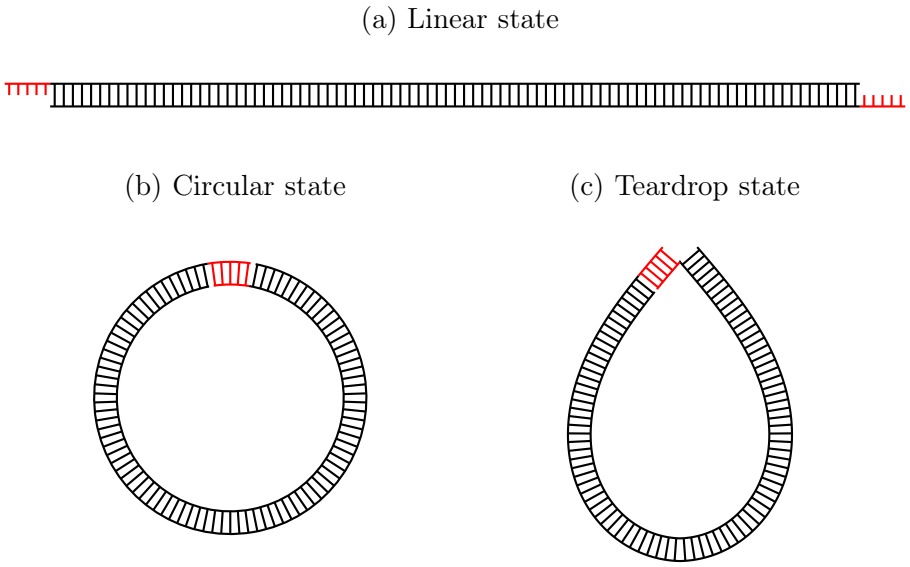


Figure 1.4: Possible effect of nicks on alternating the parallel boundary condition in DNA loop assay of 94 bp minicircle. The j-factor measurement is based on the critical assumption that DNA ends meet in parallel during the fast pre-equilibration between (a) linear and (b) circular states. There are two pre-existing nicks right after the red coloured complementary *ss*DNA overhands. Due to the presence of nicks, some hidden states may become possible, which can be revealed under sharp bending. This possibility will breakdown the parallel boundary condition assumption, thus report an anomalous looping probability once excited. Note that the teardrop state (c) is one example of such states, which is the energy minimum solution of Equation 3.3 under free boundary condition with fully hydrolyzed basepairs.

Another point, which has not been discussed in previous works, is that the interpretations of both the j-factor and the smFRET based measurements of the looping probability are under a critical assumption: upon the two cohesive *ss*DNA overhangs meet, the looped ends of the DNA satisfy the Ω boundary condition. The anomalously high looping probability is a result of comparing the measured looping probability against the predicted looping probability under such Ω boundary condition constraint. Although such Ω boundary condition was confirmed by numerous experiments for DNA $>$ 150 bp, such validation has never

been done for shorter DNA. Assuming the presence of nicks altered the *ss*DNA behaviours when sharply bent, which allows two DNA ends to meet in a non-parallel orientation, as illustrated in Figure 1.4. Then, the ring closure probability density can be several orders higher under another boundary condition, which can also explain the abnormal DNA elasticity observed in these looping experiments.

Several other efforts to probe the DNA bending response under sharp bending conditions used alternative methods. One approach is based on the classical Euler instabilities of stiff rod, which predicts modes of sudden mechanical softening at critical loads, as follow,

$$f_{\mathrm{c}}^{n} = \frac{n^2\pi^2 YI}{L^2} = \frac{n^2\pi^2 A}{\beta L^2} \tag{1.14}$$

, where Y is the Young's modulus, $I = \frac{1}{4}\pi R^4$ is DNA "area moments of inertia" of rod cross section. Shroff *at el.* [31, 32] utilized 10 bases *ss*DNA to sharply bend 25 bp *ds*DNA, achieved by the hybridization of 25 bases complimentary stand to 35 bases circular *ss*DNA loop. The FRET signal, which obtained from Cy5/Cy3 dyes labeled at the two ends of *ds*DNA, was converted to a small tensile force 6 ± 5 pN on the *ss*DNA (*i.e.*, which is also the magnitude of compressional load acting on the bent *ds*DNA), based on their single-molecule stretching calibration experiments. On the other hand, the force estimated from Euler instabilities of rod bending is much larger. As the contour length of *ss*DNA ~ 6.3 nm is a lot shorter than that of *ds*DNA region (~ 8.5 nm), the *ds*DNA must be in a bent conformation, assuming *ds*DNA is fully hybridized. According to Equation 1.14, the first onset of bending occurs at a critical load of $f_{\mathrm{c}} = 24$ pN based on the 50-nm persistence length of WLC model on this 25 bp *ds*DNA. On the contrary, the measured force is much smaller than the predicted critical value, implying again an extremely flexible *ds*DNA.

Similar active bending experiments of a short *ds*DNA fragment using *ss*DNA were also conducted by Qu *et al.* [33, 34], but in the presence of a nick in the middle of this *ds*DNA. The elastic energy of such nicked *ds*DNA has been quantified through its monomer-dimer equilibrium concentrations. Assuming the nick does not affect the geometry and elasticity of the DNA under sharp bending condi-

tion, these experiments also reported anomalously high DNA bending elasticity [33, 34, 35].

Together, these experimental evidences revealed a highly complex picture of DNA elasticity under sharp bending conditions. Because of limitations in measurement accuracies and interferences from unknown factors, it further increases the level of difficulties to study DNA micromechanics at such small scales. Even, the theoretically well-established and experimentally frequently practiced j-factor looping assay is facing challenges, due to presence of nicks. This problem also potentially influence monomer-dimer equilibration in experiments by Zocchi *et al.* [33]. Interpretations for aforementioned experiments rely on some critical assumptions that yet to be validated on DNA minicircles with counter length near A. Although defect excitations have been strongly supported by low compressional force in active bending FRET and *ss*DNA cutting in BAL-31 nuclease digestion assay [36, 37], these two approaches were conducted on much higher degrees of bending (*i.e.*, $\sim$ 64 bp minicircle), instead of biologically relevant 94 bp nucleosome loop. Overall, the bending responses of sharply bent DNA, especially at the level of bending involved in 94 bp minicircle, have remained as an unresolved problem. The previously hidden nick effects under sharp bending constraint are worth investigating as well.

1.5 DNA basepair stabilities

	A	T	G	C
A	$\overrightarrow{\texttt{AA}}$ $\underleftarrow{\texttt{TT}}$	$\overrightarrow{\texttt{AT}}$ $\underleftarrow{\texttt{TA}}$	$\left(\overrightarrow{\texttt{AG}}\ \underleftarrow{\texttt{TC}}\right)$	$\left(\overrightarrow{\texttt{AC}}\ \underleftarrow{\texttt{TG}}\right)$
T	$\overrightarrow{\texttt{TA}}$ $\underleftarrow{\texttt{AT}}$	$\left(\overrightarrow{\texttt{TT}}\ \underleftarrow{\texttt{AA}}\right)$	$\left(\overrightarrow{\texttt{TG}}\ \underleftarrow{\texttt{AC}}\right)$	$\left(\overrightarrow{\texttt{TC}}\ \underleftarrow{\texttt{AG}}\right)$
G	$\overrightarrow{\texttt{GA}}$ $\underleftarrow{\texttt{CT}}$	$\overrightarrow{\texttt{GT}}$ $\underleftarrow{\texttt{CA}}$	$\overrightarrow{\texttt{GG}}$ $\underleftarrow{\texttt{CC}}$	$\overrightarrow{\texttt{GC}}$ $\underleftarrow{\texttt{CG}}$
C	$\overrightarrow{\texttt{CA}}$ $\underleftarrow{\texttt{GT}}$	$\overrightarrow{\texttt{CT}}$ $\underleftarrow{\texttt{GA}}$	$\overrightarrow{\texttt{CG}}$ $\underleftarrow{\texttt{GC}}$	$\left(\overrightarrow{\texttt{CC}}\ \underleftarrow{\texttt{GG}}\right)$

Table 1.1: Ten unique NN parameters out of sixteen possible combinations for DNA. The doublet interactions are defined using neighbouring codes, and DNA have four types, as `A`, `T`, `G`, `C`. Thus, there are total of 16 possible doublet arrangements. However, because of DNA antiparallel double chain structure, one paired pattern without symmetry is the same as another pattern, when reads from $5'$ to $3'$ in complimentary strand. Symmetric paired patterns are identical in both strands, such as `AT/TA`, and there are totally four of them. As a result, NN steps contain $4 + \frac{4\times4-4}{2} = 10$ unique stacked pairs, as shown above. The white entries are symmetric stacked pairs, while, the coloured entries indicate the rest without symmetry. Note that those in same colour are identical. The arrows point out strand directions, and parentheses label the eliminated notations.

DNA functionalities, such as gene storages or processes, critically rely on non-bonding stabilities of the DNA duplex, which are determined by the two main components, horizontal base-pairings and vertical base-stackings [38, 39]. As we proposed, the dynamic disruptions of such non-covalent interactions are the key to understand the micromechanics and functionalities of DNA in biologic systems. Although these stabilities under physiological related constraints have not been investigated yet, their thermodynamics nature have already been extensively explored in an extreme form, which is DNA melting.

The sequence dependent basepair stabilities can be formulated in general as below [40],

$$\theta = \sum_{k=1}^{k_{\text{cut}}} \theta^{(k)} \tag{1.15}$$

, where $\theta^{(1)}$ is the singlet interactions (*i.e.*, such as base-pairings, A=T, G≡C), $\theta^{(2)}$ is the doublet interactions [*i.e.*, such as base-stackings, and this is well known as NN model] and so on. By using the thermal properties of singlet interactions alone, it fails to explain the different melting profiles of DNA with same GC content. In contrast, NN model is able to briefly predict the helix-coil transitions of DNA using only eight invariants [41].

For NN model, the four bases (A, T, G, C) define 16 possible NN steps, among which, there are ten unique stacked pairs, as AA/TT, AT/TA, TA/AT, GA/CT, GT/CA, GG/CC, GC/CG, CA/GT, CT/GA and CG/GC (Table 1.1). Their total-disruption thermal properties (enthalpy ΔH°, and entropy ΔS°) have been extensively measured using UV absorption at 268 nm through melting process and unified NN basepair parameters have been summarized by SantaLucia in 1998 [2], which are listed in the second and third column of Table 1.2.

Note that the NN parameters in the table were measured at 1 M NaCl. The melting process is highly sensitive to the changes of salt, because of the entropy under particular ionic strength, ΔS, have a strong dependence on salt concentrations. However, the salt dependency of ΔS in physiological range can be simply expressed as,

$$\Delta S = \Delta S^\circ + 0.368 \times \ln [\mathrm{Na}^+]. \tag{1.16}$$

And so, as an example, at room temperature (298.15 K) and 150 mM NaCl, the basepair Gibbs free energy changes $\Delta \mathcal{G}^{0.15\mathrm{M}}_{25^\circ\mathrm{C}}$ of those 10 NN steps can be calculated, as listed in the fourth column above.

Practically, based on the NN model the thermodynamical properties of arbitrary sequence DNA can be determined, such as melting temperature T_m or melting profile $-\frac{d\Theta(T)}{dT}$. Its T_m is a linear combination of T_{ij},

$$T_\mathrm{m} = \sum_{ij} X_{ij} \left(T^\circ_{ij} + \frac{dT_{ij}}{d \log_{10} [\mathrm{Na}^+]} \cdot \log_{10} [\mathrm{Na}^+] \right) \tag{1.17}$$

, where X_{ij} is the fraction of particular neighbour pair [41, 42]. While, the melting profile is obtained by adopting Poland–Scheraga model using forward and

NN step	ΔH° kcal/mol	ΔS° cal/K·mol	$\Delta\mathcal{G}^{0.15\mathrm{M}}_{25^\circ\mathrm{C}}$ $k_\mathrm{B}T$	T°_{ij} °C	$\frac{dT_{ij}}{d\log_{10}[\mathrm{Na}^+]}$ °C/M
AA/TT	-7.9	-22.2	-1.81	89.08	19.78
AT/TA	-7.2	-20.4	-1.53	81.85	21.00
TA/AT	-7.2	-21.3	-1.08	86.72	20.11
CA/GT	-8.5	-22.7	-2.57	103.18	17.10
GT/CA	-8.4	-22.4	-2.55	107.96	16.21
CT/GA	-7.8	-21.0	-2.24	104.43	16.87
GA/CT	-8.2	-22.2	-2.32	99.49	17.76
CG/GC	-10.6	-27.2	-3.85	124.54	13.20
GC/CG	-9.8	-24.4	-3.91	124.61	13.20
GG/CC	-8.0	-19.9	-3.13	118.49	14.18
Init. w/term G·C	0.1	-2.8			
Init. w/term A·T	2.3	4.1			
Symm. correction	0	-1.4			

Table 1.2: Unified nearest-neighbour parameters for DNA melting. The second and third columns are enthalpy and entropy for stacked basepair under complete disruptions, which were unified by SantaLucia *et al.* [2]. The bottom gives the correction terms for melting initiation from either G·C or A·T pairing, and for symmetry on self-complementary sequences. The fourth column presents the Gibbs free energy differences, $\Delta\mathcal{G}$, under 0.15 M NaCl and 25°C physiological conditions as an example. Practically, the melting temperature for particular DNA can be estimated using a linear combination of Kelvin temperatures, as $T_m = \sum_{ij} X_{ij} T_{ij}$, where X_{ij} is particular NN fraction. And T°_{ij} column shows those obtained under 1 M NaCl, with their corresponding linear temperature-salt relations listed in last column.

backward recursion relations [43, 44]. The probability of basepair α in bounding state is,

$$p(\alpha) = \frac{\mathcal{Z}_\mathrm{f}(\alpha)\,\mathcal{Z}_\mathrm{b}(\alpha)}{\mathcal{Z}} \tag{1.18}$$

, and as a result, the mean fraction of helical region is,

$$\Theta = \frac{1}{N}\sum_{\alpha=1}^{N} p(\alpha)\,. \tag{1.19}$$

According to Figure 1.5, using *pBR322* plasmid DNA as an example, the predictions from NN model fit the experimental measurements quite well, which indicates its successes in describing these large-scale DNA thermal behaviours.

More interestingly, these basepair stability parameters have been applied to

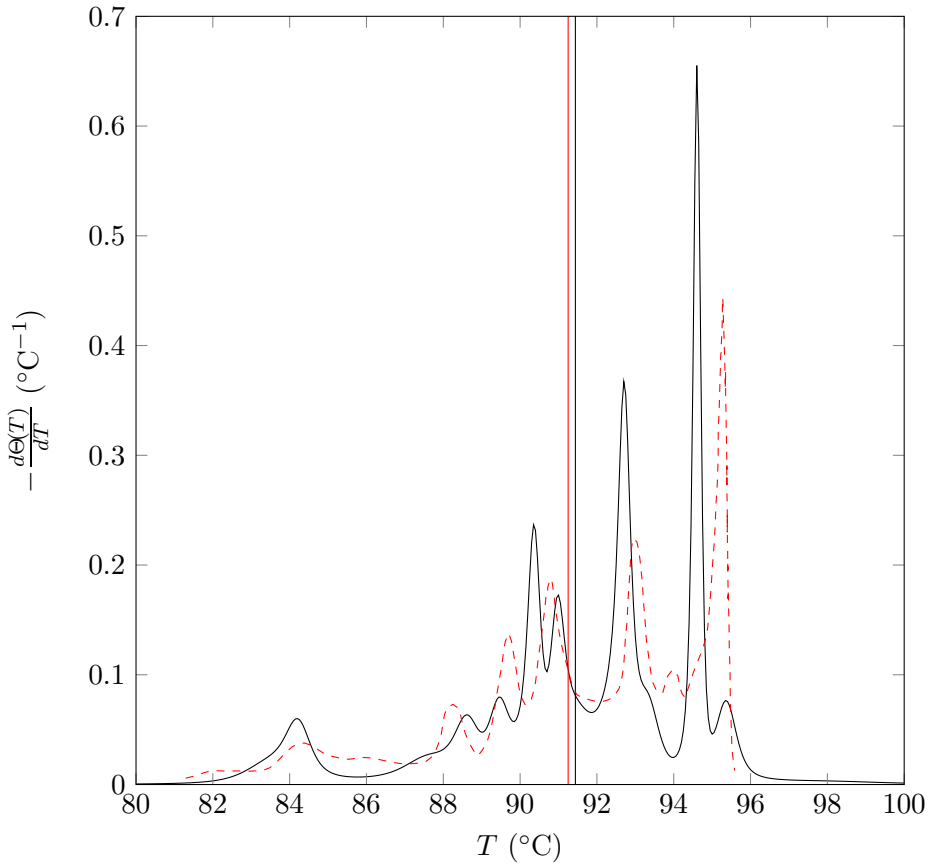


Figure 1.5: The melting profile (solid black line) and exact melting temperature (solid black vertical line) for *pBR322* plasmid DNA (4361 bp) under 150 mM NaCl shown here are obtained using `MELTSIM` [42, 45]. These predictions are in good agreement with experimental measurements presented by Vologodskii *et al.* in SSC buffer (dashed red line) [41]. While, the approximate melting temperature calculated using equation 1.17 is plotted for comparison (solid red vertical line). Note that the Poland's method ($\mathcal{O}(N^2)$) used in `MELTSIM` is accelerated by using exponential series to approximate loop functions ($\mathcal{O}(6N)$) proposed by Fixman and Freire [46].

understand various experiments involving DNA helix-coil transitions by mechanical constraints, such as DNA denaturing by unzipping force [47], or by twist-stretch force [48]. They will also be linked to the understanding of basepair disruptions observed under sharp bending constrains in my thesis studies.

1.6 DNA under constraints

The force is ubiquitous in biological systems, and it plays a critical role in lots of functional processes, such as cell division, cell adhesion or tissue formation. Mechanical forces are dynamically generated, and propagated on both cytoskeletons

and DNA. Furthermore, the cell is filled with proteins and nucleic acids, and this macromolecular crowding effect impacts many cellular activities as well. In the nucleus, the macromolecules have a high concentration of $\sim 300 - 400$ mg/ml, which corresponds to $\sim 20 - 30\%$ volume fraction [49, 50]. As a reult, DNA is constantly under lots of topological and mechanical constraints in living cells. In prokaryotes, the genome is usually a circular DNA, which is a topological domain itself. Although, in eukaryotes, the DNA molecules are mainly linear, it is known to attach to nuclear membranes, and contain many topology-isolated looped domains. In both cases, their high order structures are often negatively supercoiled and compartmentalized by attachments to scaffold proteins. More locally, DNA frequently associates with all kinds of proteins, where constraints arise all the time. For instance, eukaryotic DNA is topologically restrained by nucleosomes, while mechanical forces are applied to the chromsomes through force transmissions from the actomyosin contractions of the cytoskeleton [51, 52], and also produced by polymerases during replication and transcription [53].

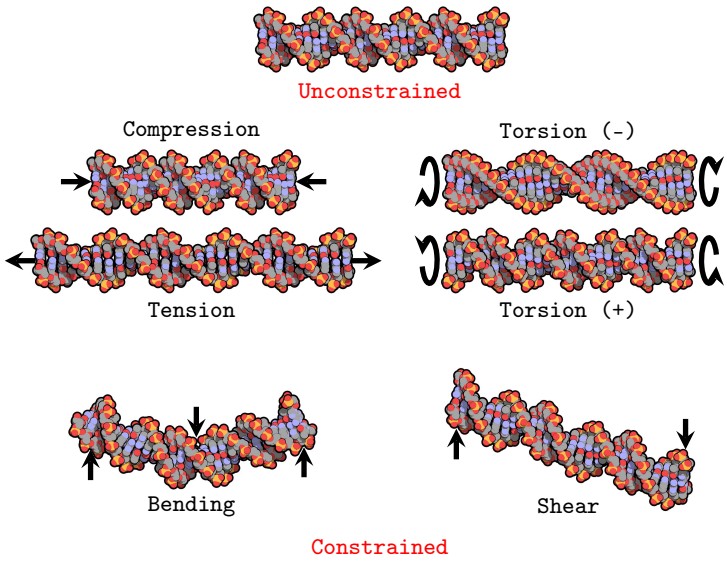


Figure 1.6: Illustrations of elementary stress types and corresponding elementary strain changes on DNA molecule. The unconstrained B-DNA is shown in detailed atomic model fro comparison. At the bottom, six possible types of elementary and idealized B-DNA stress-strain responses are listed, including compression, tension, negative, positive torsion, bending and shear, which result in negative, positive rise, negative, positive twist, $\sqrt{(\text{tilt})^2 + (\text{roll})^2}$ and $\sqrt{(\text{shift})^2 + (\text{slide})^2}$, respectively. Note that axial movements are directional, while radial movements are isotropic.

The complicated constraints can be decoupled into some elementary stresses, including tension, compression, bending, torsion and shear (Figure 1.6). The torsion is further distinguished into negative and positive torsion due to the helical nature of DNA. Actually, these categorizations are directly link to DNA material properties, and in turn to its intrinsic strain responses. These strain responses can be broken down into three translational (*i.e.*, shift, slide, rise in x, y, z-axis), and three rotational (*i.e.*, tilt, roll, twist in x, y, z-axis) movements (see Subsection 2.4.1 for details). Only those movements in z-axis (*i.e.*, axial movements) are directional, due to their non-symmetric responses to positive and negative stresses. More explicitly, tension leads to positive rise, while compression leads to negative rise. Positive and negative torsion corresponds to positive and negative twist. On the other hand, the radial movements usually assumed isotropic, and, it implies that the magnitude of stresses directly relate to the total magnitude of radial deformations regardless of directions. Therefore, shear causes lateral movements, which are the combination of shift and slide; bending induces angular movements, which are the combination of tilt and roll. Furthermore, because of the relative small bending and twisting rigidities of DNA, positive, negative torsions and bending are more commonly observed, while tension, compression and shear lead to only slight strain changes.

During DNA packaging, starting from entropically coiled DNA, further significant dimensional reductions are achieved under constrains. At first, we consider the compaction within elastic region, which predominantly results from balancing DNA bending and twisting deformations under their linear elastic limits. This process is well characterized by increasing linking number (Lk) of DNA under topological constraints,

$$\Delta\mathrm{Lk} = \Delta\mathrm{Tw} + \mathrm{Wr} \tag{1.20}$$

, where ΔTw is the change of helical turns, and Wr is the writhe. The writhe describes the spatial crossovers of DNA, resulting from local optimizations of molecular bending and twisting under certain constraints. The linking number is topologically invariant, unless DNA is covalently broken or under additional torsional stresses. The distributions of ΔTw and Wr relate to bending and torsion

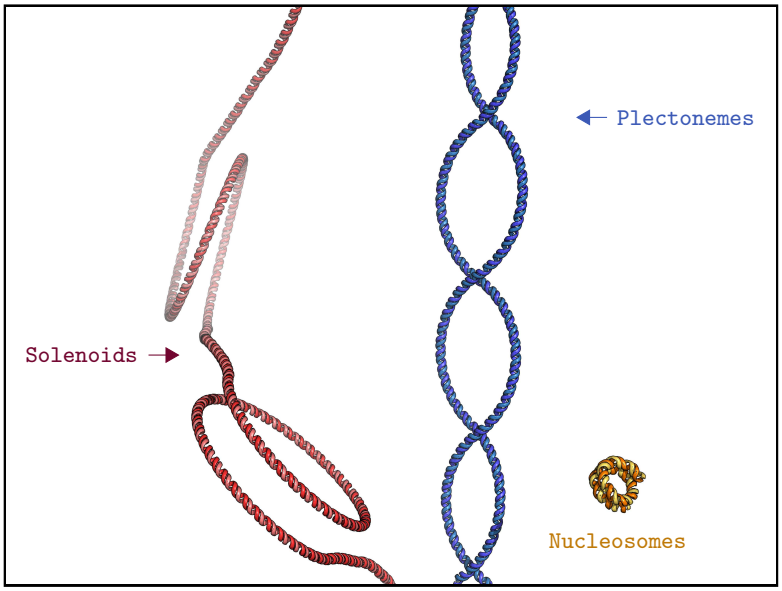


Figure 1.7: Illustrations for local shapes of plectonemes and solenoids, commonly found in negative supercoiled genomes for prokaryotes and eukaryotes, respectively. The left-handed solenoids are shown in dark red, and the right-handed plectonemes are shown in dark blue. Their dimensions are properly scaled with $R_{\text{eff}} \approx 20$ nm, which were commonly observed from *in vitro* experiments. They also corresponds to *in vivo* native supercoiling density, $\sigma \approx 0.06$ [54]. At the lower-right corner, DNA in nucleosomes are displayed for comparison, whose effective radius is ~ 5 nm. Apparently, there is a huge gap between these natural B-DNA topologies and fully packed DNA in term of energy costs.

strain energy partitions under constraints. The high order organizations of DNA molecules are facilitated by writhe formations in following two forms, plectonemes and solenoids, as shown in Figure 1.7. The genome of most organism is under negative supercoiling constraints [55], in prokaryotes, mainly in the form of right-handed plectonemes [56, 57], while in eukaryotes, mainly adapted to left-handed solenoids (*i.e.*, around histones in nucleosome).

But this compaction is limited, the homogeneous DNA solenoids [58] under condensations with effective radius, R_{eff}, in the range of $17 - 35$ nm, and plectonemes [59, 60] with opening angle $\alpha \approx 1$ rad, diameter $D \sim 6$ nm and $R_{\text{eff}} \approx \frac{D}{2(1-\sin\alpha)} \sim 20$ nm, *in vitro*. This is still much larger then that observed inside chromosomes *in vivo*; for example, DNA around nucleosomes, which are basic genome packaging units in eukaryotic cells, has $R_{\text{eff}} \approx 5$ nm. More severe bending and twisting deformations on DNA is becoming energetically unfavourable,

$$\Delta\mathcal{G} \propto \sigma^2 = \left(\frac{\Delta \text{Lk}}{\text{Lk}_0}\right)^2 \tag{1.21}$$

, where σ is the supercoiling density. The Gibbs free energy increases quadratically, and further condensations might quickly leads to homogeneity breakages.

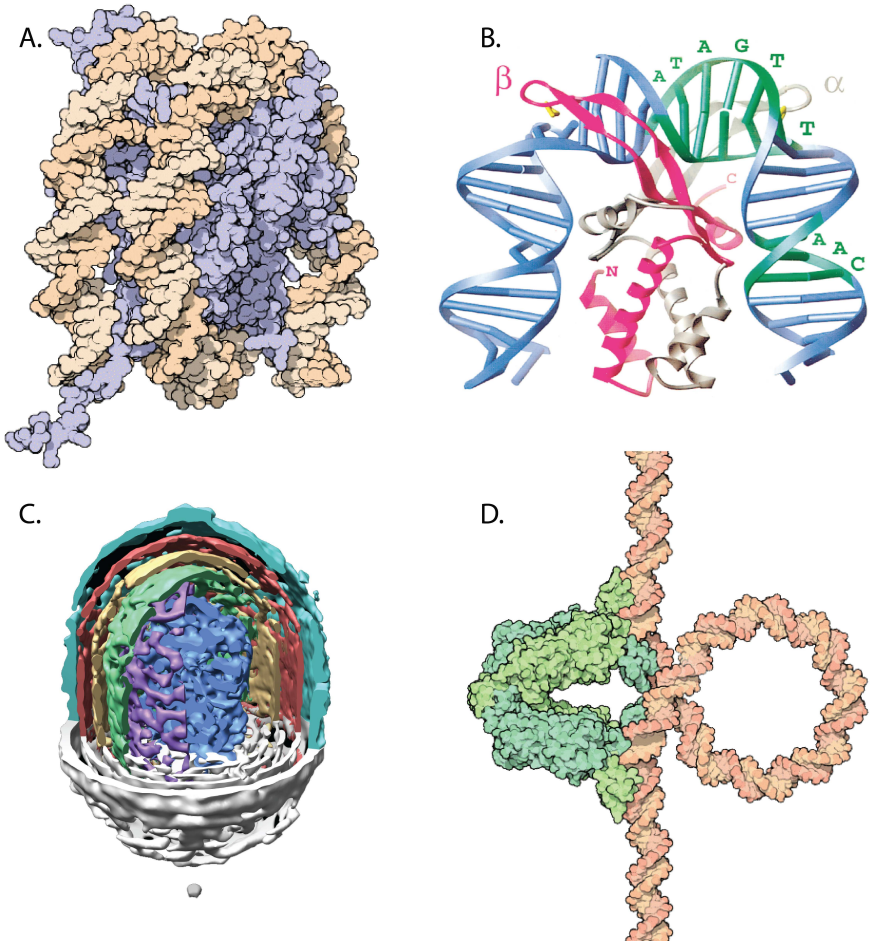


Figure 1.8: Various solved structures of biological events involving sharp bending DNA molecules. (a) The crystal structure of the nucleosome core particle, **1AOI**, with $\sim$ 147 bp DNA wrapped around [61]. (b) The crystal structure of IHF associated with nicked 35 bp H′ site of phage λ, **1IHF**, which is IHF high affinity binding site, inducing $> 160°$ bending. (c) 3D Cryo-electron microscopy imaging of ϕ29 [62]. The concentric layers of DNA are shown in different colour, whose out-most shell has a diameter of $\sim$ 33 nm. (d) The crystal strucutre of *lac* repressor with modelled 93 bp repression loop, **1LBI** [63].

Indeed, DNA conformations under sharp bending conditions are commonly found in biological events, where short pieces of DNA are highly curved, usually through protein-mediated constraints. We have explored the example of nucleosome, which relates to eukaryotic genome compaction and gene regulation, where DNA wraps around histone proteins forming $> 300°$ looping over only $\sim$ 94 bp [61, 64, 65]. While, in prokaryotes, some NAPs function as architectural factors by inducing sharp bending, which reverse the directions of DNA in very short

distances. For instance, integration host factor (IHF) intercalates into minor grooves of DNA, then introduces a $> 160^\circ$ bending on ~ 12 nm contour length [66]. Furthermore, the genome packaging processes in virus, such as *ds*DNA bacteriaphage $\phi 29$, also involve tightly organized and sharply bend DNA. The 6.6 μm DNA molecule is pumped into its small capsid (*i.e.*, 42×54 nm in dimensions [67]), and is looped inside the cavity forming multiple DNA layers, with minimum radius $R_{\text{eff}} \approx 7.2$ nm [62]. Note that there is an even extremely curved DNA toroid with diameter of ~ 60 Å, observed at the connector cavity of the phage [68], which is proposed to be able to retain inner coiled DNA. Besides above-mentioned structural functions, it has been shown that small DNA looping formations repressively regulate gene transcriptions through protein-mediated cross-linkings, *e.g.* by *lac* repressor. In which case, the two arms of this tetrameric repressor grab two binding sites ~ 93 bp apart along DNA. As a result, the small ~ 31 nm DNA in between forms a planer loop with a bending angle of $\sim 360^\circ$, and hides the *lac* operon inside to prevent transcriptions [69, 70, 63].

The entire DNA molecule in organism and many viruses on earth are dynamically packed for storage and unpacked for deciphering. Evidently, from the compaction perspective, the homogeneous B-DNA can be packed until a limitation of $R_{\text{eff}} \approx 20$ nm has reached. Nevertheless, more severe compaction have been commonly observed with R_{eff} as low as 3 nm, such as DNA packed in phage collar. Then, from the elasticity perspective, the stiffness of B-DNA, modelled as homogeneous WLC polymer, has been consistently determined to be $A \approx 50$ nm under weak constraints at intermediate-to-long length scales. While, under more extreme constraints and short length scales, the apparent A is much smaller than 50 nm as revealed in recent cyclization experiments. Further, it has been shown that single-strand-specific endonucleases were able to cleave DNA minicicles with size of ~ 64 bp, suggesting that in such sharply bent DNA, melted DNA bubbles forms [36]. It means that B-DNA cannot retain its homogeneity all the time, therefore, dynamically and locally adapts to strong constraints by inducing homogeneity breakages. In other words, B-DNA gives up its B-form (*i.e.*, breaks) to compensate strong constraints, such as sharp bending, and subsequently to reduce the energy costs.

1.7 Generalized DNA polymer model

The homogeneity breakage of DNA under sharp bending condition do not obey traditional WLC model, which treats DNA as a homogeneous thin rod with harmonic bending potential (*i.e.*, linear elastic response). As demonstrated in several previous theoretical studies [30, 4, 5, 71], homogeneity breakage of DNA can be treated as excitation of mechanical defects, which can in general be described by nonlinear DNA bending elastic response. One feasible type of such defect is a kinked DNA basepair step. As pointed by Crick and Klug in 1975 [6], such a kinked basepair step can be a local energy minimum. Another possibility is DNA basepair melting, creating a flexible hinge in the DNA [30, 5]. These two types of defects have similar mechanical effect, as they both allow DNA to form a large bending angle at defects, thus relaxing DNA overall bending. Below I briefly review why mechanical defect excitation can be modelled by a generalized polymer model with nonlinear bending elasticity.

As demonstrated by Yan *et al.* [30, 4], the bending energy of a DNA molecule subject to mechanical excitation can be generally described as,

$$E\left(n_i;\hat{t}_i,\hat{t}_{i+1}\right)=\delta_{n_i,0}E^0\left(\hat{t}_i,\hat{t}_{i+1}\right)+\delta_{n_i,1}\left(E^1\left(\hat{t}_i,\hat{t}_{i+1}\right)+\mu\right) \tag{1.22}$$

, where $E\left(n_i;\hat{t}_i,\hat{t}_{i+1}\right)$ is the vertex energy of the i^{th} vertex in a discretized polymer chain, $n_i=0$ denotes a vertex in the B-form, while $n_i=1$ indicates a defected vertex. E^0 and E^1 correspond to the vertex energies in intact state and in defected state, respectively. μ is an energy cost associated with defect excitation. And $\delta_{i,j}=\begin{cases}0, & \text{if } i\neq j\\ 1, & \text{if } i=j\end{cases}$ is the Kronecker delta function. The vertex energies in different states are,

$$\begin{aligned}\beta E^0&=\frac{a}{2}\left(\hat{t}_i-\hat{t}_{i+1}\right)^2=a\left(1-\cos\theta\right)\\ \beta E^1&=\begin{cases}\frac{a'}{2}\left(\hat{t}_i-\hat{t}_{i+1}\right)^2=a'\left(1-\cos\theta\right), & \text{in case of hinge}\\ \frac{a'}{2}\left(\hat{t}_i\cdot\hat{t}_{i+1}-\cos\gamma\right)^2=\frac{a'}{2}\left(\cos\theta-\cos\gamma\right)^2, & \text{in case of kink}\end{cases}\end{aligned} \tag{1.23}$$

, where E^0 and E^1 can be written as functions of either tangent vectors, $\left(\hat{t}_i,\hat{t}_{i+1}\right)$

or bending angle, θ, while γ is kinked energy minimal angle; and $a = \frac{A}{l}$ together with $a' = \frac{A'}{l}$ are dimensionless vertex bending rigidities for different states.

By summing the Ising index n_i in calculation of the partition function, the above excitation model has been shown to be equivalent to a generalized polymer model with an nonlinear bending elasticity [4],

$$\beta E_{\text{gen}} = -\ln\left(\exp\left(-\beta E^0\right) + \exp\left(-\beta E^1 - \beta\mu\right)\right). \tag{1.24}$$

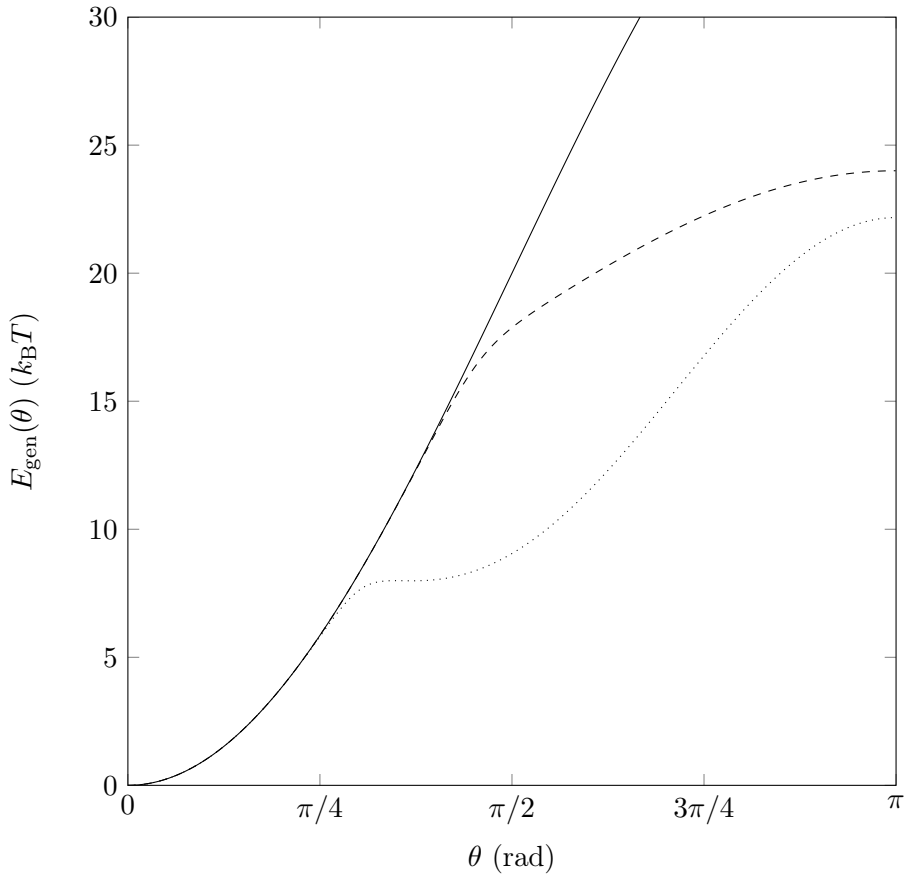


Figure 1.9: At a fixed segment contour length of $l \ll A$ ($l = 2.5$ nm), the effective DNA bending energy is calculated for (1) B-DNA with $A = 50$ nm (solid line, based on WLC moldel), (2) DNA subject to excitation of a softening defect with $A' = 15$ nm, $\mu = 12\ k_{\text{B}}T$ (dashed black line, based on Equation 1.24 with hinged E^1) and (3) DNA subject to excitations of an intrinsic kinking defect with $A' = 30$ nm, $\mu = 8\ k_{\text{B}}T$, $\gamma = 68°$ (dotted black line, based on Equation 1.24 with kinked E^1). Both energy profiles (black) for softening and intrinsic kinking defect excitation models contain sudden energy cost reductions not far beyond μ and deviation from the WLC harmonic energy profile.

Figure 1.9 shows representative vertex energy profiles for B-form DNA, vertex subject to flexible hinge excitation, and vertex subject to kink excitation. As can

been seen from the figure, for vertex subject to flexible hinge or kink excitation, the bending energy profiles deviate from that of B-form when the bending angle exceeds certain threshold angle θ_{c} determined by the equation, $E^0(\theta_{\mathrm{c}}) - E^1(\theta_{\mathrm{c}}) = \mu$.

Here we note that similar defect excitation models were proposed by several groups independently, but were named differently. For example, the kinkable WLC (KWLC) model proposed by Wiggins in 2005 [5] is equivalent to the flexible hinge excitation model by setting the vertex bending stiffness to be zero ($a' = 0$). As another example, the empirical linear sub-elastic chain (LSEC) model, where its energy takes the simple form *vs.* bending angle, as, $\beta E_{\mathrm{LSEC}}(\theta; l) = \alpha\,|\theta|$ ($\alpha = 6.8\ k_{\mathrm{B}}T$ for $l = 2.5$ nm) [71, 22], can also be approximated by the flexible hinge excitation model with appropriately chosen parameters.

Chapter 2

Molecular dynamics and DNA structural analysis

2.1 Introduction

MD simulation is a method to simulate the real-time physical motions of molecules at atomic level based on Newton's second law. Benefit from the rapid expansion of computational power, nowadays, MD simulations are applicable to larger systems with longer simulation time, and utilized to probe certain interesting biological problems. Quite uniquely, it can provide direct theoretical approaches to complicated biological events, through obtaining detailed behaviours at atomic resolution, as well as offering great manipulative capabilities.

In this chapter, we focus on simulation and analysis methods used in our DNA micromechanical studies. We start with some MD simulation basics, and DNA force fields to provide an overall picture about MD simulation. While, simulation associated subjects, including DNA initial modelling, structure analysis and advanced sampling are introduced in latter part. DNA conformational analysis, which is a key component for bridging simulated raw data to our results, is the reverse processes of DNA initial building; it is explained side-by-side with DNA initial modelling section for better understanding on both topics.

2.2 Molecular motion in discretized time

For a system with N atoms initiated from certain state, it self-evolves driven by thermal fluctuations and intrinsic interactions according to Newton's equations of motion. In MD simulation, the current state is defined by sets of position vectors, denoted by $\boldsymbol{r}^N$, which is usually modelled from solved structures by crystallography or nuclear magnetic resonance (NMR); and by sets of velocity vectors, denoted by $\boldsymbol{v}^N$, which were initiated randomly from Maxwell-Boltzmann distribution at certain temperature T,

$$p(\boldsymbol{v}_i) = \sqrt{\frac{m_i}{2\pi k_{\mathrm{B}} T}} \exp\left(-\frac{m_i \boldsymbol{v}_i^2}{2k_{\mathrm{B}} T}\right) \tag{2.1}$$

, where i indexes particular atom, and m_i is its atomic mass. Then, its time evolution is uniquely determined through Hamiltonian dynamics, which is a re-formulation of Newtonian dynamics.

Under the Hamiltonian of system, which is the combination of kinetic energy, $\mathcal{K}$ and potential energy $\mathcal{V}$,

$$\mathcal{H}\left(\boldsymbol{p}^N, \boldsymbol{q}^N\right) = \mathcal{K}\left(\boldsymbol{p}^N\right) + \mathcal{V}\left(\boldsymbol{q}^N\right) \tag{2.2}$$

, the motion of atoms obeys,

$$\begin{aligned} \dot{\boldsymbol{q}}_i &= \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}_i} \\ \dot{\boldsymbol{p}}_i &= -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}_i} \end{aligned} \tag{2.3}$$

, where $\boldsymbol{q}_i$ is the generalized coordinates, and $\boldsymbol{p}_i$ is the generalized momenta. In Cartesian coordinates with conservative potential energy, the Hamiltonian is simplified, $\mathcal{H}\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right) = \sum_i \frac{\boldsymbol{p}_i^2}{2m_i} + \mathcal{V}\left(\boldsymbol{r}^N\right)$, and we can rewrite Equation 2.3 into more familiar forms, as,

$$\begin{aligned} \dot{\boldsymbol{r}}_i &= \frac{\boldsymbol{p}_i}{m_i} = \boldsymbol{v}_i \\ \dot{\boldsymbol{p}}_i &= -\frac{\partial \mathcal{V}\left(\boldsymbol{r}^N\right)}{\partial \boldsymbol{r}_i} = \boldsymbol{f}_i. \end{aligned} \tag{2.4}$$

Now, the MD trajectories can be populated by numerically integrating the $6N$ first-order differential equations, Equation 2.4, step-by-step (*i.e.*, with small time step δt) in discretized time based on finite difference methods. Here, we

are going to introduce the two frequently used algorithms for MD integration schemes, the leap-frog [72] and velocity Verlet [73] integrators. The leap-frog algorithm updates the "half-way" velocity and stepwise position alternatively, as,

$$\begin{aligned} \boldsymbol{v}\left(t+\frac{\delta t}{2}\right) &= \boldsymbol{v}\left(t-\frac{\delta t}{2}\right)+\frac{\delta t}{m}\boldsymbol{f}(t) \\ \boldsymbol{r}\left(t+\delta t\right) &= \boldsymbol{r}(t)+\delta t\,\boldsymbol{v}\left(t+\frac{\delta t}{2}\right). \end{aligned} \tag{2.5}$$

While, the velocity Verlet algorithm simultaneously solve the stepwise position and velocity, with the help of half-step velocity,

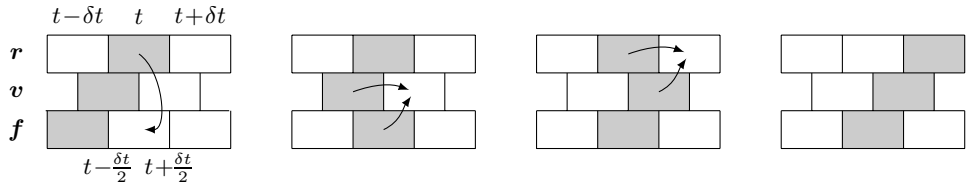
$$\begin{aligned} \boldsymbol{r}\left(t+\delta t\right) &= \boldsymbol{r}(t)+\delta t\,\boldsymbol{v}(t)+\frac{\delta t^2}{2m}\boldsymbol{f}(t) \\ \boldsymbol{v}\left(t+\delta t\right) &= \boldsymbol{v}(t)+\frac{\delta t}{2m}\left(\boldsymbol{f}(t)+\boldsymbol{f}\left(t+\delta t\right)\right) \end{aligned} \tag{2.6}$$

, where the relation $\boldsymbol{v}\left(t+\frac{1}{2}\delta t\right) = \boldsymbol{v}(t)+\frac{\delta t}{2m}\boldsymbol{f}(t)$ is substituted to get above iterative equations. The updating schemes for these algorithms are explicitly illustrated in Figure 2.1. Note that the subscripts are ignored in above equations for simplicity. Then, these generated time evolutions of system are constantly adjusted using varies constraints to obtain better equilibrated ensembles under Boltzmann distribution, through center-of-mass motion removal, temperature and pressure coupling, *etc.*

The leap-frog and velocity Verlet integrators gain their popularities in MD simulation due to their time reversibility and symplectic nature. The time reversibility means the generated path in phase space can be exactly traced backward by setting time interval to $-\delta t$. This property is the fundamental prerequisite for system to achieve equilibrium ensemble, and practically important for the steady behaviours of long time simulations. That is the reason why it is a preferable choice for the slow convergent biomolecular dynamics, although its cumulated error is relatively large (*i.e.*, $\mathcal{O}\left(\delta t^2\right)$) against other high order approximation solutions, for instance, fourth-order Runge-Kutta (*i.e.*, $\mathcal{O}\left(\delta t^4\right)$). Furthermore, these symplectic algorithms preserves certain important properties of original Hamiltonian system, such as, volume of phase space.

Due to the simultaneously evaluated $\boldsymbol{r}$ and $\boldsymbol{v}$ at same t, velocity Verlet initiates naturally, but leap-frog requires knowledge about previous velocity, which

(a) Leap-frog scheme



(b) Velocity Verlet scheme

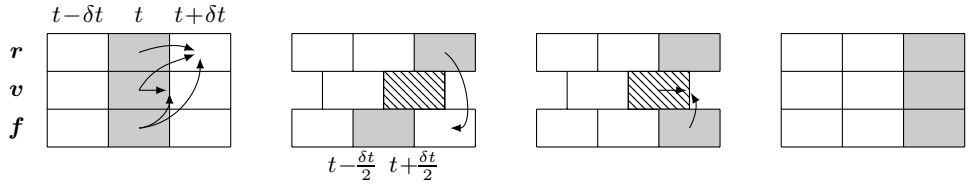


Figure 2.1: The different versions of Verlet algorithm [74], (a) leap-frog form and (b) velocity Verlet form. Under leap-frog, the evolutions of position and velocity have a half-step time difference, but those of velocity Verlet are at the same pace. The force are called once in (a) but twice in (b), which cause troubles to the latter under large N system.

lead to their non-identical trajectories. More importantly, some detailed refinements on ensembles, like pressure, are only possible through velocity Verlet. On the other hand, leap-frog is more efficient for large system, because of its half-size global communication calls on forces. Therefore, generally speaking, we use leap-frog algorithm for NTV ensemble, and velocity Verlet integrator for NPT ensemble in our simulations.

2.3 DNA force field

A leftover quantity, the potential energy, $\mathcal{V}(\boldsymbol{r}^N)$, that is barely touched during last section, is our knowledge about the system. It is assumed to well reflect the behaviours of nature, and theoretically approachable through quantum mechanics (QM). However, due to its computational limit, for biological system, we still need to stay in the classical regime, and using a series of empirical energy functions to approximate the system intrinsic potential energy. Such approximation includes physically meaningful potential energy forms (Equation 2.7) and a set of experimentally determined parameters. Together with the topologies of

molecules, they are called "force field".

$$\begin{aligned}\mathcal{V}(\boldsymbol{r}^N) = &\sum_{\text{Bonds}} K_r(r-b_0)^2 + \sum_{\text{Angles}} K_\theta(\theta-\theta_0)^2 \\ &+ \sum_{\text{Dihedral}} \sum_{n} \frac{\mathcal{V}_n}{2}(1+\cos(n\phi-\gamma)) \\ &+ \sum_{i<j} \left(4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)\right) \\ &+ \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}.\end{aligned} \tag{2.7}$$

In Equation 2.7, this total potential energy are composed of bonded and non-bonded terms, where first three are sums of bonded potentials, representing two, three and four body covalent interactions, respectively; last two are sums of non-bonded potentials, including van der Waals (VDW) and electrostatic interactions. These interactions are expressed in commonly adapted forms, and are illustrated in Figure 2.2. In the following paragraphs, each interaction and corresponding parameters are explicitly described in detail.

The bonded interactions are covalently based, containing stretch, bending, torsion responses for particular bond. The bond stretching is usually described by harmonic potential, where the energy increases quadratically with the magnitude of deviations against the equilibrium length, b_0. Sometimes, more realistic, but computation inefficient functions are used, such as Morse potential. For describing the bond angle vibrations, harmonic functions with energy minimal θ_0 are frequently used, but other versions are also possible, for instance, the cosine based angle potential in GROMACS-96 force field. Regarding the four-body interactions, the dihedral angle (*i.e.*, torsional angle) often prefers several angle ranges to avoid steric congestions, *e.g.*, syn, anti; cis, trans, gauche+, gauche−, *etc.* Thus, its potential energy is wavelike, which is commonly approximated by simple periodic form, Ryckaert-Bellemans function or several truncated terms (*i.e.*, $n = 3$ or 4) of its Fourier series, generally expressed as the third term in Equation 2.7, where γ denotes a phase shift. Additional "improper" dihedral terms are incorporated into force fields to restrain the out-of-plane motions, like those in DNA bases. For AMBER force fields, they are expressed in the same way as proper dihedral, while taking a harmonic form in CHARMM force

fields. Besides these independent terms, coupling effects can be handled using cross terms, such as stretch-stretch, bending-bending, stretch-bending, *etc.* For example, CHARMM force fields have a Urey-Bradley term, which represents stretch-bending coupling.

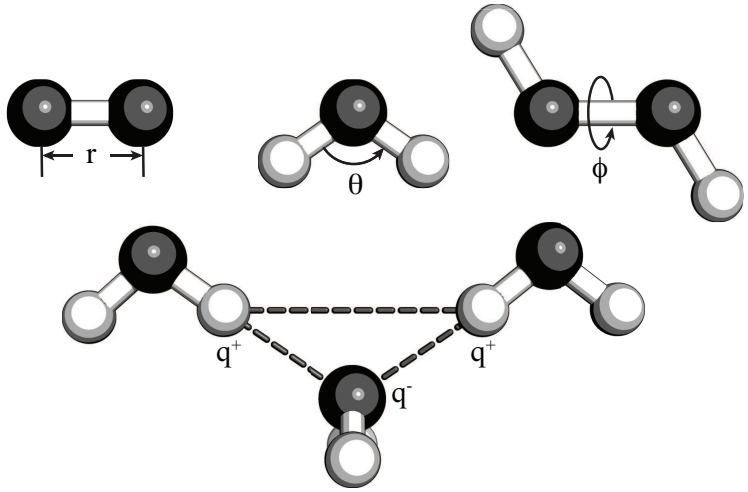


Figure 2.2: Illustrations on different empirical functions in force field. Bonded interactions: (a) $1-2$ interactions. The stretching of bond follows Hook's law, which equilibrates at bond length, b_0, with a stiffness of twice K_r. (b) $1-3$ interactions. The bending of bond is in harmonic form as well, whose energy is $K_\theta\left(\theta-\theta_0\right)^2$. (c) $1-4$ interactions. Torsional energy is often periodic, taking various types even in single force field. Proper dihedral interactions reflect the torsional constrains in sequential linked atoms, $ijkl$; the torsion angle ϕ is defined by dihedral angle between ijk and jkl plane. Improper dihedral interactions are sometimes defined on non-sequential linked atoms to restrain certain geometry. Non-bonded interactions: (d) pairwise VDW and Coulomb interactions. The figure uses electrostatic interactions as an example, where partial point charges locate at the center of atoms, representing the polarized electron cloud. The like charges repel each other and unlike charges attract each other, while their forces are centro-symmetric.

The non-bonded interactions are pairwise electromagnetic forces between atoms, even without requiring shared electrons. The VDW interactions are represented by the Lennard-Jones interactions (*i.e.*, $6-12$ potential), containing a repulsive, r_{ij}^{-12}, and an attractive terms, r_{ij}^{-6}, which have an energy minimum when their inter-atom distance equals the sum of VDW radii of the two atoms. Occasionally, the repulsion forces are described by an exponential form, which yields the robust, realistic, but expensive Buckingham potential. On the other hand, the electrostatic effects are modelled using Coulomb interactions between

partial point charges at the center of atoms, as the last term in Equation 2.7, where its coefficient normally written in relative dielectric constant, $\epsilon_r = 4\pi\epsilon_0$. Although the pairwise interactions are also applicable to boned atoms, they are assumed to be folded into two-body and three-body interactions already. As a result, we only consider the pairs that are three or more than three bonds away from each other. However, for $1-4$ non-bonded interactions, they are certainly coupled with torsion constraints. To compromise the overestimation effects, scale factors are commonly employed for these four-body non-bonded interactions, for instance, in AMBER force fields, 1/2.0 and 1/1.2 are used to scale down $1-4$ VDW and $1-4$ electrostatic interactions respectively. From computational point of view, the exhausted summations of these pairwise interactions are inefficient, especially for large system. Practically, these pairwise terms are only evaluated at short range with the help of neighbour lists, while long-range interactions are tackled with special techniques. In `GROMACS` package, dispersion corrections are integrated for the cut-off effects of long-range Lennard-Jones potentials. Respect to long-range electrostatic interactions, Particle-Mesh Ewald method [75, 76] are used to speed up the reciprocal summations.

Next, in order to better understand and select DNA force fields, we are going to outline the processes for force field parameterizations. It is a very rich and challenging subject to approach the potential energy of system by fitting above analytical functions with large amount of parameters, through incorporating different kinds of experimental data, such as known liquid solid properties, vibrational frequencies from Raman spectroscopy, structures from NMR, crystallography and *ab initio* QM calculations, *etc.* Atom types of particular force field are defined based on their targeting system; for biomolecules, such as protein or nucleic acids, $\sim$ 40 to 60 atom types are needed. Same element may have multiple atom types, and each describes certain hybrid orbital resulting from association with particular neighbours. Interactions involving different atom types have different parameters. Note that if certain parameters are deviate from optimization a lot, the potential energy can still be restored from corrections on some other parameters. So, the relatively straightforward bend, angle parameters are firstly fixed. Then, the parameters for VDW, Coulomb interactions were

carefully determined on top of b_0, θ_0, K_r and K_θ. Finally, the coefficients for error dominant torsional energy were tuned at last, through extensive comparisons against relative large-scale experimental observations and/or *ab initio* calculated energy landscapes for specific system. Force fields are updated and corrected constantly, given some new experimental observations or unexpected simulated behaviours, while these modifications are usually on these dihedral interactions.

For DNA molecules, we are focusing on the development of AMBER force fields, because its latest DNA version, ParmBSC0, is currently used under our simulations. In 1994, Cornell *et al.* [77] presented explicitly stated Parm94, which was a major modifications based on Weiner *et al.* [78, 79] force field and targeted its applications on biomolecular simulations. It equipped with improved charge models based on 6-31G* basis set, and recalculated VDW parameters using liquid simulations. These provided profound base for subsequent AMBER force fields. Regarding DNA related dihedral potentials, corresponding parameters in Parm94 were deliberately determined through studying free energy *vs.* sugar pucker, backbone angle γ and base angle χ. Notably, the expected preference of C2'-endo over C3'-endo sugar puckering outperformed early versions of CHARMM force fields. Later, Parm98 [80] modified some torsional potentials, which led to better DNA morphologies that resemble crystal structures in twist, rise, minor and major groove width, *etc.* Nevertheless, Parm98 was known to increase the energy barrier for C2'-endo to C3'-endo transition, which over-stabilized *B*-form. This problem was corrected by Parm99 [81], which obtained superior sugar-puckering properties and χ angles on top of Parm98. In early 2000, Zakrzewska *et al.* [82] reported a massive irreversible α/γ transitions away from expected gauche−/gauche+ state in a 50 ns MD simulation. Its amendment gave rise to the newest ParmBSC0 [83], which well fitted the simulated α/γ potential energy to *ab initio* QM calculations through introducing an additional atom type CI and its associated torsion terms. More importantly, ParmBSC0 improved the helical shapes and stabilities of *B*-DNA, which prolonged the possible simulation timescales up to $\sim$ 200 ns.

2.4 DNA initial structure generation

In order to initialize an efficient DNA MD simulation, we need to design their full-atomic structures. These initial guesses are critical, because a "good" guess helps to shorten reaction path and reduce simulation time. Generally speaking, a reasonable initial conformation should have smooth energy distributions among all atoms, and should be as close as possible to final conformation of interest. In this section, we devote many efforts in understanding the geometries and arrangments of DNA building blocks (*i.e.*, Watson-Crick basepairs for *A*, *B*-DNA), in order to achieve better initial DNA structures.

2.4.1 Basepair reference frame and orientation

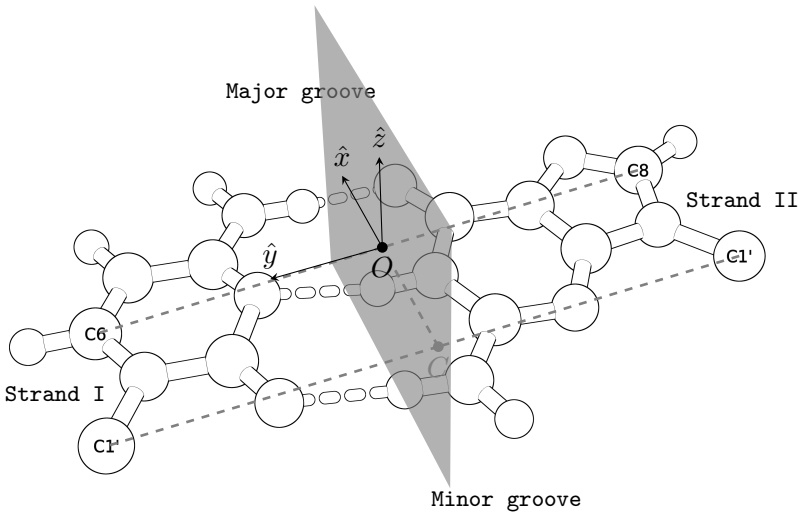


Figure 2.3: Here we use C≡G basepair as example, and only complementary bases are shown. The gray plane, which is the perpendicular bisector of the line segment $\overline{\text{(C1' C1')}}$ at the midpoint C, intersects with the line segment $\overline{\text{(C6 C8)}}$ at O. x-axis directs from C to O. y-axis is parallel to $\overline{\text{(C1' C1')}}$, pointing towards the Strand I. z-axis is $\hat{z} = \hat{x} \times \hat{y}$.

To generate the whole DNA molecule, ideal Watson-Crick basepairs (A=T, T=A, G≡C and C≡G) are sequentially arranged and stacked one after another. In order to do that, we need some geometry definations. Given an ideal Watson-Crick basepair, a reference frame is required to assign its three-dimensional orientation as a whole. The standard reference frame, suggested by Olson *et al.* [84], follows the guidelines proposed on EMBO Workshop on DNA Curvature

and Bending, 1988 [85]. It is defined using the coordinates of four atoms, C6 from pyrimidine (`C` and `T`), C8 from purine (`G` and `A`), and two sugar C1' atoms. The pseudo-dyad axis of basepair (*i.e.*, x-axis) runs from the midpoint between two C1' atoms towards the major groove, passing through the intersection point of $\overline{\text{(C1' C1')}}$ perpendicular bisection plane and $\overline{\text{(C6 C8)}}$ line. Then this intersection point is defined as the origin of reference frame. Staring from the origin, the y-axis points to the sequence strand (*i.e.*, Strand I) and parallels with $\overline{\text{(C1' C1')}}$ line. The z-axis of this right-handed triad is perpendicular with xy-plane.

Starting from the 5′ end of Strand I to its 3′ end, we number the bases on Strand I from 1 to N, and from $2N$ to $N+1$ for bases on Strand II. Along the same direction, the basepairs, as well as the reference frames, are enumerated from 1 to N. For $(i+1)^{\text{th}}$ basepair, the positions of all atoms are fixed within its reference frame, and oriented as a group in 3D against that of i^{th} basepair. This spatial arrangement can be described by six orientation parameters (*i.e.*, sequential basepair parameters), three rotational parameters, tilt, roll, twist, denoted by τ, ρ and Ω, and three translational parameters, shift, slide, rise, denoted by D_x, D_y and D_z along x, y, z-axis respectively (Figure 2.4). The axial directions mark the positive directions of translational parameters, while the rotational parameters obey right-handed rules.

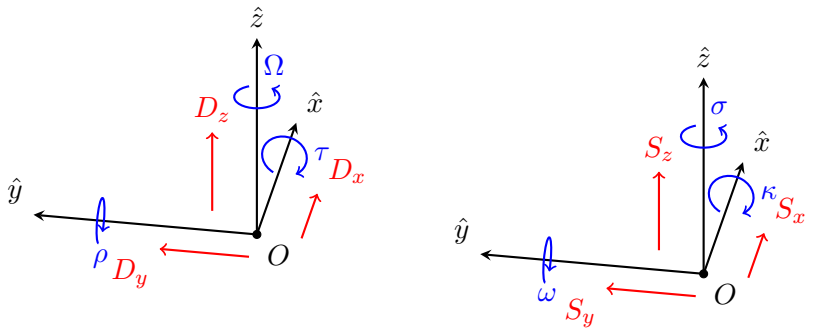


Figure 2.4: In the left coordinates, the six sequential basepair parameters describe the relative orientation between adjacent local basepairs. The rotational parameters, including tilt (τ), roll (ρ) and twist (Ω), are coloured in blue, and the translational parameters, including shift (D_x), slide (D_y) and rise (D_z), are coloured in red. While, in the right coordinates, the six complimentary base parameters describe the relative orientation of the two bases within basepair. The rotational parameters, including buckle (κ), propeller twist (ω) and opening (σ), are coloured in blue, and the translational parameters, including shear (S_x), stretch (S_y) and stagger (S_z), are coloured in red. The arrow directions mark the positive directions of parameters.

2.4.2 DNA helix computation scheme

Here we introduce the Cambridge University Engineering Department Helix computation Scheme (CEHS) [86] for achieving target basepair arrangements from given aforementioned parameter sets. Since the non-commutativity among rotational parameters, additional constrains are required for unique descriptions of spatial arrangements from τ, ρ and Ω. One of such constraints is the invariance of parameter values when changing directions. In other words, the set of parameters should be the same no matter arranging $(i+1)^{\text{th}}$ basepair related to i^{th} basepair or *vise versa*. Based on the recommendation by EMBO workshop [85], the concept of middle frame (M) was used to attain such reversibility, which is the "half-way" reference frame located right between i^{th} and $(i+1)^{\text{th}}$ reference frame. With the help of middle frame, we describe the detailed processes for carrying out CEHS as Figure 2.5, then summarize it mathematically later.

1. We, firstly, prepare to bend the DNA, where the total bending (*i.e.*, Roll-Tilt angle) is the combination of tilt and roll, as,

$$\Gamma = \sqrt{\tau^2 + \rho^2}. \tag{2.8}$$

   Because the bending is directional, which is within a particular plane in 3D, we need to find a new axis first, that is perpendicular to this plane, and functions as "hinge" in the bending process. This axis is defined as Roll-Tilt axis (H), which is determined by relative magnitudes and signs of roll, tilt angles, as follow,

$$\hat{H} = \frac{\rho}{\Gamma}\hat{y}_{\text{M}} + \frac{\tau}{\Gamma}\hat{x}_{\text{M}} \tag{2.9}$$

   , where $\hat{y}_{\text{M}}$ and $\hat{x}_{\text{M}}$ is the y, x-axis unit vector of middle frame. And this $\hat{H}$ is inclined with y-axis at an angle $\phi = \tan^{-1}\left(\frac{\tau}{\rho}\right)$.
2. Starting from the middle frame, we align y-axis with $\hat{H}$ by rotating $-\phi$ about z-axis. Then, we rotate $(i+1)^{\text{th}}$ basepair half Roll-Tilt angle about $\hat{H}$ positively, and i^{th} basepair same amount negatively to get total bending Γ. This also ensures the correct relative distributions of tilt and roll.

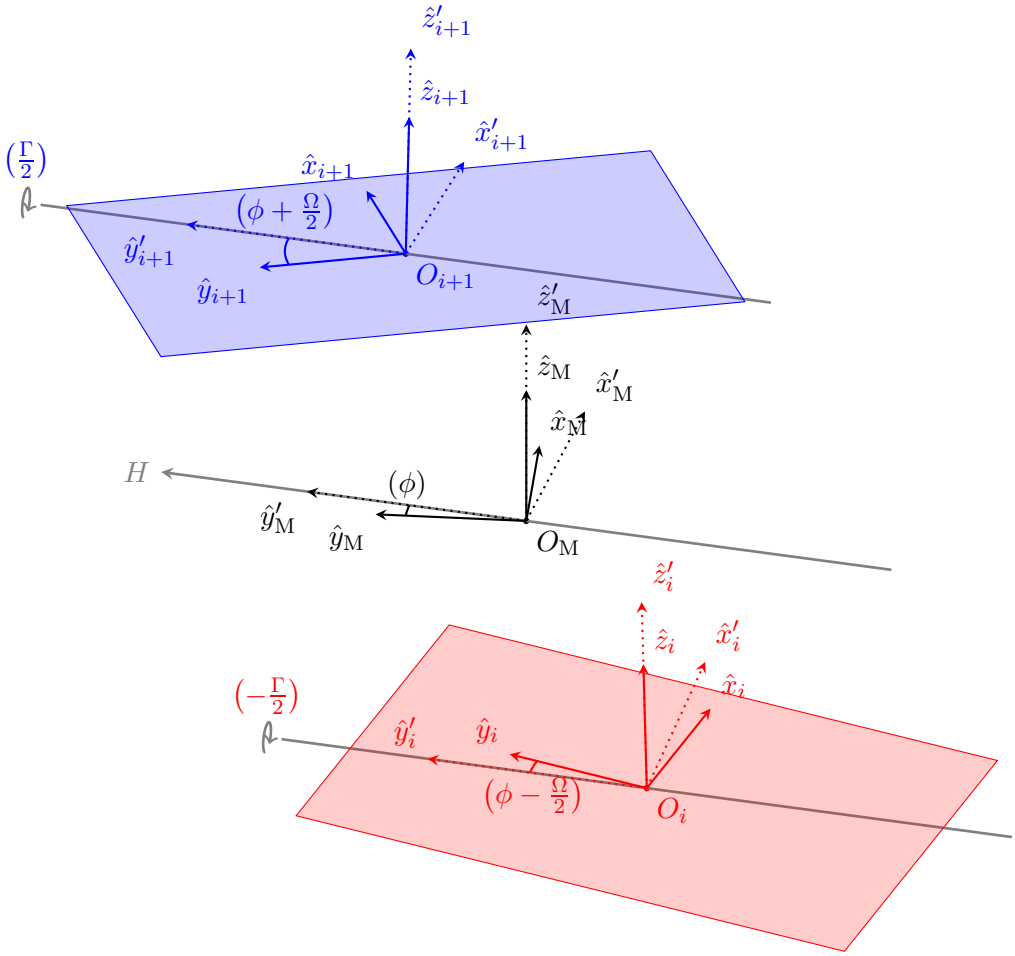


Figure 2.5: The CEHS constructive illustrations for building sequential basepairs spatial arrangements starting from the "middle frame" (solid black triad). For clear demonstrations, the i^{th} (red) and $(i+1)^{\text{th}}$ (blue) basepairs are moved apart by same amount before any rotations. The "hinge" (H in gray; Roll-Tilt axis) is the bending axis, which is inclined at $\hat{y}_{\mathrm{M}}$ with (ϕ). First, we rotate the solid black triad against $\hat{z}_{\mathrm{M}}$ by $(-\phi)$, which aligns $\hat{y}'_{\mathrm{M}}$ with H and gets the dotted black triad. Then, we bend it about H by $\left(\frac{\Gamma}{2}\right)$ and $\left(-\frac{\Gamma}{2}\right)$ to generate the dotted blue and red triad, respectively. Finally, the reference frame of $(i+1)^{\text{th}}$ basepair (solid blue triad) is achieved by rotating $\left(\phi + \frac{\Omega}{2}\right)$ around $\hat{z}'_{i+1}$, while that of i^{th} basepair (solid red triad) is achieved by rotating $\left(\phi - \frac{\Omega}{2}\right)$ around $\hat{z}'_i$. The translation is communicative, O_{i+1}, O_i is resulted from moving O_{M} by half of D_x, D_y, D_z in $\hat{x}_{\mathrm{M}}$, $\hat{y}_{\mathrm{M}}$, $\hat{z}_{\mathrm{M}}$ directions positively and negatively, respectively. Note that intermediate unit triads, $\hat{x}'$, $\hat{y}'$, $\hat{z}'$, are not in unit length for clear illustration.

3. Then, we finish the bending by restoring the hinge alignment processes. Starting from the dotted temporary frames in Figure 2.5, we take a ϕ angle turn on each basepairs about the updated z-axis.
4. Now, we twist the DNA by rotating basepairs apart from each other. This is proceeded in the same way as last step, but taking half twist angle positively

for $(i+1)^{\text{th}}$ basepair, and half twist angle negatively for i^{th} basepair, which yields total twist of Ω.

5. Finally, we translate the $(i+1)^{\text{th}}$ and i^{th} basepair by $\left[\frac{D_x}{2}, \frac{D_y}{2}, \frac{D_z}{2}\right]^T$, and $\left[-\frac{D_x}{2}, -\frac{D_y}{2}, -\frac{D_z}{2}\right]^T$ away from origin in middle frame, respectively.

As we can see, CEHS applies reversible and symmetric rotational and translational movements on both basepairs from the middle frame. It firstly bends DNA, secondly twists it and finally moves basepairs apart. This scheme ensures the correctness of most significant twist (*i.e.*, large positive angles), while putting the correlative effects into minor components, tilt and roll (*i.e.*, usually around 0°). Mathematically, CEHS represents the 3D orientations of rigid bodies in terms of *zyz*-Euler angle, and can be expressed using transformation matrix. Therefore, the procedures above are written as the equation group below [86, 87],

$$\begin{aligned}
\mathcal{T}_{i+1} &= \mathcal{R}^{\mathrm{M}}_{i+1}\mathcal{T}_{\mathrm{M}} &&= \left[\mathcal{R}^z(-\phi)\,\mathcal{R}^y\left(\frac{\Gamma}{2}\right)\mathcal{R}^z\left(\phi+\frac{\Omega}{2}\right)\right]\mathcal{T}_{\mathrm{M}} \\
\mathcal{T}_{i} &= \mathcal{R}^{\mathrm{M}}_{i}\mathcal{T}_{\mathrm{M}} &&= \left[\mathcal{R}^z(-\phi)\,\mathcal{R}^y\left(-\frac{\Gamma}{2}\right)\mathcal{R}^z\left(\phi-\frac{\Omega}{2}\right)\right]\mathcal{T}_{\mathrm{M}} \\
O_{i+1} &= \frac{1}{2}\left[D_x\hat{x}_{\mathrm{M}} + D_y\hat{y}_{\mathrm{M}} + D_z\hat{z}_{\mathrm{M}}\right] + O_M \\
O_{i} &= -\frac{1}{2}\left[D_x\hat{x}_{\mathrm{M}} + D_y\hat{y}_{\mathrm{M}} + D_z\hat{z}_{\mathrm{M}}\right] + O_M
\end{aligned} \tag{2.10}$$

, where $\mathcal{T} = [\hat{x}|\hat{y}|\hat{z}]$ is the direction cosine matrix, $\mathcal{R}$ is elementary rotation matrix here, O is the origin coordinates; their superscripts mean against particular vector or frame, and subscripts mean of particular vector or frame.

DNA structure construction processes are much easier, if it is built from one basepair to next without altering the spatial arrangements of previous dismissed basepairs. The scheme are more operative by setting the reference to i^{th} basepair, instead of the middle frame. So, we rewrite Equation 2.10, by plugging in the inverse relationship from M back to i^{th} reference frame [*i.e.*, $\mathcal{R}^i_{\mathrm{M}} = \left[\mathcal{R}^{\mathrm{M}}_i\right]^{-1} = \left[\mathcal{R}^z\left(\frac{\Omega}{2}-\phi\right)\mathcal{R}^y\left(\frac{\Gamma}{2}\right)\mathcal{R}^z(\phi)\right]$] in front of first equation, into recursive forms, as,

$$\begin{aligned}
\mathcal{T}_{i+1} &= \mathcal{R}^i_{\mathrm{M}}\mathcal{R}^{\mathrm{M}}_{i+1}\mathcal{T}_i &&= \left[\mathcal{R}^z\left(\frac{\Omega}{2}-\phi\right)\mathcal{R}^y(\Gamma)\,\mathcal{R}^z\left(\phi+\frac{\Omega}{2}\right)\right]\mathcal{T}_i \\
\mathcal{T}_{\mathrm{M}} &= \mathcal{R}^i_{\mathrm{M}}\mathcal{T}_i &&= \left[\mathcal{R}^z\left(\frac{\Omega}{2}-\phi\right)\mathcal{R}^y\left(\frac{\Gamma}{2}\right)\mathcal{R}^z(\phi)\right]\mathcal{T}_i \\
O_{i+1} &= \left[D_x\hat{x}_{\mathrm{M}} + D_y\hat{y}_{\mathrm{M}} + D_z\hat{z}_{\mathrm{M}}\right] + O_i.
\end{aligned} \tag{2.11}$$

2.4.3 Bending *B*-DNA as example

The orientation scheme described above is generally applicable to any twist-dominant helical arrangements. The values of parameters, as well as idealized coordinates within single basepairs define the final DNA structures. For different types of DNA, different building blocks and sequential basepair parameters should be used. Here is a set of parameters that represents the structure of canonical *B*-DNA, listed in Table 2.1.

The two columns of this table, list the sequential basepair and complimentary base parameters for standard *B*-form DNA. We have extensively explained the sequential basepair parameters before, which are parameters that represents the six degree of freedom between adjacent basepairs. However, we have omitted the fact that, within a basepair, the two complimentary bases attached to each backbone also orient with each other with six degree of freedom. The complimentary base parameters are six similar parameters (*i.e.*, three rotational and three translational parameters) that represent this relative spatial arrangement, as shown in Figure 2.4. Furthermore, the CEHS, used in case of basepair reference frame, is also directly applicable to these parameters, by simply reorienting the base triad, x, (-z), y-axis as x, y, z-axis. Then the orientation scheme can be mathematically expressed similar to Equation 2.11, as follow,

$$\begin{aligned}
\mathcal{T}_{i,\mathrm{I}} &= \mathcal{R}^{i,\mathrm{II}}_{\mathrm{M}'}\mathcal{R}^{\mathrm{M}'}_{i,\mathrm{I}}\mathcal{T}_{i,\mathrm{II}} &&= \left[\mathcal{R}^{z}\left(\frac{\omega}{2}-\phi'\right)\mathcal{R}^{y}\left(\gamma\right)\mathcal{R}^{z}\left(\phi'+\frac{\omega}{2}\right)\right]\mathcal{T}_{i,\mathrm{II}} \\
\mathcal{T}_{\mathrm{M}'} &= \mathcal{R}^{i,\mathrm{II}}_{\mathrm{M}'}\mathcal{T}_{i,\mathrm{II}} &&= \left[\mathcal{R}^{z}\left(\frac{\omega}{2}-\phi'\right)\mathcal{R}^{y}\left(\frac{\gamma}{2}\right)\mathcal{R}^{z}\left(\phi'\right)\right]\mathcal{T}_{i,\mathrm{II}} \\
O_{i,\mathrm{I}} &= [S_x\hat{x}_{\mathrm{M}'}+S_y\hat{y}_{\mathrm{M}'}-S_z\hat{z}_{\mathrm{M}'}]+O_{i,\mathrm{II}}
\end{aligned} \tag{2.12}$$

, where I, II indicate the base attached to particular strand, and M′ is the middle frame between two base reference frames. This middle frame is defined as basepair reference frame in the DNA analytical procedures. Due to the reorientation of triad, the "bending angle" (*i.e.*, Buckle-Opening angle, γ) is the combination of rotations in x and (-z)-axis, $\gamma=\sqrt{\kappa^2+\sigma^2}$. And, the "hinge" (*i.e.*, Buckle-Opening axis, h) inclines with (-z)-axis at $\phi'=\tan^{-1}\left(\frac{\kappa}{-\sigma}\right)$.

In practice, generalized *B*-DNA basepairs coordinates are usually utilized, which are either direct solved structures, such as the coordinates in the Landolt-

Sequential basepair para.		**Complimentary base para.**	
Tilt (τ)	$-0.1^\circ \pm 2.5^\circ$	Buckle (κ)	$0.5^\circ \pm 6.7^\circ$
Roll (ρ)	$0.6^\circ \pm 5.2^\circ$	Propeller twist (ω)	$-11.4^\circ \pm 5.3^\circ$
Twist (Ω)	$36.0^\circ \pm 6.8^\circ$	Opening (σ)	$0.6^\circ \pm 3.1^\circ$
Shift (D_x)	-0.02 ± 0.45 Å	Shear (S_x)	0.00 ± 0.21 Å
Slide (D_y)	0.23 ± 0.81 Å	Stretch (S_y)	-0.15 ± 0.12 Å
Rise (D_z)	3.32 ± 0.19 Å	Stagger (S_z)	0.09 ± 0.19 Å

Table 2.1: *B*-DNA average sequential basepair parameters are listed in the left column, with respective standard deviations. While, the second column lists that of complimentary base parameters. The data is calculated from multiple high resolution (< 2.0 Å) crystallographic structures using `X3DNA` [84].

Börnstein database [88, 89] or orientationally optimized using bases coordinates (*e.g.* by Olson *at el.* [84]). These idealized basepairs simplifies building processes; thus, complementary base parameters can be considered as fixed around the averages, as shown second column in Table 2.1. As a result, we only need to focus on the sequential basepair parameters. Now, using these building blocks and the sequential basepair parameters, we can easily build a straight *B*-DNA with any particular sequence one basepair at a time, following the CEHS. Note that the most significant sequential basepair parameters are twist and rise. As long as these two parameters are around the optimal values, while others are not far from zero, we can generate a "good" initial *B*-form DNA structures.

Nevertheless, we are not satisfied with only building the straight *B*-DNA. By introducing some periodical changes (*i.e.*, following DNA helical turns) to sequential basepair parameters, we can induce various kinds of gradual changes to DNA global conformations, such as bending or unwinding. Using bending as example, the significant twist and rise are fixed to maintain *B*-form, while the bending sensitive tilt and roll are slightly modified. Normally, we want to introduce a consistent bending, which requires a constant Roll-Tilt angle, while keeping the hinge pointing to similar directions for all basepairs. We can approximate this using following relationship,

$$\begin{aligned} \tau_i &= \Gamma_0 \sin\left((i-1)\,\Omega + \phi_0^{\mathrm{M}_1}\right) \\ \rho_i &= \Gamma_0 \cos\left((i-1)\,\Omega + \phi_0^{\mathrm{M}_1}\right) \end{aligned} \tag{2.13}$$

, where Γ_0 is the average bending angle for each basepair, and $\phi_0^{M_1}$ is approximate the inclining angle of each hinge with y-axis of the 1st middle frame.

Then, we bent the E6-94 sequence 94 bp B-DNA into planar minicircle, to mimic the looped structures in DNA looping experiments by Cloutier and Widom [3]. There are 94 basepair steps in total, including 94th to 1st closing basepair step. Using Equation 2.13, the circular shape is achieved by setting $94 \times \Gamma_0 = \pi$, and the integer number of helical turns (*i.e.*, multiple of $\sim$ 10.5 bp) ensures a proper closing orientation between the two ends. With additional minor modifications of other parameters (*i.e.*, rise and twist for better ends matching), we can get the targeted minicircle as shown below.

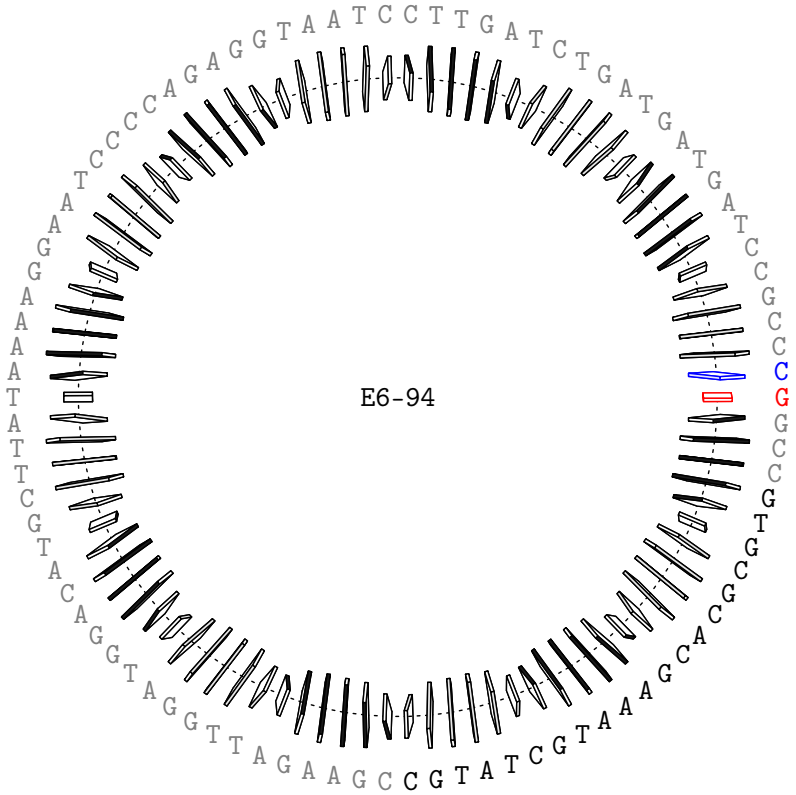


Figure 2.6: The block view of 94 bp E6-94 sequence mini circle. Each basepair is presented as a rectangle, whose long sides are parallel with y-axis of particular reference frame, short sides are parallel with its x-axis, and minor groove edges is coloured in black. The DNA, with top strand sequence labeled at outer circle, starts from 1st bp at 3-o'clock (red), runs clockwise, and ends to 94th bp (blue). The particular 20 bp partial sequence in black was used for simulations in Chapter 3, 4.

2.5 DNA conformational analysis

Once the MD simulations initiate, thousands of atoms start to bounce thermally back and forth in femtosecond timescales, thus DNA overall conformations dynamically deviate from their initials towards equilibrated states. Now, we have to quantify these structural evolutions by translating these large amount, high frequency atomic trajectories to macroscopic parameterized fluctuations, such as bending or end-to-end distance dynamics. In short, this aim is achieved by grouping highly correlated atoms, assigning representative triads and evaluating spatial arrangements. Because these DNA conformational analysis are essential in bridging simulation raw data to our results, we are going to comprehensively illustrate those processes in this section.

2.5.1 Base, basepair reference frames for fluctuating DNA

At each time step, we have an instant DNA configuration (*i.e.*, $\boldsymbol{r}^N$, where $\boldsymbol{r}$ is position vector and N is number of DNA atoms). It is intuitive to represent this DNA configuration using its centerline. Practically, we treat each Watson-Crick basepair of DNA as a block, as shown in Figure 2.6, and representing its 3D orientation using a triad originated at the basepair center. These discretized triads along DNA form the trace of DNA configuration in Cartesian coordinate. These representative triads are exactly the basepair reference frames used in last section. However, in this case, the atoms are depart from the ideal Watson-Crick basepairs used for building up DNA. Therefore, it is nontrivial to define these basepair reference frames from $\boldsymbol{r}^N$.

Extracting the overall DNA conformations from MD trajectories is a process of reducing degrees of freedom, by selecting configuration significant components, grouping highly correlated atoms, ignoring their relative motions and representing them with reference frames. Here we explicitly illustrate these procedures. Firstly, we focus on purine and pyrimidine bases, because they are pairing, stacking with each other and forming the central core of DNA. The bases are more significant in basepair centroid determinations, while, the more flexible sugar and phosphate backbone are flapping in peripheral. Moreover, for individual bases,

their structures are never deviate much from the "standard", because of relative rigid conjugated rings. As a result, we prefer to fit the standard bases [90], with preassigned base reference frames attached [84], to each observed instant base atomic arrangements by minimizing the sum of squares of their residual errors. This least-square fitting was implemented by Horn in 1987 [91] to find a closed-form solution of their absolute orientation, whose deviation is simplified by representing rotations with unit quaternions, see Appendix C for details. Then, based on two base reference frames of complimentary bases within each Watson-Crick basepair, their middle frame (*i.e.*, the half-way rotational triad between two reference frames as defined in Section 2.4) is obtained as basepair reference frame. This way, we can describe instant DNA global shape using sets of discretized triads.

2.5.2 DNA orientation analysis by CEHS

Based on sets of reference frames (each denoting a local orientation of DNA), more intuitive representations of DNA spatial arrangements, such as bending or twisting, are further calculated using CEHS. Given two reference frames, their relative spatial arrangements can described by three translational and three rotational parameters, as shown in Section 2.4. Due to the anisotropic nature of DNA, some parameters are more significant to the global conformations, while some are more sensitive to certain types of constraints. Here, we are going to conduct orientation analysis to decouple those parameters by reversibly applying the CEHS. In other words, CEHS, which is introduced before for building up relative orientation with given parameters, is suitable to assess orientation parameters with given triads, as well.

Here is the procedures to calculate the complimentary base parameters, using the two complimentary base reference frames within a Watson-Crick basepair assigned in Subsection 2.5.1. The geometries of reference frames are similar to Figure 2.5, but,

1. triads are reoriented, with original x, y, z-axis replaced by x, (-z), y-axis, in order to decouple the most significant propeller twist at first,
2. base of Strand II is at bottom, and that of Strand I is at top, following the

positive direction of y-axis.

Firstly, Buckle-Opening angle, which is the combination of rotations in x, (-z)-axis, is determined using the y-axes as,

$$\gamma = \cos^{-1}\left(\hat{y}_{i,\mathrm{II}} \cdot \hat{y}_{i,\mathrm{I}}\right) \tag{2.14}$$

, where i indicates the basepair number, and I, II denote the strand. Next, the Buckle-Opening axis is obtained using,

$$\hat{h} = \hat{y}_{i,\mathrm{II}} \times \hat{y}_{i,\mathrm{I}}. \tag{2.15}$$

Then, the two reference frames, which can be expressed using direction cosine matrices, $\mathcal{T}_{i,\mathrm{I}}$ and $\mathcal{T}_{i,\mathrm{II}}$, are rotated negative and positive half Buckle-Opening angle respectively, to coincide their y-axis,

$$\begin{aligned} \mathcal{T}'_{i,\mathrm{II}} &= \mathcal{R}_{\hat{h}}\left(\frac{\gamma}{2}\right)\mathcal{T}_{i,\mathrm{II}} \\ \mathcal{T}'_{i,\mathrm{I}} &= \mathcal{R}_{\hat{h}}\left(-\frac{\gamma}{2}\right)\mathcal{T}_{i,\mathrm{I}} \end{aligned} \tag{2.16}$$

, where $\mathcal{R}_{\hat{e}}(\theta)$ represents the rotational operation of θ angle around normalized arbitrary axis $\hat{e} = [e_x, e_y, e_z]^T$, in following explicit form,

$$\begin{bmatrix} e_x^2 + \left(e_y^2 + e_z^2\right)\cos\theta & e_x e_y\left(1-\cos\theta\right) - e_z\sin\theta & e_x e_z\left(1-\cos\theta\right) + e_y\sin\theta \\ e_x e_y\left(1-\cos\theta\right) + e_z\sin\theta & e_y^2 + \left(e_x^2 + e_z^2\right)\cos\theta & e_y e_z\left(1-\cos\theta\right) - e_x\sin\theta \\ e_x e_z\left(1-\cos\theta\right) - e_y\sin\theta & e_y e_z\left(1-\cos\theta\right) + e_x\sin\theta & e_z^2 + \left(e_x^2 + e_y^2\right)\cos\theta \end{bmatrix}. \tag{2.17}$$

The middle frame (M$'$) of two complementary base reference frames (*i.e.*, basepair reference frames) is defined as $\mathcal{T}_{\mathrm{M}'}$, which locates right between $\mathcal{T}'_{i,\mathrm{I}}$ and $\mathcal{T}'_{i,\mathrm{II}}$. It can be achieved by summing up the directional cosine matrices, then normalizing column-wise. As a result, $\mathcal{T}_{\mathrm{M}'} = [\hat{x}_{\mathrm{M}'}|\hat{y}_{\mathrm{M}'}|\hat{z}_{\mathrm{M}'}]$ originated at $O_{\mathrm{M}'}$ is,

$$\begin{aligned} \hat{y}_{\mathrm{M}'} &= \frac{\hat{y}'_{i,\mathrm{II}} + \hat{y}'_{i,\mathrm{I}}}{\|\hat{y}'_{i,\mathrm{II}} + \hat{y}'_{i,\mathrm{I}}\|} \\ O_{\mathrm{M}'} &= \frac{1}{2}\left[O_{i,\mathrm{II}} + O_{i,\mathrm{I}}\right] \end{aligned} \tag{2.18}$$

, where $\hat{x}_{\mathrm{M}'}$ and $\hat{z}_{\mathrm{M}'}$ can be calculated in similar way.

Through above procedures, the $x'z'$-planes of two transformed frames are

parallel to each other, and propeller twist is directly obtained by,

$$|\omega| = \cos^{-1}\left(\hat{x}'_{i,\mathrm{II}} \cdot \hat{x}'_{i,\mathrm{I}}\right) \tag{2.19}$$

, where $\omega > 0$, if $\left[\hat{x}'_{i,\mathrm{II}} \times \hat{x}'_{i,\mathrm{I}}\right] \cdot \hat{y}_{\mathrm{M}'} > 0$, and $\omega < 0$, if $\left[\hat{x}'_{i,\mathrm{II}} \times \hat{x}'_{i,\mathrm{I}}\right] \cdot \hat{y}_{\mathrm{M}'} < 0$. Then, the incline angle between Buckle-Opening axis and (-z)-axis is,

$$|\phi'| = \cos^{-1}\left(\hat{h} \cdot -\hat{z}'_{M}\right) \tag{2.20}$$

, where $\phi' > 0$, if $\left[\hat{h} \times \hat{z}_{\mathrm{M}'}\right] \cdot \hat{y}_{\mathrm{M}'} < 0$, and $\phi' < 0$, if $\left[\hat{h} \times \hat{z}_{\mathrm{M}'}\right] \cdot \hat{y}_{\mathrm{M}'} > 0$. Based on ϕ', the buckle and opening rotation are decoupled from the combined Buckle-Opening angle as,

$$\kappa = \gamma \sin\phi', \ \sigma = -\gamma \cos\phi'. \tag{2.21}$$

Finally, the translational parameters, shear, stretch and stagger, are separated from their linear combinations using,

$$[S_x, S_y, -S_z] = [O_{i,\mathrm{I}} - O_{i,\mathrm{II}}]^T \mathcal{T}_{\mathrm{M}'}. \tag{2.22}$$

Based on the adjacent basepair reference frames, $\mathcal{T}_{\mathrm{M}'}$ for i^{th} and $(i+1)^{\mathrm{th}}$ basepairs, their sequential basepair parameters, which describes their six degree rigid-body relative orientation, are calculated in similar manner, as below. And their reference frame geometry is shown in Figure 2.5.

1. The total bending (*i.e.*, Roll-Tilt angle) is obtained by z-axis deflection,

$$\Gamma = \cos^{-1}\left(\hat{z}_i \cdot \hat{z}_{i+1}\right) \tag{2.23}$$

   , where i stands for the basepair number.

2. The hinge axis (*i.e.*, Roll-Tilt axis), which DNA bends about, is perpendicular to both z-axles,

$$\hat{H} = \hat{z}_i \times \hat{z}_{i+1}. \tag{2.24}$$

3. The two reference frames contra-rotate about $\hat{H}$ by $\frac{\Gamma}{2}$ each to eliminate the

bending, and their transformed triads are,

$$\begin{aligned} \mathcal{T}_i' &= \mathcal{R}_{\hat{H}}\left(\frac{\Gamma}{2}\right)\mathcal{T}_i \\ \mathcal{T}_{i+1}' &= \mathcal{R}_{\hat{H}}\left(-\frac{\Gamma}{2}\right)\mathcal{T}_{i+1}. \end{aligned} \tag{2.25}$$

4. The middle frame (M) right between the two transformed basepair reference frames, $\mathcal{T}_i'$ and $\mathcal{T}_{i+1}'$, is represented by $\mathcal{T}_{\mathrm{M}} = [\hat{x}_{\mathrm{M}}|\hat{y}_{\mathrm{M}}|\hat{z}_{\mathrm{M}}]$ originated at O_{M}, which is obtained through,

$$\begin{aligned} \hat{z}_{\mathrm{M}} &= \frac{\hat{z}_i' + \hat{z}_{i+1}'}{\|\hat{z}_i' + \hat{z}_{i+1}'\|} \\ O_{\mathrm{M}} &= \frac{1}{2}\left[O_i + O_{i+1}\right] \end{aligned} \tag{2.26}$$

   , where $\hat{x}_{\mathrm{M}}$ and $\hat{y}_{\mathrm{M}}$ is averaged and normalized, too.

5. The magnitude of twist is quantified using,

$$|\Omega| = \cos^{-1}\left(\hat{y}_i' \cdot \hat{y}_{i+1}'\right) \tag{2.27}$$

   , while, its sign is determined in such way: $\Omega > 0$, if $\left[\hat{y}_i' \times \hat{y}_{i+1}'\right] \cdot \hat{z}_{\mathrm{M}} > 0$, and $\Omega < 0$, if $\left[\hat{y}_i' \times \hat{y}_{i+1}'\right] \cdot \hat{z}_{\mathrm{M}} < 0$.

6. the magnitude of incline angle between the Tilt-Roll axis and y_{M} is,

$$|\phi| = \cos^{-1}\left(\hat{H} \cdot \hat{y}_{\mathrm{M}}\right) \tag{2.28}$$

   , which is positive, when $\left[\hat{H} \times \hat{y}_{\mathrm{M}}\right] \cdot \hat{z}_{\mathrm{M}} > 0$; and is negative, when $\left[\hat{H} \times \hat{y}_{\mathrm{M}}\right] \cdot \hat{z}_{\mathrm{M}} < 0$.

7. The ϕ indicates the relative contributions of tilt and roll to total bending, which isolates the correlated rotations in x and y-axis in following way,

$$\tau = \Gamma \sin\phi, \;\; \rho = \Gamma \cos\phi. \tag{2.29}$$

8. The independent, shift, slide and rise, are directly obtained by projecting the displacement of basepair reference frames onto middle frame triad, as,

$$[D_x, D_y, D_z] = [O_{i+1} - O_i]^T \mathcal{T}_{\mathrm{M}}. \tag{2.30}$$

2.6 Advanced sampling

For non-interactive, thermal equilibrated system connected with a heat bath, the probability density function of particular configuration follows the Boltzmann distribution,

$$\rho(\boldsymbol{r}^N) \propto \exp\left(-\beta\mathcal{H}(\boldsymbol{r}^N)\right) \tag{2.31}$$

, where $\mathcal{H}$ is the Hamiltonian, which in general includes internal energy and external energy. Then, most system properties of interest can be achieved through corresponding ensemble averages, for instant, the likelihood of system along particular reaction coordinate, ξ,

$$\rho(\xi) = \frac{\int \delta\left(\xi(\boldsymbol{r}^N) - \xi\right) \exp\left(-\beta\mathcal{H}(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}}{\int \exp\left(-\beta\mathcal{H}(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}} = \left\langle \delta\left(\xi(\boldsymbol{r}^N) - \xi\right)\right\rangle \tag{2.32}$$

, where δ is the Dirac delta function. In this thesis, we are focusing on configuration space only by ignoring the momentum term, but the formulas are extensible to entire phase space.

By assuming the ergodicity of system, the ensemble average is the same as time average, for example, Equation 2.32 is equivalent to,

$$\rho(\xi) = \lim_{\tau\to\infty} \frac{1}{\tau} \int_{t=0}^{\tau} \rho\left(\xi(t)\right) dt\,. \tag{2.33}$$

As a result, it is possible to study the system properties by directly populating trajectories using MD simulations. However, an exhausted sampling is never possible for typical biological systems. The sampling process is limited by computational power and high dimensionality. Furthermore, the system is usually trapped in some local energy minimums, blocked by nearby energy barriers within its frustrated energy landscape. While, the rare events, where targeted properties may reside in, are commonly insufficiently explored. So the techniques, which accelerate the sampling process and enlarge the sampling areas, are required. Here we focus on umbrella sampling technique, which is extensively used in Chapter 4 and 5 to inspect DNA micromechnical properties.

2.6.1 Umbrella sampling

Statistically, umbrella sampling [92] is an reweighting sampling method, which weights particular events more to enhance the sampling at targeted regions of intrinsic distributions; physically, it applies a biased potential to constraint the system near certain configurations. As a result, it significantly improve the efficiency of energy landscape explorations. Here we are going to derive the central relationship, which restores the Boltzmann distribution from simulated non-physical statistics.

The generalized Hamiltonian with biased potentials is in form of,

$$\mathcal{H}_{\{\lambda\}}(\boldsymbol{r}^N) = \lambda_0 \mathcal{H}_0(\boldsymbol{r}^N) + \sum_{i=1}^{W} \lambda_i \mathcal{V}_i(\boldsymbol{r}^N) \tag{2.34}$$

, where $\mathcal{H}_0$ is the intrinsic Hamiltonian of system, $\mathcal{V}_i$, for $i = 1, 2, \cdots, W$ are generalized restrain potentials (*i.e.*, external energy). While, $\{\lambda\} = \{\lambda_0, \lambda_1, \cdots, \lambda_W\}$ is the coupling parameter set, which indicates the applied potentials, (*e.g.*, $\{0\}$ represents the unbiased case, where $\lambda_0 = 1$ and others are zeros. Note that λ_0 is always unity.)

$$\begin{aligned}
&\rho_{\{\lambda\},\beta}(\xi) \\
&\quad = \frac{\int \delta\left(\xi(\boldsymbol{r}^N) - \xi\right) \exp\left(-\beta \mathcal{H}_{\{\lambda\}}(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}}{\int \exp\left(-\beta \mathcal{H}_{\{\lambda\}}(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}} \\
&\quad = \frac{\mathcal{Z}_{\{0\},\beta}}{\mathcal{Z}_{\{\lambda\},\beta}} \cdot \frac{\int \delta\left(\xi(\boldsymbol{r}^N) - \xi\right) \exp\left(-\beta \mathcal{H}_0(\boldsymbol{r}^N)\right) \prod_{i=1}^{W} \exp\left(-\beta \lambda_i \mathcal{V}_i\left(\xi(\boldsymbol{r}^N)\right)\right) d^N\boldsymbol{r}}{\int \exp\left(-\beta \mathcal{H}_0(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}} \\
&\quad = \left(\prod_{i=1}^{W} \exp\left(-\beta \lambda_i \mathcal{V}_i(\xi)\right)\right) \cdot \frac{\mathcal{Z}_{\{0\},\beta}}{\mathcal{Z}_{\{\lambda\},\beta}} \cdot \frac{\int \delta\left(\xi(\boldsymbol{r}^N) - \xi\right) \exp\left(-\beta \mathcal{H}_0(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}}{\int \exp\left(-\beta \mathcal{H}_0(\boldsymbol{r}^N)\right) d^N\boldsymbol{r}} \\
&\quad = \left(\prod_{i=1}^{W} \exp\left(-\beta \lambda_i \mathcal{V}_i(\xi)\right)\right) \cdot \frac{\mathcal{Z}_{\{0\},\beta}}{\mathcal{Z}_{\{\lambda\},\beta}} \cdot \rho_{\{0\},\beta}(\xi)\,.
\end{aligned} \tag{2.35}$$

The Boltzmann distribution is usually represented against particular reaction coordinate, ξ. So, the restrain potentials are selected to forcibly sampling

along ξ, where the functions of ξ only are commonly used, $\mathcal{V}_i(\boldsymbol{r}^N) = \mathcal{V}_i\left(\xi(\boldsymbol{r}^N)\right)$, for $i = 1, 2, \cdots, W$. Now, the probability density function at particular reaction coordinate from biased sampling can be expressed as ensemble average (Equation 2.32). Then it links to that from unbiased sampling, following above Equation 2.35, where $\mathcal{Z}$ is the partition function of particular system.

Finally, the unbiased generalized free energy, $\mathcal{F} = -k_{\mathrm{B}}T \ln \mathcal{Z}$, is analytical expressed by,

$$\mathcal{F}_{\{0\},\beta}(\xi) = -\beta^{-1} \ln\left(\rho_{\{\lambda\},\beta}(\xi)\right) - \sum_{i=1}^{W} \lambda_i \mathcal{V}_i(\xi) + \mathcal{F}_{\{\lambda\},\beta} \tag{2.36}$$

, with a weight factor $\mathcal{F}_{\{\lambda\},\beta} = -\beta^{-1} \ln \mathcal{Z}_{\{\lambda\},\beta}$, which is dependent of coupling parameter set, and is independent of reaction coordinate. Illustratively, corrected free energy curves obtained (*i.e.*, subtracting applied potentials) from differently constraint simulations under same temperature offset each other by a constant. $\mathcal{F}_{\{\lambda\}_i,\beta_i}$ bring them together, which can be estimated using methods, such as Weighted Histogram Analysis Method (WHAM).

2.6.2 Weighted Histogram Analysis Method

WHAM [93, 94] utilizes all the information obtained during biased sampling processes to retrieve the underlining unbiased Boltzmann distribution. The key is to estimate the relative weights among different attempts. More explicitly, $\mathcal{F}_{\{\lambda\},\beta}$ are iteratively evaluated to minimize the variations of intrinsic number of microstates, Ω. Here we are going to derive the recursive equations that combines all the statistics to reconstruct the free energy landscape.

The Hamiltonian is further expressed into $\mathcal{H}_{\{\lambda\}} = \sum_{i=0}^{W} \lambda_i \mathcal{V}_i$. During the j^{th} run, the bin size scaled number of microstates around generalized coordinate is estimated by,

$$\tilde{\Omega}_j\left(\{\mathcal{V}\}, \xi\right) = n_j\left(\{\mathcal{V}\}, \xi\right) \exp\left(\left(\sum_{i=0}^{W} \beta_j \lambda_{i,j} \mathcal{V}_i\right) - \beta_j \tilde{\mathcal{F}}_j\right) \tag{2.37}$$

, where n_j is the occurrence within particular histogram, and $\tilde{\mathcal{F}}_j$ (*i.e.*, the short form of $\tilde{\mathcal{F}}_{\{\lambda\}_j,\beta_j}$) is the free energy of j^{th} system. The optimized $\tilde{\Omega}\left(\{\mathcal{V}\}, \xi\right)$ from

total of R runs is achieved by tuning a set of direct weighting probabilities, w_i, for $i = 1, 2, \cdots, R$, as,

$$\tilde{\Omega}\left(\{\mathcal{V}\}, \xi\right) = \sum_{j=1}^{R} w_j\left(\{\mathcal{V}\}, \xi\right) \tilde{\Omega}_j\left(\{\mathcal{V}\}, \xi\right). \tag{2.38}$$

, where $\sum_{j=1}^{R} w_j\left(\{\mathcal{V}\}, \xi\right) = 1$.

By minimizing the statistical error of $\delta^2\tilde{\Omega}\left(\{\mathcal{V}\}, \xi\right)$ against w_j, we can solve w_j and yield the recursive WHAM equations,

$$\begin{aligned}
\tilde{p}_{\{\lambda\},\beta}\left(\{\mathcal{V}\}, \xi\right) &= \frac{\sum_{k=1}^{R} g_k^{-1} n_k\left(\{\mathcal{V}\}, \xi\right) \exp\left(-\beta \sum_{i=0}^{W} \lambda_i \mathcal{V}_i\right)}{\sum_{m=1}^{R} N_m g_m^{-1} \exp\left(-\sum_{i=0}^{W} \beta_m \lambda_{i,m} \mathcal{V}_i + \beta_m \tilde{\mathcal{F}}_m\right)} \\
\exp\left(-\beta_j \tilde{\mathcal{F}}_j\right) &= \sum_{\{\mathcal{V}\},\xi} \tilde{p}_{\{\lambda\}_j,\beta_j}\left(\{\mathcal{V}\}, \xi\right)
\end{aligned} \tag{2.39}$$

, where $\tilde{p}_{\{\lambda\},\beta}\left(\{\mathcal{V}\}, \xi\right)$ is the constrained probability around generalized coordinate, N_m is the total count of events in m^{th} run, and $g_m = 1 + 2\tau_m$, τ_m is the integrated correlation time. When all restrain potentials are functions of reaction coordinates ξ only, [*e.g.* the harmonic potentials are frequently used $\mathcal{V}_i\left(\xi\left(\boldsymbol{r}^N\right)\right) = \frac{\kappa_i}{2}\left(\xi\left(\boldsymbol{r}^N\right) - \xi_i\right)^2$, $i = 1, 2, \cdots, W$], and all simulations are conducted under same $T = \left(k_{\mathrm{B}}\beta\right)^{-1}$, the above equations can be significantly reduced from $(W + 2)$ to 1 dimensions as following,

$$\begin{aligned}
\tilde{p}_{\{\lambda\},\beta}(\xi) &= \frac{\exp\left(-\beta \sum_{i=1}^{W} \lambda_i \mathcal{V}_i(\xi)\right) \sum_{k=1}^{R} g_k^{-1} n_k(\xi)}{\sum_{m=1}^{R} N_m g_m^{-1} \exp\left(-\beta \sum_{i=1}^{W} \lambda_{i,m} \mathcal{V}_i(\xi) + \beta \tilde{\mathcal{F}}_m\right)} \\
\exp\left(-\beta \tilde{\mathcal{F}}_j\right) &= \sum_{\xi} \tilde{p}_{\{\lambda\}_j,\beta_j}(\xi).
\end{aligned} \tag{2.40}$$

Lastly, the unbiased free energy difference profile, $\Delta\mathcal{F}(\xi)$, relate to a reference state (*i.e.*, global energy minimum state is used in this thesis) with reaction coordinate, ξ_0, is readily evaluated from the non-constraint probabilities as below,

$$\Delta\mathcal{F}_{\{0\},\beta}(\xi) = -\beta^{-1} \ln\left(\frac{\tilde{p}_{\{0\},\beta}(\xi)}{\tilde{p}_{\{0\},\beta}(\xi_0)}\right). \tag{2.41}$$

Chapter 3

DNA defects induced by strong bending

3.1 Introduction

As we mentioned in Chapter 1, the knowledge of DNA homogeneity breakages under constraints are critical to the understanding of DNA compaction biologically, and the refinements of WLC polymer model physically. Although the DNA constitutive failures (*i.e.*, defects) have been proposed and theoretically studied before [6, 30, 5], direct and systematic investigations of its atomic nature and intrinsic properties are still lacking.

Hereby, in this chapter, we are focusing on answering the questions associated with the occurrences of defects within *B*-form DNA under bending constraints in atomic-level resolutions. In short, we approached these by applying the bending constraints through connecting a contractile spring to the two ends of a smoothly bent DNA initial. By tuning up the spring constant (κ), we altered the strengths of bending from weak to strong, then, obtained, analyzed and compared their full-atom dynamical behaviours using MD simulations.

From these results, we observed the distinctive bending behaviours under strong bending conditions, where localized sharp bends (*i.e.*, kinks) present at the middle of DNA fragments, against the "expected" uniform smooth bending under weak bending constraints. Zooming in at their atomic details revealed the existences of defects, in form of hydrogen bonding and basepair stacking disrup-

tions. Moreover, the spatial, temporal correlations between basepair disruptions and kink formations indicate that the defects induced by strong bending directly lead to localized sharp bends, and in turn relax the initiating bending constraints.

Furthermore, we statistically summarized that the locations of defects always occur at the AT-rich center of our sequence. Replacing this middle region with higher GC content sequence do not significantly influence the defect central localizations, which means sequences only have minor effects on defect generations, while bending geometry is the dominate factor.

3.2 Unconstrained MD to simulate classical *B*-form DNA

In order to justify the simulation methods, as well as to obtain the control dynamics of classic *B*-DNA, an unconstrained MD simulation has been conducted as benchmark on the 20 bp sequence specific DNA fragment. This particular sequence is extracted from the 94 bp high affinity nucleosome positioning sequence, E6-94, used in the cyclization experiments by Cloutier and Widom [3], as listed bellow,

$$\begin{aligned} &5' - \texttt{GTGCGCACGAAATGCTATGC} - 3' \\ &3' - \texttt{CACGCGTGCTTTACGATACG} - 5'. \end{aligned} \tag{3.1}$$

The positions of basepairs, counted from the $5'$ end of the top strand as 1 to $3'$ end as 20, were indexed by i. The detailed location of above sequence in E6-94 is shown in Figure 2.6.

This simulation was prepared and ran using the latest GROMACS package (version 4.5) [95, 96] under the newest Parm99 force field with ParmBSC0 corrections [80, 83]. Before starting the MD simulation, a basic simulation unit (*i.e.*, unit cell) was properly generated. Firstly, a straight 20 bp *B*-form DNA was produced as initial following Section 2.4 with the help of X3DNA [97]. Secondly, this initial structure was centered within a cuboid, where its first principle axis parallels to longest edge (whose length is DNA length plus twice specific buffer length, $d_{\mathrm{b}} = 1.6$ nm), and second principle axis parallels with second longest edge (whose length is DNA diameter plus twice the buffer length). Next, this unit cell

was further prepared by filling the cuboid with TIP3P water [98], neutralizing the negative charges on DNA using sodium counter-ions, and replacing water molecules by sodium chloride to achieve 150 mM ionic strength. Lastly, it was finalized by energy minimization using the steepest descent method to remove any energy unfavourable close contacts, then by thermolization using 200 ps velocity rescaling and 200 ps Parrinello-Rahman pressure coupling simulations to adjust its temperature and volume [99, 100]. Based on such prepared unit cell, a 70 ns MD simulation, without any constraints applied to DNA, was executed using periodic boundary conditions, under NVT ensemble, with constant temperature of 300 K and volume of ~ 288 nm^3.

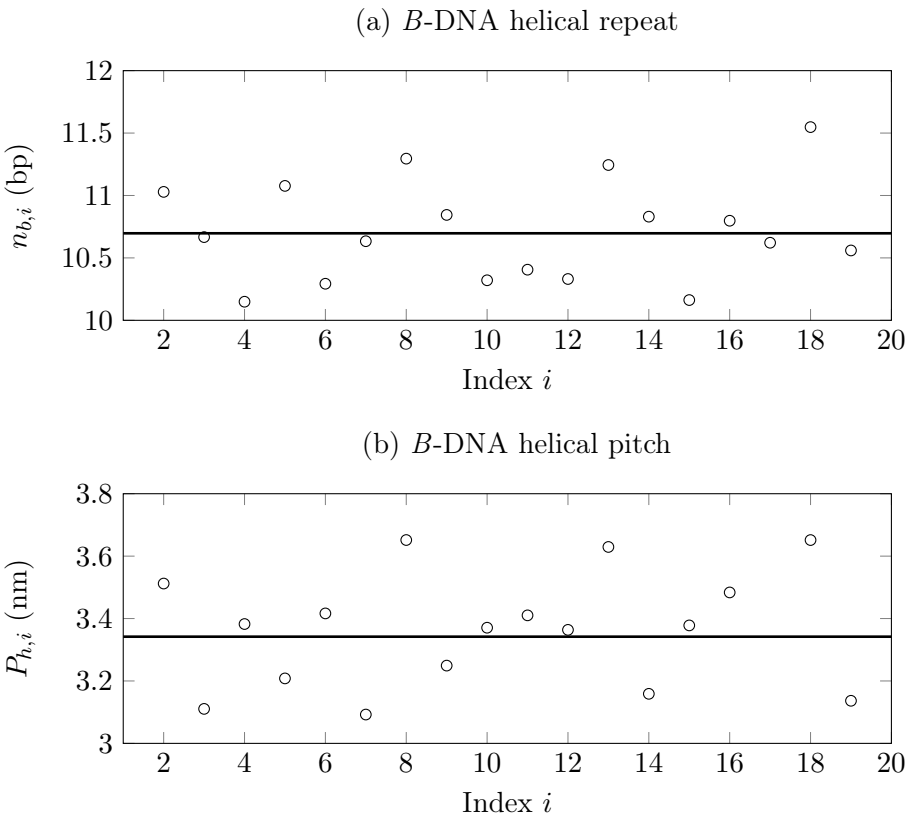


Figure 3.1: Helical repeat, $n_{b,i}$ (a) and helical pitch, $P_{h,i}$ (b) along DNA are derived using average twist and rise at particular site i over the last 20 ns of 70 ns simulation. Black lines show their global mean obtained by $n_b = \frac{2\pi}{\langle\Omega\rangle}$ and $P_h = \frac{2\pi\langle D_z\rangle}{\langle\Omega\rangle}$ at 10.70 ± 0.07 bp and 3.34 ± 0.03 nm, respectively, where Ω is twist, D_z is rise per basepair step, and the values after $\pm$ sign are corresponding standard errors calculated from uncorrelated structure representatives. $n_{b,i}$ and $P_{h,i}$ are all around their global mean, which indicates the homogeneity of DNA.

The DNA conformations were collected per 1000 time steps (*i.e.*, sampling time interval $\Delta t = 1000 \times 2$ fs $= 2$ ps) from the last 20 ns out of the 70 ns

simulation trajectories. These structural representatives are considered as equilibrated ensembles, that represent the current stable state of DNA (*i.e.*, in this case, normal *B*-DNA energy minimum state). This can be justified from stabilized structural and energy dynamics, such as RMSD or total potential energy fluctuations.

In order to verify that we achieved the targeted conformational state of *B*-DNA, we, firstly, analyzed their sequential basepair parameters following Section 2.5, based on those last 20 ns structural representatives. Then, we further calculated their helical parameters both globally and locally, including helical repeat ($n_b = \frac{2\pi}{\langle\Omega\rangle}$) and helical pitch ($P_h = \frac{2\pi\langle D_z\rangle}{\langle\Omega\rangle}$, where Ω is twist, and D_z is rise). The results show that DNA, in current MD simulation, assumed a regular straight helical structure, with $n_b = 10.70 \pm 1.53$ bp and $P_h = 3.34 \pm 0.67$ nm, that resembles the experimentally obtained classical structure, (where the values after $\pm$ sign are corresponding standard deviations, while their standard errors are less than 1%; refer to Figure 3.1 for more details). Thus, we conclude that an unconstrained benchmark of *B*-DNA has been obtained as control, through a valid MD simulation method with appropriate force fields.

3.3 Basepair disruptions at sharp bends induced by strong springs

3.3.1 Simulation methods with bending constraints

In order to apply the bending constraints, a different unit cell was set up, with compressional force added to the two ends of DNA. A smoothly bent conformation of 20 bp DNA was built, using the same sequence, as shown in Figure 3.2. This curved initial speeds up the simulation, by omitting the transitions from compression to bending, when compressional force is applied. Next, it was centered in rhombic dodecahedron unit cell, with its inscribed sphere diameter equals to the largest DNA extension observed in control simulation, plus twice the buffer length. This near spherical unit cell can tessellate space by translating itself, hold any minor deformed DNA conformations, allow free rotational movements of DNA, and reduce computational costs (*i.e.*, $\sim 71\%$ of cubic unit

cell volume). Then, this simulation unit was finalized exactly following the same procedures mentioned in last section.

The compressional force was realized by connecting DNA terminal ends with a zero length contractile spring. Its spring constant κ is the adjustable parameter (in unit of pN/nm), where larger value of κ corresponds to stronger compressional force. The spring attaches to the second and the second-last basepairs of the DNA fragment, and the spring force is distributed among their base atoms according to atomic weights. More analytically, a harmonic potential, $\mathcal{V} = \frac{\kappa}{2}d^2$ is directly added between the atom-mass weighted centers of their purine and pyrimidine bases. After these, the constrained MD simulations were conducted under NVT ensembles, with constant temperature of 300 K and volume of ~ 1170 nm^3. The conformation evolutions under different bending constraints were obtained by tuning κ in respective MD simulations.

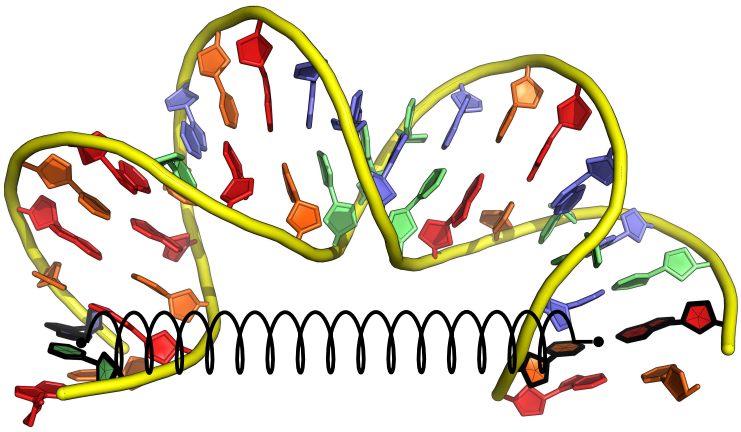


Figure 3.2: Initial smoothly bent DNA conformation generated by `X3DNA`. This initial conformation has an overall bending angle of $\sim 160°$. A zero-length spring is connected to the bases of second and second-last basepairs (highlighted by black outlines) to actively pull the DNA ends inward. Note that the nucleotides are coloured by sequence, `A` in blue, `T` in green, `G` in red and `C` in orange, while backbones are coloured in yellow, in all snapshots across this entire thesis.

3.3.2 Distinctive behaviours under different bending

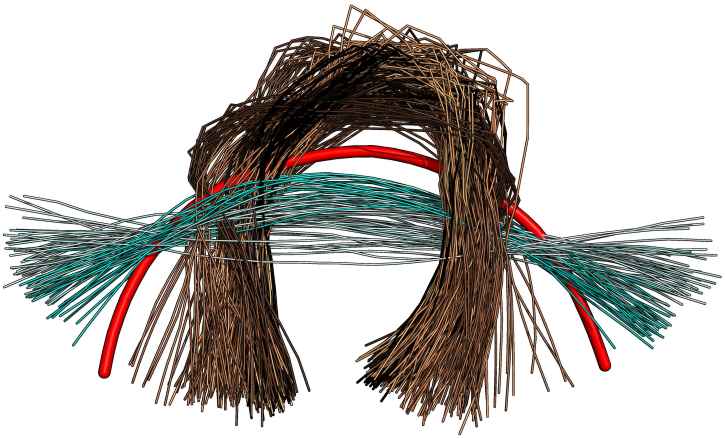
Multiple conformational trajectories under different bending constraints have been extracted from multiple MD simulations, which were conducted following

the instructions in last section with various κ ranging from 8 to 85 pN/nm. Two types of distinctive bending behaviours were observed, through direct visualizations, as well as monitoring DNA end-to-end distances. Succinctly, under weak bending, explicitly when $\kappa < 20.0$ pN/nm, DNA fragments are slightly, uniformly bent, and the degrees of bending increases as κ increases. While, under strong bending, explicitly when $\kappa > 25.0$ pN/nm, DNA fragments are severely, unevenly bent, with their two ends physically colliding into each other.

An optimal, global helical axis of DNA is a smooth curve, which roughly coincides with the centerline of DNA fragment and briefly reflects its overall conformation. The helical axis for each DNA conformation in MD trajectories has been determined using `Curves+` algorithm [101, 102]. For clearer visualizations, 20 helical axes of DNA structures at $t = 51, 52, \cdots, 70$ ns where chosen to form a representative ensemble for each individual MD simulation. All these 14 representative ensembles (*i.e.*, 280 helical axes in total) were superimposed and drew on Figure 3.3(a), which are coloured from light to dark as κ increases. Here we use cyan for cases under weak bending; and use copper for cases under strong bending. All 40 helical axes from two simulations with $\kappa < 20.0$ pN/nm are slightly bent and more extend, which indicates that DNA straightened out from its initial structure during their conformational evolutions. On the other hand, all 240 helical axes from twelve simulations with $\kappa > 25.0$ pN/nm are severely bent, and more or less overlap with each other, regardless of their κ values. Compared with the red helical axis of the initial, we can see that the terminals of these 240 DNA fragments were pulled closer by contractile springs during simulations. More importantly, the cyan DNA fragments are homogeneously bent, in contrast, the copper ones cannot sustain their homogeneities, resulting in sharp bends in the middle region, and slight bends in their two arms.

The end-to-end distance d is defined as distance between center-of-mass of the terminal atom groups (*i.e.*, atoms of purine and pyrimidine bases in 2^{nd} and 19^{th} basepairs), which is a natural indicator of bending. The mean end-to-end distance of equilibrated DNA fragment under various κ was calculated by averaging over the $10,000$ conformations from last 20 ns for each simulation. Under weak bending conditions, the mean end-to-end distances are in the range

(a) Distinctive helical axis ensembles under different bending



(b) Distinctive end-to-end distances under different bending

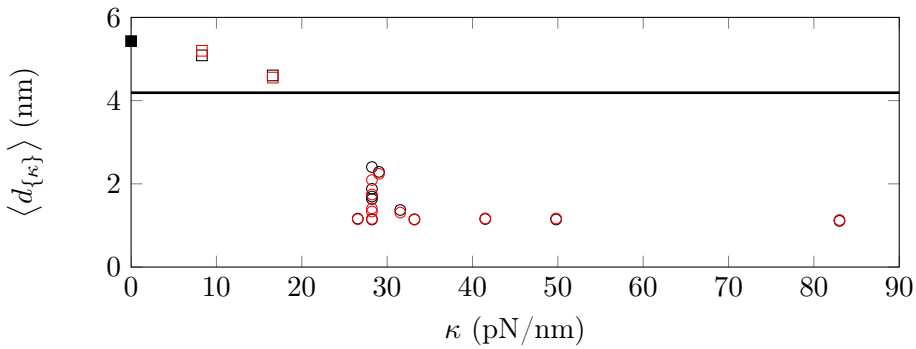


Figure 3.3: Overview of distinctive DNA bending behaviours under weak and strong bending. (a) Superimpositions of equilibrated DNA helical axes collected per ns in last 20 ns for each simulation. The fourteen independent MD simulations were all initiated from 70 ns same initial (represented by thick red helical axis), and their corresponding stabilized "centerlines" were coloured cyan for weak spring constants κ = 0.0, 8.3, 16.6 pN/nm; and copper for strong spring constants κ = 26.6, 28.2_{I}, 28.2_{II}, 28.2_{III}, 28.2_{IV}, 28.2_{V}, 29.0, 31.5, 33.2, 41.5, 49.8, 83.0 pN/nm, respectively. Note that there are five independent simulations for κ = 28.2 pN/nm. When $\kappa < 20.0$ pN/nm, their center lines are uniformly bent and more straight than initial. While, when $\kappa > 25$ pN/nm, their center lines are non-uniformly bent and much more curved than initial. (b) Mean end-to-end distances $\langle d_{\{\kappa\}} \rangle$ under various κ averaged over 50 to 70 ns for each simulation. $\langle d_{\{0\}} \rangle$ from the unconstrained simulation (■) is shown as control. $\langle d_{\{\kappa\}} \rangle$ with $\kappa < 20.0$ pN/nm (□) are longer than d_{ini} (red line), shorted than control, and negatively correlated with κ. $\langle d_{\{\kappa\}} \rangle$ with $\kappa > 25.0$ pN/nm (○) are much shorter than d_{ini} and uncorrelated with κ. Further evolutions of d were obtained till 100 ns for comparison purpose, whose $\langle d_{\{\kappa\}} \rangle$ with $\kappa < 20.0$ pN/nm (□) and with $\kappa > 25.0$ pN/nm (○) over 80 to 100 ns remain similar.

of $4.6 < \langle d_{\{\kappa\}} \rangle < 5.2$ nm, when spring constants of $8.3 \leq \kappa \leq 16.6$ pN/nm were used. These $\langle d_{\{\kappa\}} \rangle$ are slightly larger than the end-to-end distance of initial, $d_{\text{ini}} \approx 4.2$ nm, but smaller than that obtained from unconstrained simulation (*i.e.*, $\kappa = 0$ pN/nm, denoted by $\{0\}$), $\langle d_{\{0\}} \rangle \approx 5.43$ nm. The trend, $\langle d_{\{16.6\}} \rangle < \langle d_{\{8.3\}} \rangle < \langle d_{\{0\}} \rangle$, follows typical semi-flexible polymer bending responses, DNA is deformed more when bent harder. Under strong bending conditions, DNA mean end-to-end distances are $1.0 < \langle d_{\{\kappa\}} \rangle < 2.5$ nm under spring constants $26.2 \leq \kappa \leq 83.0$ pN/nm, which are much smaller than d_{ini}, and insensitive to the change of κ. Considering the ~ 2 nm DNA width and volume exclusion effects, these distances mean that the DNA ends physically collide into each other.

These dynamics of end-to-end distances under different bending constraints have reached steady states within our 70 ns simulation, where drastic changes of d from d_{ini} were usually finished within first 20 ns. Due to the gap between MD simulation and experimental timescales, we further extended these 14 independent simulations till 100 ns to test the equilibration of conformational evolutions. This time length is approaching the testing time limit for ParmBSC0 force field [83], while its accumulated error becomes more significant. As shown in Figure 3.3(b), the resultant $\langle d_{\{\kappa\}} \rangle$ averaged over 80 to 100 ns (red data) are more or less the same as those obtained over 50 to 70 ns (black data). This indicates that end-to-end distances have been properly equilibrated under different κ, while the observed distinctive trends under weak and strong bending constrains are time invariant.

3.3.3 Hydrogen bonding and base stacking disruptions induced by strong springs

The severely bent DNA fragments observed under strong bending conditions in last section, are unexpected from WLC polymer model point of view, and can induce DNA further compaction *in vivo*. The questions, then, naturally arise: what are the structural differences that lead to these distinctive bending? If the *B*-DNA still intact under weak bending constraints, are there defects induced by strong springs? If yes, in what forms? In this section, We are going to approach these questions by zooming into the detailed atomic structures of those

severely bent DNA. Their basepair integrities were closely examined through analyzing the two major non-covalent interactions, horizontal hydrogen bonding and vertical basepair stacking.

Basepair integrity checks on intact DNA under weak bending

As control, the "standard" hydrogen bonding and basepair stacking profiles for *B*-DNA were extracted from the unconstrained simulation. Next, the same methods were applied to simulations under weak bending conditions to confirm their basepair integrities.

To quantify the horizontal hydrogen bonding, in each basepair the inter-distances of atoms involved in hydrogen bonds formations were extracted, as $h_{i,j}$, where i denotes the basepair index and j denotes the j^{th} hydrogen bond in the particular basepair. Note, $j = 1, 2$ for `A=T` and $j = 1, 2, 3$ for `G≡C`. And the minimal and maximal values of $h_{i,j}$ within particular basepair were calculated for each time step, and their equilibrated values over the last 20 ns, $\langle \min(h_{i,j}) \rangle$ and $\langle \max(h_{i,j}) \rangle$ against i were obtained as the hydrogen bonding profile. The profile obtained from unconstrained *B*-DNA simulation [red lines in Figure 3.4(a), with corresponding standard deviations as error bars] shows uniform close associations among all pairing bases, where 95% confidence intervals of $\min(h_{i,j})$ and $\max(h_{i,j})$ are 0.196 ± 0.025 and 0.361 ± 0.056 nm, respectively, regardless of the sequences. Hereafter, a basepair is considered as a Watson-Crick basepair when all of its hydrogen bonds are within the optimal ranges, more explicitly, when $0.17 < \min(h_{i,j}) < 0.23$ nm and $0.30 < \max(h_{i,j}) < 0.42$ nm.

The vertical basepair stacking can be quantified through their overlapping areas, which can be calculated using `X3DNA` by projecting the i^{th} and $(i+1)^{\text{th}}$ basepair into their middle frame (for $i = 2, 3, \cdots, N-2$, where N is the number of basepairs), under the prerequisite that their inter-distance is not further than 4.5 Å. Otherwise, they are considered as totally unstacked. So, for each DNA conformation, the stacking areas within individual strands, $S^{\text{I}}_{i,i+1}$ [*i.e.*, between i and $(i+1)$ bases in Strand I] and $S^{\text{II}}_{i,i+1}$ [*i.e.*, between $(2N-i)$ and $(2N-i+1)$ bases in Strand II], were obtained through strand-wise base-base overlapping areas, while the cross-strand overlaps were ignored due to their minor stacking

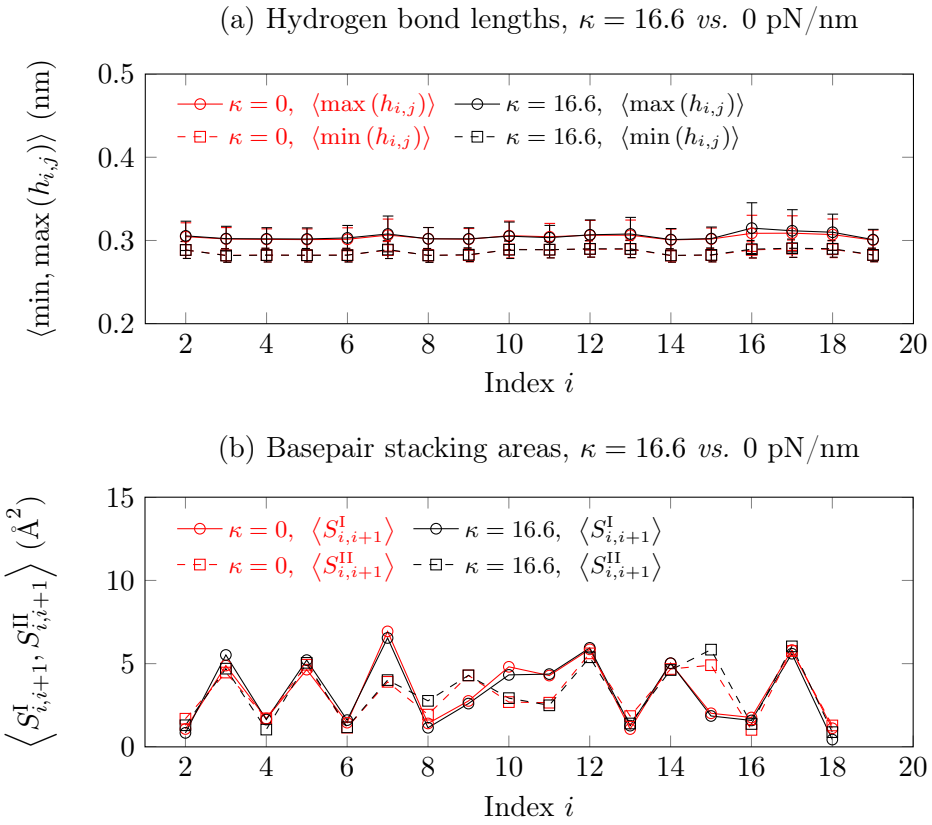


Figure 3.4: Basepair integrity analysis under weak bending constraints: (a) The hydrogen bonding profile, $\langle \min, \max(h_{i,j}) \rangle$ *vs.* $i = 2, 3, \cdots, 19$ averaged over the last 20 out of 70 ns simulation with $\kappa = 16.6$ pN/nm (black lines) against that of unconstrained DNA ($\kappa = 0$ pN/nm, red lines). (b) The basepair stacking profile, $\left\langle S^{\mathrm{I}}_{i,i+1}, S^{\mathrm{II}}_{i,i+1} \right\rangle$ *vs.* $i = 2, 3, \cdots, 18$ averaged over the last 20 ns of the same simulation (black lines) against that of control as well (red lines). These basepair integrity profiles under $\kappa = 16.6$ pN/nm coincide with those of control, which reveals the undisturbed non-covalent interactions, specifically, hydrogen bonding and basepair stacking, inside DNA fragments constrained by weak contractile springs.

effects. For the unstacked bases, their $S^{\mathrm{I}}_{i,i+1}$ or $S^{\mathrm{II}}_{i,i+1}$ were set to be 0 Å^2. Then, the average values over the last 20 ns, $\left\langle S^{\mathrm{I}}_{i,i+1} \right\rangle$ and $\left\langle S^{\mathrm{II}}_{i,i+1} \right\rangle$ against the basepire index i were acquired as the basepair stacking profile. The control profile from unconstrained simulation [red lines in Figure 3.4(b)] presents a wide range of stacking areas with large fluctuations from 1 up to 7 Å^2 in sequence dependent zigzag manner. However, none of the basepair is unstacked.

After obtaining the non-covalent interaction profiles for the control simulation, we considered the cases under weak bending, and selected the $\kappa = 16.6$ pN/nm simulation for demonstration. Its hydrogen bonding and basepair stacking profiles completely overlap with those of control as shown by black lines in

Figure 3.4. It suggests that the two major non-covalent interactions are not altered at all when $\kappa < 20$ pN/nm. In other words, the DNA fragments under weak bending constraints are still intact B-form DNA.

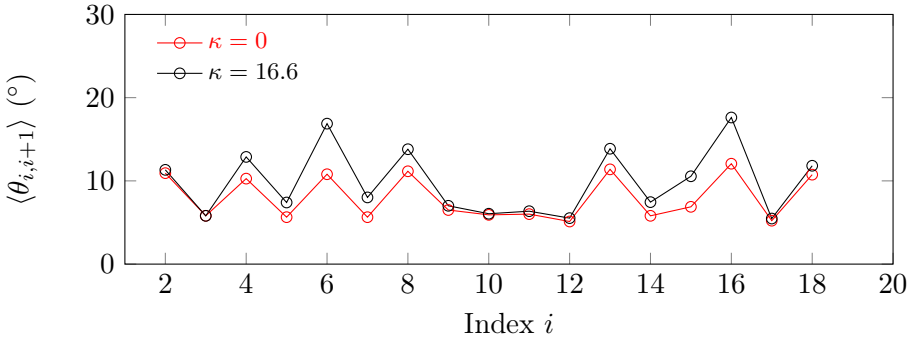


Figure 3.5: Local deformation analysis under weak bending constraints: the bending angular scan profile, $\langle\theta_{i,i+1}\rangle$ *vs.* $i = 2, 3, \cdots, 18$ averaged over the last 20 ns under $\kappa = 16.6$ pN/nm (black line) against that under $\kappa = 0$ pN/nm (red line). The overall larger bending angles along the whole DNA compared with those of unconstrained straight DNA give rise to a consistent and homogeneous directional bending.

In order to further analyze the local bending deformations of intact DNA, its bending angular scan over particular contour length along whole DNA was conducted. First of all, a right-handed standard reference frame [84] was attached to each Watson-Crick basepair (Figure 2.3), with $\hat{x}_i$ pointing to the major groove, $\hat{y}_i$ pointing to the backbone of Strand I, and $\hat{z}_i = \hat{x}_i \times \hat{y}_i$ describing the normal direction of each basepair, where i denotes the basepair index. Then, their stepwise bending angles between i^{th} and $(i + \Delta)^{\text{th}}$ basepair was measured by,

$$\theta_{i,i+\Delta} = \cos^{-1}\left(\hat{z}_i \cdot \hat{z}_{i+\Delta}\right) \tag{3.2}$$

, where $i = 2, 3, \cdots, 19-\Delta$. The Δ defines the contour length of separation in the unit of bp, and here, we set $\Delta = 1$ to check the very local bending deformations between every adjacent basepairs. Note that the normal direction is only defined on Watson-Crick basepair. As a result, this bending angular scan method is not directly applicable to DNA containing non Watson-Crick basepair.

Figure 3.5 shows the very local bending angular scan profile of unconstrained DNA, $\langle\theta_{i,i+1}\rangle$, for $i = 2, 3, \cdots, 18$ (red line) averaged over the last 20 ns of control simulation, which varies in a zigzag manner in the range $5° - 13°$. These positive

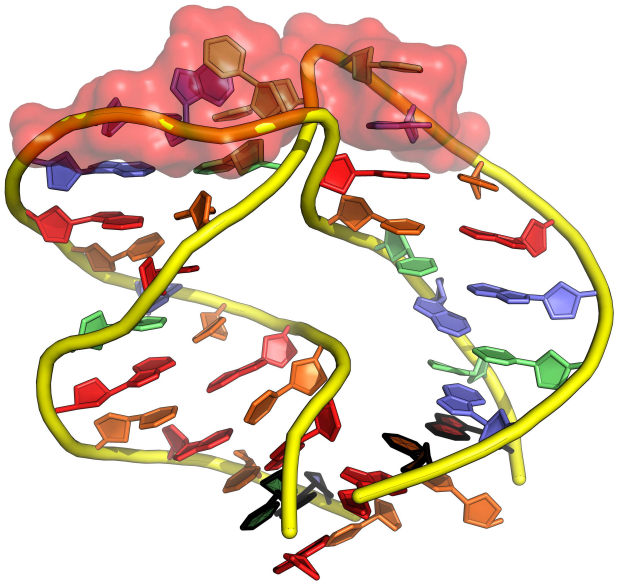
values of bending angles do not mean a consistent bending of DNA, but a helical 3D arrangement of basepair normal directions, which still results in a overall straight conformation. This profile is negatively correlated with the basepair stacking profile [red lines in Figure 3.4(b)], reflecting the geometric correlations between bending and projection. In the case of $\kappa = 16.6$ pN/nm, the bending angular scan profile (black line in Figure 3.5) is overall similar to that of control, but with slightly larger bending angles across the entire DNA, ranging from 5° to 18°. These larger bending angles result in a consistent directional bending of DNA in the same direction as the initial [Figure 3.3(a)]. Note that these bending angles weight differently to overall bending at different positions, due to the helical nature of DNA. More interestingly, the bending angles generally increase more at weakly stacked basepairs. Furthermore, there are no significant outliers, which quantitatively confirm the uniform bending observed in last section.

As we have setup the basepair integrity analysis methods, and quantitatively confirmed the intactness of DNA basepairs under weak bending conditions. Next, we can simply move to the strong bending constrained cases by applying the same methods to reveal their atomic causes of very distinctive bending behaviours in detail.

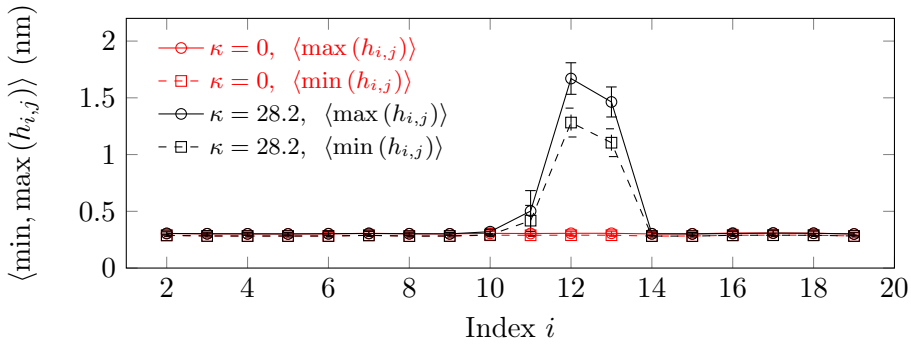
Case I: basepair disruptions under $\kappa = 28.2$ pN/nm

Figure 3.6(a) is a conformation snapshot obtained at 60 ns of a simulation with $\kappa = 28.2$ pN/nm, which contains a highly localized sharp bend near the middle of the DNA. And the normal secondary helical structures of *B*-DNA is locally disturbed within the sharp bend region, which was highlighted by red surfaces. Figure 3.6(b) shows its hydrogen bonding profile averaged over the last 20 out of 70 ns simulation. The much larger values of $\langle \min, \max(h_{i,j}) \rangle$ for $i = 11, 12, 13$, which are significantly deviated from those of control (red lines), clearly indicate that the hydrogen bonding in $11^{\text{th}} - 13^{\text{th}}$ basepairs are disrupted, while others are still intact. And Figure 3.6(c) shows its basepair stacking profile, averaged over the last 20 ns as well. The 0 Å^2 values of $\left\langle S^{\text{I}}_{i,i+1}, S^{\text{II}}_{i,i+1} \right\rangle$ for $i = 11, 13$ represent the total base unstacking between 11^{th} and 12^{th} basepairs, as well as, between 13^{th} and 14^{th} basepairs in both strands. It is interesting to note that

(a) Conformational snapshot at 60 ns, $\kappa = 28.2$ pN/nm



(b) Hydrogen bond lengths, $\kappa = 28.2$ *vs.* 0 pN/nm



(c) Basepair stacking areas, $\kappa = 28.2$ *vs.* 0 pN/nm

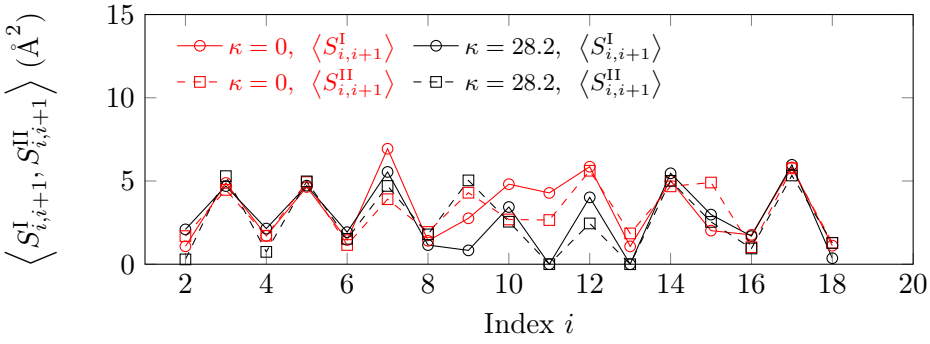


Figure 3.6: (a) A snapshot of a severely bent DNA conformation at 60 ns constrained using a strong contractile spring with $\kappa = 28.2$ pN/nm, which contains a large local kink around defects (highlighted by red surfaces) in the middle. (b) $\langle \min, \max(h_{i,j}) \rangle$ *vs.* i averaged over the last 20 ns reveals disrupted hydrogen bonding in three basepairs $i = 11, 12, 13$. (c) Equilibrated $\left\langle S^{\mathrm{I}}_{i,i+1}, S^{\mathrm{II}}_{i,i+1} \right\rangle$ *vs.* i reports totally disrupted basepair stacking in both strands at two locations $i = 11, 13$. The profiles of unconstrained DNA were plotted in red as controls in (b) and (c).

the level of hydrogen bonding disruptions at the 11^{th} basepairs is significantly lower than those at the $12^{\text{th}} - 13^{\text{th}}$ basepairs. Detailed analysis of the $h_{i,j}$ time trace for $i = 11$ shows that the 11^{th} basepair was under dynamic fluctuations between base-paired and disrupted states (see Figure 3.8, row 2), which results in smaller average values than those of total disruptions.

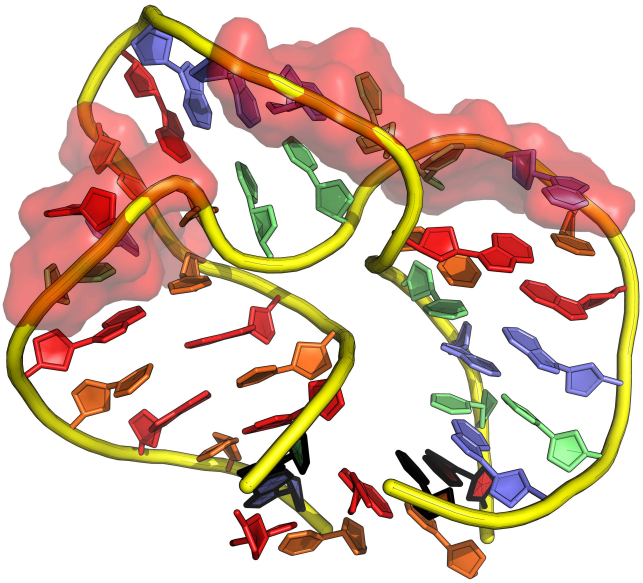
Case II: basepair disruptions under $\kappa = 33.2$ pN/nm

Similar analysis on another representative simulation with $\kappa = 33.2$ pN/nm reveals two regions with disrupted basepairs [Figure 3.7(b), 3.7(c)]. In the first region between $6^{\text{th}} - 9^{\text{th}}$ basepairs, the hydrogen bonding of 7^{th} and 8^{th} basepairs are disrupted, while the base stacking between 6^{th} and 7^{th} basepairs, as well as between 8^{th} and 9^{th} basepairs are totally unstacked in Strand II and I, respectively. In the second region $12^{\text{th}} - 14^{\text{th}}$ basepairs, the hydrogen bonding of 12^{th} and 13^{th} basepairs are disrupted, while the base stacking between 12^{th} and 13^{th} basepairs, as well as between 13^{th} and 14^{th} basepairs are totally unstacked in Strand I and both strands, respectively. The snapshot of DNA conformation at 60 ns visualizes the localized sharp bends and disturbed secondary helical structures (highlighted by red surfaces) in the aforementioned regions. The lower level of basepair disruptions observed at the $6^{\text{th}} - 9^{\text{th}}$ basepairs against those at $12^{\text{th}} - 14^{\text{th}}$ basepairs can be explained by milder disruptions of non-covalent interactions. Similar to previous case, the hydrogen bonding in 7^{th} and 8^{th} basepairs were under dynamic fluctuations between partially closed and opened states (see $h_{i,j}$ time traces for $i = 7, 8$ in Figure 3.9, rows $2 - 3$ for detail), while, their basepair stacking were both partially disturbed with 0 Å^2 of $\left\langle S^{\text{I}}_{i,i+1}, S^{\text{II}}_{i,i+1} \right\rangle$ in only one of their strands.

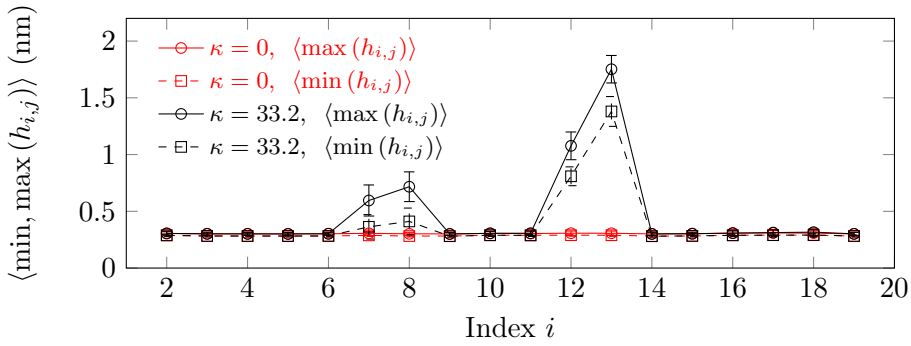
3.3.4 Localized sharp bends caused by basepair disruptions

As summarized from last subsection, the local defects at the middle of DNA in form of basepair disruptions have been induced by strong contractile springs. Then, in this subsection, we are going to explore the consequences of defects to DNA overall bending deformations.

(a) Conformational snapshot at 60 ns, $\kappa = 33.2$ pN/nm



(b) Hydrogen bond lengths, $\kappa = 33.2$ *vs.* 0 pN/nm



(c) Basepair stacking areas, $\kappa = 33.2$ *vs.* 0 pN/nm

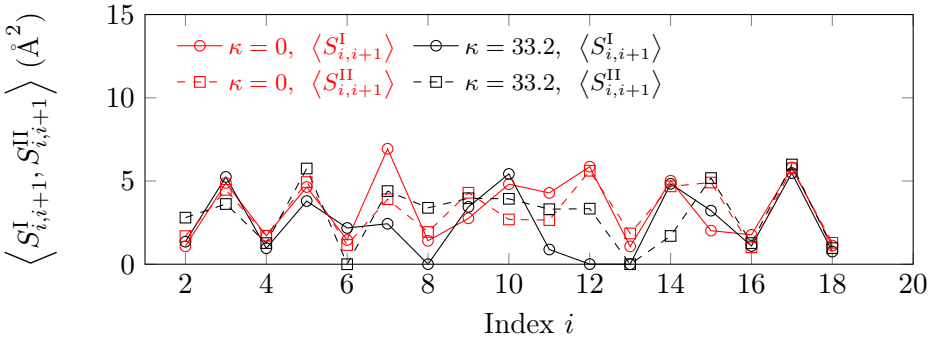


Figure 3.7: (a) A snapshot of a severely bent DNA conformation at 60 ns constrained using a strong contractile spring with $\kappa = 33.2$ pN/nm, which contains two large local kinks around defects (highlighted by red surfaces) in $6^{\mathrm{th}} - 9^{\mathrm{th}}$ and $12^{\mathrm{th}} - 14^{\mathrm{th}}$ basepairs. (b) $\langle \min, \max(h_{i,j}) \rangle$ *vs.* i averaged over the last 20 ns reveals disrupted hydrogen bonding in four basepairs $i = 7, 8, 12, 13$. (c) Equilibrated $\left\langle S^{\mathrm{I}}_{i,i+1}, S^{\mathrm{II}}_{i,i+1} \right\rangle$ *vs.* i reports totally disrupted basepair stacking at four locations $i = 6, 8, 12, 13$, in Strand II, I, I and both strands, respectively. The profiles of unconstrained DNA were plotted in red as controls in (b) and (c).

Case I: local bending deformations under $\kappa = 28.2$ pN/nm

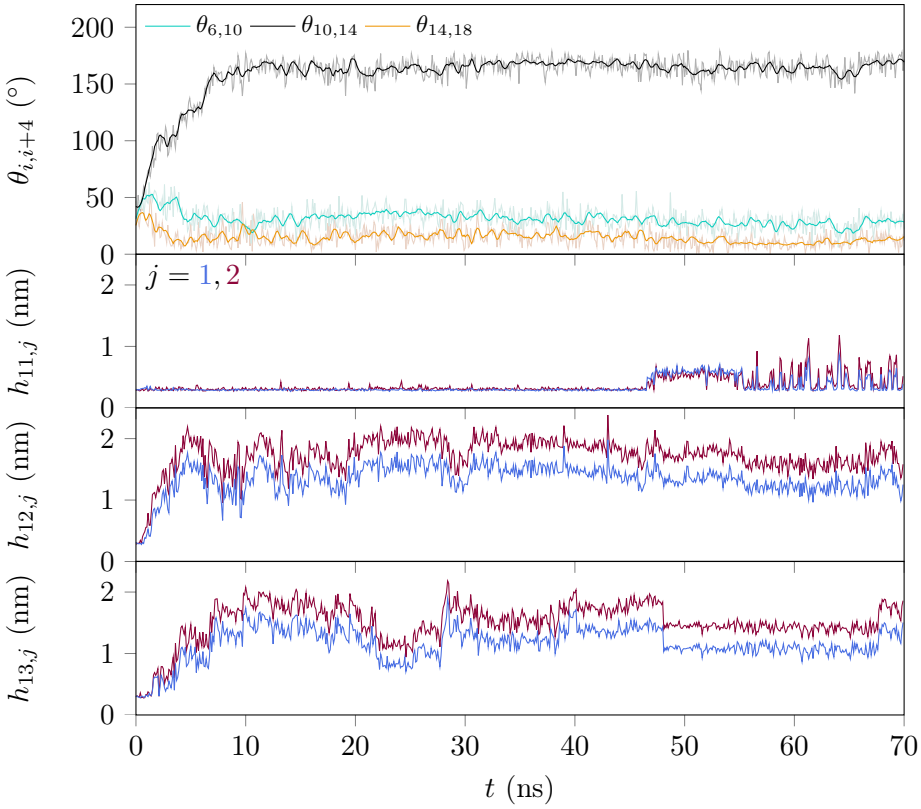


Figure 3.8: 70 ns dynamics of local bending deformations and hydrogen bonding disruptions under $\kappa = 28.2$ pN/nm. Row 1: time evolution of $\theta_{10,14}$ (black) enclosing three disrupted basepairs at $i = 11, 12, 13$, which shows the kink development around defected region. The bending angle evolutions of two intact regions with same length, $\theta_{6,10}$ (cyan) and $\theta_{14,18}$ (orange), are shown for comparison. Rows $2 - 4$: time evolutions of $h_{i,j}$ for the three disrupted basepairs $i = 11, 12, 13$, which are all `AT` basepairs and two atom-atom distances are involved ($j = 1$ in dark blue and $j = 2$ in dark red).

Because local bending angles, $\theta_{i,i+\Delta}$, are only definable by using normal directions of intact i^{th} and $(i + \Delta)^{\text{th}}$ basepairs (Equation 3.2). We extracted the time evolutions of bending angles between two boundary Watson-Crick basepairs that straddle several targeted DNA regions with same contour length to study their bending deformations. The first row in Figure 3.8 shows 70 ns evolution of the bending angle, $\theta_{10,14}$ (black), between the intact 10^{th} and 14^{th} basepairs, which enclose the region affected by the $11^{\text{th}} - 13^{\text{th}}$ basepair-disruption in the simulation with $\kappa = 28.2$ pN/nm (*i.e.*, the case I in last subsection). Evolution from an initial value of $\sim 40°$ toward larger DNA bending angle began immediately after the simulation started till saturation was reached within 10 ns, and

remained at that high level of bending at $\sim 170^\circ$ till the end. It indicates the formation of sharp bend in the middle of DNA around the disrupted basepairs.

For comparison, the bending angle evolutions of two unaffected regions with same length (*i.e.*, $\Delta = 4$), $\theta_{6,10}$ (cyan) and $\theta_{14,18}$ (orange), were plotted. Synchronized with the DNA kink formation of $\theta_{10,14}$, the bending angles of these two regions were relaxed from the similar initial values of $\sim 40^\circ$ to lower values of $\sim 30^\circ$ and $\sim 15^\circ$ within 10 ns, respectively, and remained at these low levels of bending throughout the rest of simulation. These results suggest that the kink formation of $\theta_{10,14}$ contributed to relax the rest of DNA to a more straight conformation.

Rows $2-4$ in Figure 3.8 show the time evolutions of the inter-distances of atoms involved in hydrogen bonding, $h_{i,j}$, for the three defected `AT` basepairs $i = 11, 12, 13$ and $j = 1, 2$. The time traces for the $12^{\text{th}} - 13^{\text{th}}$ basepairs show that both of their hydrogen bonds opened up significantly within 5 ns, and then remained in the disrupted state throughout the remaining simulation. As time-courses of these basepairs disruptions are highly correlated with its local sharp bend formation and the rises of $h_{i,j}$ are slightly prior to that of bending angles, we conclude that disruptions of basepairs are the cause of local kink development in this simulation. Furthermore, the time traces for the slightly disrupted 11^{th} basepair reveal that it maintained Watson-Crick base-paired in the first $\sim$ 48 ns. Later, the hydrogen bonding was disrupted between $\sim$ 48 and $\sim$ 56 ns. After $\sim$ 56 ns, its basepairing fluctuated between disrupted and intact states. Although the disruption at 11^{th} basepair was trivial to the overall DNA bending deformations, it proposes the potential of recoveries and deteriorations within $\sim$ 10s ns timescales for the defects.

Case II: local bending deformations under $\kappa = 33.2$ pN/nm

Figure 3.9 shows similar bending angle and base-pairing inter-distance dynamics obtained from the simulation with $\kappa = 33.2$ pN/nm (*i.e.*, case II in last section). The first row contains the bending angle evolutions between two nearest intact basepairs enclosing the disrupted basepairs at two defected regions, $\theta_{6,9}$ (red) and $\theta_{11,14}$ (black). Within 10 ns after simulation started, the bending angles of these

two disrupted regions evolved from initial values of $\sim 40°$ to greater than $70°$ and finally stabilized at $\sim 80°$ and $\sim 140°$, respectively. Synchronizing with the kink formations at these two regions, the bending angles at intact basepair regions with the same length (*i.e.*, $\Delta = 3$), $\theta_{3,6}$ (cyan) and $\theta_{14,17}$ (orange), relaxed from initial bending angles of $\sim 40°$ to smaller than $20°$. These observations also show that DNA locally bends at defects, and subsequently relaxes the rest.

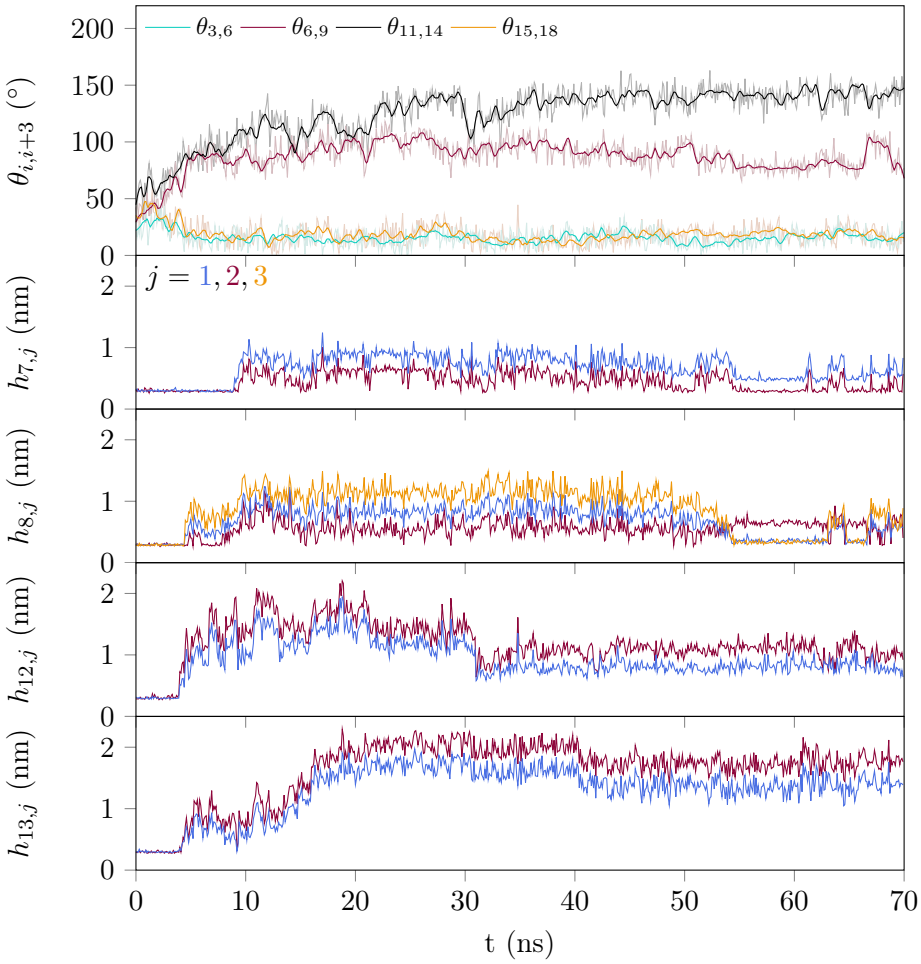


Figure 3.9: 70 ns dynamics of local bending deformations and hydrogen bonding disruptions under $\kappa = 33.2$ pN/nm. Row 1: time evolutions of $\theta_{6,9}$ (red) and $\theta_{11,14}$ (black) enclosing two disrupted basepairs each (*i.e.*, $i = 7, 8$ and $i = 12, 13$ respectively) in two different regions, which illustrate the kink developments around defected sites. The bending angle evolutions of two intact regions with same length, $\theta_{3,6}$ (cyan) and $\theta_{15,18}$ (orange), are shown for comparison. Rows $2 - 5$: time evolutions of $h_{i,j}$ for the four disrupted basepairs $i = 7, 8, 12, 13$, which are all `AT` basepairs and two atom-atom distances are involved ($j = 1$ in dark blue and $j = 2$ in dark red), except for 8^{th} `GC` basepair, which includes an additional base-paring inter-distance ($j = 3$ in orange).

Rows $2 - 5$ in Figure 3.9 show the evolutions of $h_{i,j}$ for all four affected basepairs $i = 7, 8, 12, 13$. In the first defected region $6^{\text{th}} - 9^{\text{th}}$ basepairs, the

dynamics of $h_{8,j}$ are positive correlate with and slightly ahead of the evolution of $\theta_{6,9}$. It illustrates that the initial kink formation of $\theta_{6,9}$ at ~ 5 ns was caused by disruption of the hydrogen bonds of the 8^{th} basepair (*i.e.*, `GC` basepair with three hydrogen bonds). At ~ 10 ns, the other basepair, 7^{th} basepair, in this region was also disrupted. In the period of $10 - 50$ ns, the two basepairs remained in the disrupted state. In the last ~ 20 ns, both the 7^{th} and 8^{th} basepairs fluctuated between the disrupted state and a nearly intact state (*i.e.*, the hydrogen bonds formed, but the distances are slightly larger than the totally intact state). As the nearly intact state predominated ($\sim 70\%$) in the last 20 ns, the averaged level of disruption for the two basepairs became small, consistent with the low level of overall basepair disruption revealed in Figure 3.7(b). On the other hand, in the second defected region $11^{\text{th}} - 14^{\text{th}}$ basepairs, the $h_{i,j}$, where $i = 12, 13$ and $j = 1, 2$, were increased significantly at ~ 5 ns, leading and correlating with the kink formation of $\theta_{11,14}$. The basepairs remained in the disrupted state throughout the rest of the simulation.

In summary, the twelve independent simulations with $\kappa > 25$ pN/nm are all analogous to these two representative cases, in which strong bending constraints induce usually one defect in the middle of DNA, seldom two defects (*i.e.*, only in the case with $\kappa = 33.2$ pN/nm, see Figure 3.10 for details). These defects cause DNA to immediately deform and generate local sharp bends in the respective regions. Then, the kink formations absorb bending constraints and in turn relax the rest of DNA. Under our bending constraints, the defects are mainly disrupted basepairs with both disturbed hydrogen bonding and basepair stacking. However, we also observed large kinks at sites where the hydrogen bonding are nearly intact (*e.g.*, the last 20 ns in Figure 3.9, rows $2 - 3$).

3.4 Effects of DNA sequence on the localization of defects

Figure 3.10(a) plots the hydrogen bonding profiles, $\langle \min(h_{i,j}) \rangle$ and $\langle \max(h_{i,j}) \rangle$ *vs.* i averaged over the last 20 ns, for all twelve independent simulations with $\kappa > 25.0$ pN/nm. This "histogram" reveals that the disrupted basepairs appeared

around the same region near DNA center that happen to be AT-rich (*i.e.*, $5' -$ AAAT $- 3'$, the $10^{\text{th}} - 13^{\text{th}}$ basepairs). So it is interesting to investigate the causes for this consistent localization of defects.

One possible cause to the central localization of basepair disruptions is the largest curvature at the center under bending constraints. Assuming the two ends of the homogeneous WLC polymer meets before any defects appears, we can orient its elastic energy minimal rigid path on xy-plane by setting its terminal cross point as origin, center point on the positive side of y-axis, then this rigid path breaks into two reflection symmetric halves. By defining the angle between unit tangent vector $\hat{t}(s)$ (alone the half path in Quadrant I) and x-axis as $\upsilon(s)$, we have its relationship with curvature [103] as,

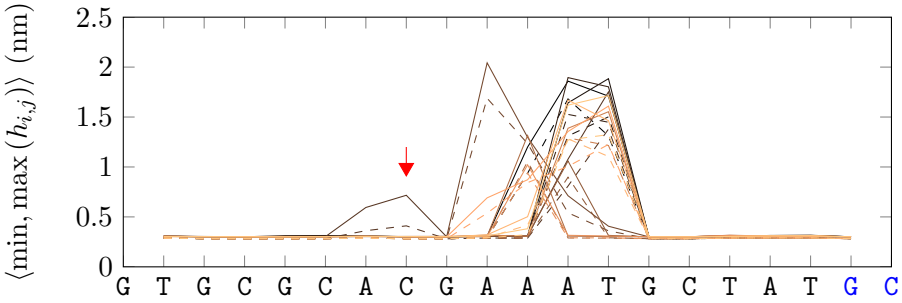
$$L^2 \left(\frac{\partial \hat{t}(s)}{\partial s} \right)^2 = -\lambda \cos\left(\upsilon(s)\right) + c \tag{3.3}$$

, where L is the contour length, $\lambda > 0$ is a Lagrange multiplier, and $c > 0$ is integration constant. Following above equation, the curvature is maximized at the center, as $\upsilon\left(\frac{L}{2}\right) = \pi$ by symmetry.

Alternatively, it may be due to the less stable AT non-covalent interactions in the middle of our DNA. Based on the unified NN basepair parameters in Table 1.2, melting AT next to AT basepair (*i.e.*, $\Delta\mathcal{G} < 2\ k_{\text{B}}T$) is generally easier than melting AT next to GC or melting GC next to AT basepairs (*i.e.*, $3 > \Delta\mathcal{G} > 2\ k_{\text{B}}T$), while melting GC next GC basepairs is hardest (*i.e.*, $\Delta\mathcal{G} > 3\ k_{\text{B}}T$). Note that the signs of free energy are reversed from SantaLucia's representations to denote cost.

To see which factor predominates, we shifted the entire sequence tail-to-head by 2 bp and replaced the central AT-rich island at $10^{\text{th}} - 13^{\text{th}}$ basepairs with $5' -$ CGAA $- 3'$. Based on this new sequence, a new initial structure, which is similar in shape with Figure 3.2, was built using the set of complimentary base and sequential basepair parameters extracted from the original-sequence initial. Five independent 70 ns simulations under different bending constraints with $\kappa > 25$ pN/nm were conducted following the same procedures in Subsection 3.3.1. The overlay of their hydrogen bonding profiles [Figure 3.10(b)] shows that the basepair

(a) Hydrogen bonding profiles for original sequence



(b) Hydrogen bonding profiles for modified sequence

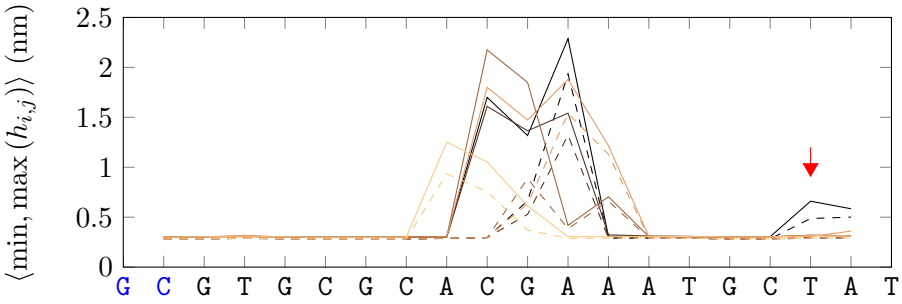


Figure 3.10: Hydrogen bonding profiles of the defected DNA with original sequence $5' - \texttt{GTGCGCACGAAATGCTATGC} - 3'$ and modified sequence $5' - \texttt{GCGTGCGCACGAAATGCTAT} - 3'$. Overlay of $\langle \min(h_{i,j}) \rangle$ (dashed) and $\langle \max(h_{i,j}) \rangle$ (solid) along the DNA sequence, averaged over the last 20 ns for (a) twelve independent simulations with the original sequence and (b) five independent simulations with the modified sequence, under various bending constraints with $\kappa > 25.0$ pN/nm. These hydrogen bonding profiles were coloured from light to dark copper as κ increases, respectively (*i.e.*, $\kappa = 26.6, 28.2_{\mathrm{I}}, 28.2_{\mathrm{II}}, 28.2_{\mathrm{III}}, 28.2_{\mathrm{IV}}, 28.2_{\mathrm{V}}, 29.0, 31.5, 33.2, 41.5, 49.8, 83.0$ pN/nm for original sequence, while $\kappa = 28.2, 31.5, 33.2, 41.5, 49.8$ pN/nm for modified sequence). The modified sequence was generated from the original sequence by removing its tailing $5' - \texttt{GC} - 3'$ and plugging it back to its front, which offset the AT-rich region (*i.e.*, its $10^{\mathrm{th}} - 13^{\mathrm{th}}$ basepairs) away from its center. The arrow in panel (a) indicates an additional disrupted region slightly off center in one of the twelve simulations, while the arrow in panel (b) points out the AT end peeling in one of the five simulations.

disruptions still occur at the central region, mainly at the $10^{\mathrm{th}} - 11^{\mathrm{th}}$ basepair (*i.e.*, GC base-pairing), and 12^{th} basepair (*i.e.*, AT base-pairing). Taken together, these results suggest that the central localization of the basepair disruptions is mainly caused by the high curvature at the center, while the sequence effects are minor under our bending constructs.

3.5 Discussion

In this chapter, we have examined the mechanical responses of short DNA fragments under sharp bending constraint using MD simulations. In this type of study, various degrees of DNA bending was induced by connecting the two DNA ends with zero-length springs and tuning the spring constants. We find that mechanical defects indeed can be excited when DNA is sharply bent enough, which involve disruptions of hydrogen bonds and base stacking over several basepairs, and are mostly localized in the middle of this short DNA. At these defected sites, large local kinks are formed immediately after non-covalent disruptions, which relaxed the bending in the rest of DNA, and retained them in normal *B*-form. These observations suggest that these defects are flexible in nature, which might be related to several recent experiments that reported anomalous elasticity of tightly bent DNA, for example, the anomalously high probabilities observed for < 110 bp DNA looping [3, 29, 104] and large bending angles of < 30 nm DNA using AFM imaging [22]. In the following chapter, we are going to analyze the impacts of defect excitations on the overall mechanical properties of DNA during bending.

Chapter 4

Micromechanical properties of DNA with defect excitations

4.1 Introduction

From mechanical point of view, we have extensively studied the processes of DNA deformations under compressional load in last chapter, by directly observing their structural and morphological evolutions at atomic level resolutions. In a nutshell, these micromechanical processes initiate from the homogeneity breakages of DNA, and, as defects generate, propagate and stabilize in forms of hydrogen bonding and basepair stacking disruptions, result in heterogeneous bending of DNA. In this chapter, we are going to further investigate the micromechanical properties [105] of DNA under bending constraints by broadly sampling its configuration space to obtain integrated and quantified results.

During DNA bending, this thorough conformational sampling processes against reaction coordinates, end-to-end distances (d) in this case, was accelerated using umbrella sampling simulations [92]. More explicitly, multiple MD simulations constrained by contractile springs with various intrinsic lengths (l_k) were conducted, during which their end-to-end distances were forcibly restrained to and fluctuated near particular l_k. The resulting equilibrated statistics, under various bending conditions, ranging from weak gradually to strong, were, then, used to reconstruct a DNA free energy difference profile, $\Delta\mathcal{A}(d)$, based on WHAM [93]. By differentiating this free energy difference profile, we further

acquired mechanical characteristic force-extension curve, $f(d)$, for DNA without and with defect excitations.

4.2 Nanosecond timescale importance sampling

4.2.1 Umbrella sampling simulations

We use the end-to-end distance, d, (*i.e.*, the distance between center-of-mass of paired bases in the second and second-last basepairs, which globally represents the degree of bending) as our reaction coordinate to study the DNA micromechanical properties during bending. The non-constrained B-DNA simulations usually trapped around its global energy minimum state with instant d not far from its most probable value, denoted by d_0. Although the system is ergodic, direct sampling at much smaller d is always insufficient within our timescales. In order to speed up the sampling process, we utilize the contractile springs again, but with various intrinsic length of l_k and spring constant κ_k. As a result, besides providing compressional loads, they forcibly collapse DNA ends to fluctuate near targeted l_k. It is equivalent to apply a series of biasing potentials $\mathcal{V}_k = \frac{\kappa_k}{2}(d - l_k)^2$ to boost the sampling at particular reaction coordinates. This method is umbrella sampling simulations.

Twelve pre-bent DNA conformations with different end-to-end distances, which evenly span over a range of ~ 2.8 to 5.3 nm, were extracted from previous simulations (*i.e.*, those two described in Chapter 3 with zero length contractile springs and $\kappa = 8.3$ and 28.2 pN/nm, specifically). These conformations were set as initials for independent umbrella sampling simulations, with their end-to-end distances equal to intrinsic lengths of respective constraining springs. Here these simulations are indexed by k, for $k = 1, 2, \cdots, 12$, ordered by decreasing l_k. In the rest of the chapter, a quantity with subscription of $\{k\}$ denotes that it was obtained through k^{th} simulation biased by additional potential $\mathcal{V}_k$. For spring constant, we applied the same $\kappa_{\text{u}} = 248.9$ pN/nm across all these simulations. The choice of spring constant is critical to the sampling process. Larger values of spring constant result in very local samplings. Although sufficient for particular d, they only explore a narrow range of reaction coordinate. On the other hand,

smaller values of spring constant yield wider distributions, but lack the abilities to overcome local energy barriers. Ideally, a set of locally adaptive κ_k achieves sufficient sampling for each d in a overlapped wide range, while brings any energy barriers down to the level with energy differences of only several $k_{\mathrm{B}}T$. Due to the complexity and inefficiency of the recursive searching process for optimized κ_k, we stopped at several trials, which finalized a reasonable choice of κ_{u} for all twelve independent simulations.

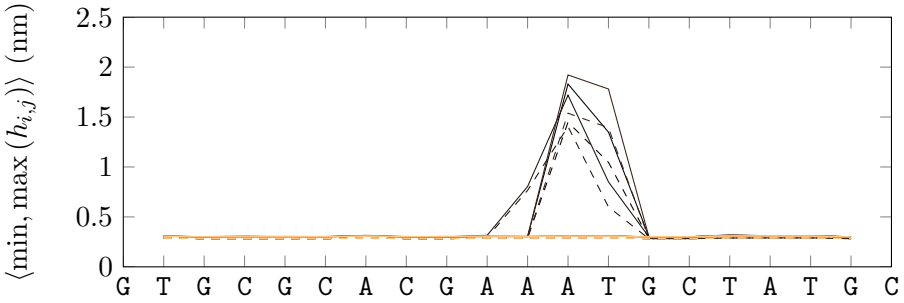
Now, for each independent simulation, following the setup procedures described in Subsection 3.3.1, the k^{th} initial was centered at the same rhombic dodecahedron unit cell used before, surrounded by TIP3P water and 150 mM NaCl, prepared by energy minimizations and constrained by the k^{th} contractile spring with harmonic potential of $\mathcal{V}_k = \frac{\kappa_{\mathrm{u}}}{2}(d - l_k)^2$. Then, twelve MD simulations were conducted in parallel under NTV ensemble at constant temperature of 300 K and volume of ~ 1170 nm^3.

4.2.2 Reconstructing the unbiased sampling

Among these initial conformations, there are 10 intact DNA ($k = 1, 2, \cdots, 10$) and 2 defect-excited DNA containing disrupted $12^{\mathrm{th}} - 13^{\mathrm{th}}$ basepairs ($k = 11, 12$). Although some of them were selected from non-equilibrated portions of trajectories, they are still easier to reach equilibrated configurations, compared against generated uniformly curved DNA as an example. For each MD simulation, total of 50 ns trajectories were produced, and $10,000$ representatives were collected from the last 20 ns with sampling interval of 2 ps. Based on equilibrated hydrogen bonding profiles in Figure 4.1(a), nine out of ten intact DNA remain in B-form ($k = 1, 2, \cdots, 9$), while the 10^{th} simulation developed a defect. The two simulations, which started from defected initials, preserved their defects. And all the three most constrained DNA contain disrupted hydrogen bonding in the same middle region of $11^{\mathrm{th}} - 13^{\mathrm{th}}$ basepairs ($k = 10, 11, 12$) throughout their last 20 ns dynamics.

For each simulation, the direct probability density function, $\rho_{\{k\}}(d)$, was plotted in Figure 4.1(b). The biased distributions for defect-excited DNA ($k = 10, 11, 12$), all locate at the small d region, while, for the rest nine from intact

(a) Hydrogen bonding profiles for umbrella sampling



(b) Direct probability density functions for umbrella sampling

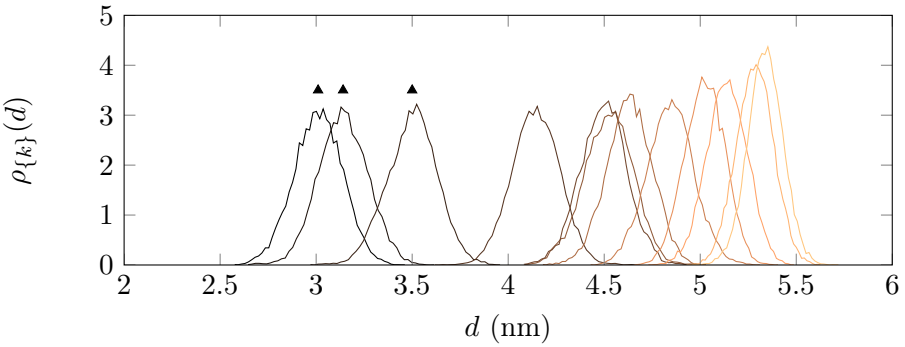


Figure 4.1: Umbrella sampling with twelve simulations constrained by different potentials (a) Hydrogen bonding profiles of their equilibrated conformations averaged over the last 20 ns. Solid lines show $\langle \max(h_{i,j}) \rangle$ and the dashed lines show $\langle \min(h_{i,j}) \rangle$. Three most bent DNA contains disrupted hydrogen bonds in the middle region, while the other nine are intact *B*-DNA with various degrees of bending. (b) Biased probability density functions against end-to-end distances for the twelve simulations obtained from collected samples in the last 20 ns each. The distributions indicated with ▲ are from the three defect-excited DNA, while the rest are obtained from the nine intact DNA. $\rho_{\{k\}}(d)$ overlap with each other in the entire d range from $\sim$ 2.5 to 5.7 nm. Lines are coloured from light to dark as intrinsic contractile spring lengths l_k decreases (*i.e.*, l_k = 5.27, 5.18, 4.94, 4.79, 4.56, 4.31, 4.17, 4.16, 3.80, 3.37, 3.01, 2.85 nm respectively) in both (a) and (b).

DNA, group together at the large d region. Generally, they overlapped the entire region of end-to-end distance, from $\sim$ 2.5 to 5.7 nm. Relatively speaking, there is a gap near $\sim$ 3.8 nm, at where the transitions between intact and defect-excited DNA occur. We have not successfully obtain a simulation, during which the conformation fluctuated back and forth between two states. It indicates that the sampling near transition coordinate is insufficient within our simulation time; because the energy barrier has not been flattened enough to the thermally excited level.

Theoretically, the unbiased probability relates to the k^{th} biased probability by a factor (Equation 2.35), which is the product of a reaction coordinate de-

pendent term, $\exp\left(\frac{\beta\kappa_u}{2}(d-l_k)^2\right)$ and a reaction coordinate independent term, $\frac{\mathcal{Z}_{\{k\}}}{\mathcal{Z}_{\{0\}}}$, where $\mathcal{Z}$ is the partition function. The second term can be estimated by considering all the statistics acquired from different simulations, using WHAM method introduced in Subsection 2.6.2. Practically, we used the algorithm by Grossfield [106], which is implemented by `g_wham` [107] in `GROMACS` package, to reconstruct the unbiased probabilities, $\tilde{p}_{\{0\}}(d)$, which minimizes its statistical error $\delta^2\tilde{p}_{\{0\}}(d)$ through recursively solving WHAM equations (Equation 2.40). Due to the insufficient sampling at the transition coordinate, we separately evaluated the unbiased probabilities at 200 values of d each, for the two distinctive states.

4.3 Free energy difference profile with defect excitations

Under NTV ensemble, the Helmholtz free energy difference profile of DNA with defect excitations, reference to the global minimum state, was obtained by,

$$\Delta\mathcal{A}(d) = -\beta^{-1}\ln\left(\tilde{p}_{\{0\}}(d)\right) + \mathcal{A}_{\text{offset}} \tag{4.1}$$

, as shown in Figure 4.2. The constant $\mathcal{A}_{\text{offset}}$ for B-form DNA is chosen to set the most possible d_0 at 0 k_BT, and we get the $\Delta\mathcal{A}(d)$ ($\bigcirc$) in the range of $3.5 < d < 5.7$ nm for intact DNA. Then, a cubic spline interpolation was applied to obtain a smooth continuous version of free energy difference profile, which was superimposed with previous discretized version in a smaller range of $3.9 < d < 5.6$ to avoid the scattered data at boundaries. For basepair disrupted DNA, the unknown offset is chosen, so that the two profiles match each other as much as possible at the overlapped region. Thus, similarly, a discretized version of $\Delta\mathcal{A}(d)$ ($\square$) was obtained in the range of $2.5 < d < 4.0$ nm, with a smooth continuous version overlaid in a smaller and more confident range of $2.8 < d < 3.8$ nm for defected-excited DNA.

The free energy difference profile contained a single energy minimum at $d_0 \approx 5.43$ nm, which is the same as $\langle d_{\{0\}}\rangle$ obtained from unconstrained simulation in last Chapter. This implies energy well is symmetric around d_0 within

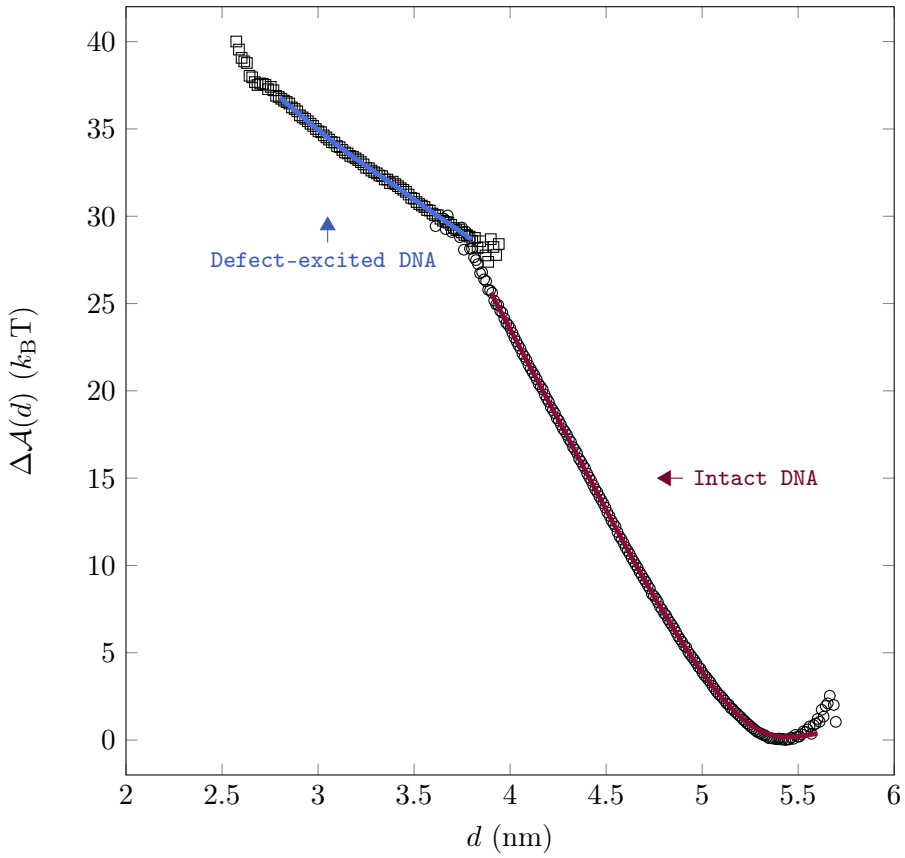


Figure 4.2: Free energy difference profile for DNA with defect excitations. Discretized $\Delta\mathcal{A}(d)$, reference to global minimum state, separately obtained for intact (○) and defect-excited DNA (□) are plotted against end-to-end distances. Smoothed and continuous cubic-spline interpolations for these two states are overlaid, drawn in dark red for B-form DNA, dark blue for basepair disrupted DNA. This profile has the energy minimal at $d_0 \approx 5.43$ nm.

small deviations, which is correct based on our $\Delta\mathcal{A}(d)$. Furthermore, this d_0 corresponds to $\sim$ 0.32 nm extension per basepair step, which is similar to the experimentally determined rise, $D_z = 0.332 \pm 0.019$ nm in Table 2.1. Then, as d decreases, the free energy of DNA climbs up rapidly during bending. However, after the transition coordinate at $\sim$ 3.8 nm, the slope suddenly drops, which suggests that DNA becomes easier to bend once broken.

4.4 Force-extension curve with defect excitations

The force-extension relationship of DNA with defect excitations can be obtained using two approaches. Firstly, from continuous free energy difference profile, continuous force-extension curve is obtained by derivative, $f(d) = -\frac{\partial\Delta\mathcal{A}(d)}{\partial d}$. Secondly, forces on springs can be directly read out through Hooke's law for each

independent simulation. Then, the force at corresponding equilibrated mean end-to-end distance, $\langle d_{\{k\}} \rangle$, is estimated by the average force, $f\left(\langle d_{\{k\}} \rangle\right) = \langle f_{\{k\}} \rangle = \langle \kappa_{\mathrm{u}} \left(d_{\{k\}} - l_k\right) \rangle$. Two continuous force profiles were calculated using the first approach, and are shown using solid lines in Figure 4.3. The three points (□) were evaluated based on the second approach using last 20 ns representatives of three basepair disrupted DNA simulations, for $k = 10, 11, 12$. They more or less locate on its continuous force-extension curve for defect-excited DNA. While, the other nine $f\left(\langle d_{\{k\}} \rangle\right)$, for $k = 1, 2, \cdots, 9$, (○) also agree with $f(d)$ for intact DNA.

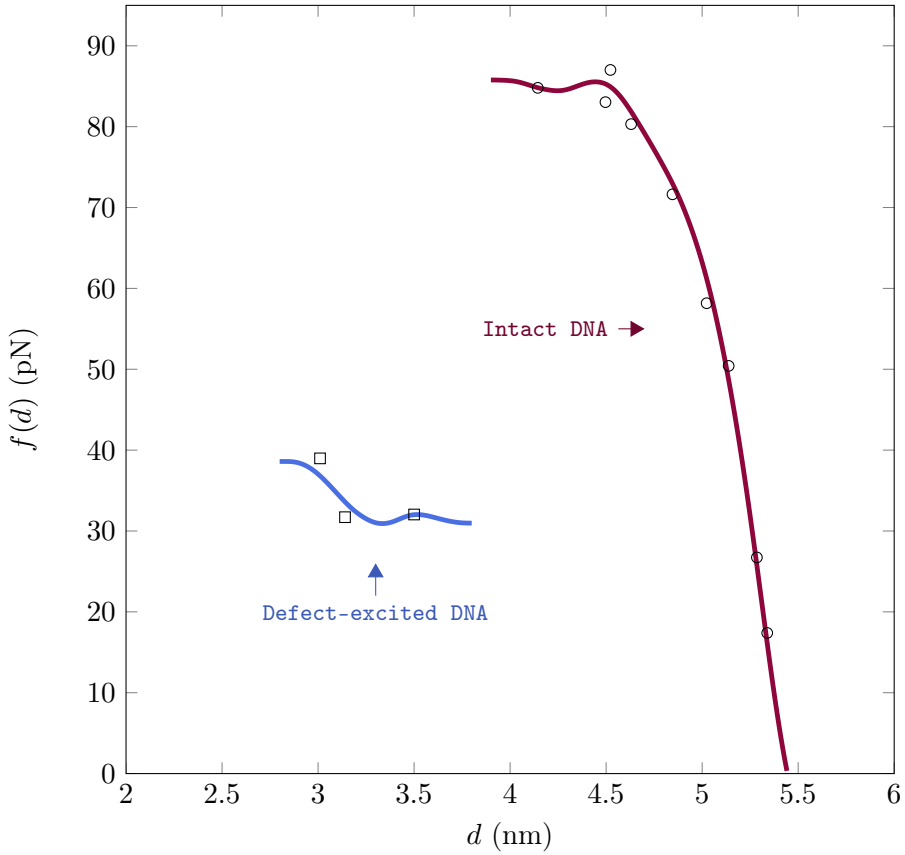


Figure 4.3: Force-extension curve for DNA with defect excitations. Continuous $f(d)$ for defect-excited DNA locate at small d is drawn in dark blue, while that for intact DNA at large d is plotted in dark red. The three points of $\langle f_{\{k\}} \rangle$ against $\langle d_{\{k\}} \rangle$ (□), for $k = 10, 11, 12$, averaged from the equilibrated fluctuations of defect containing DNA simulations, are roughly on $f(d)$ in dark blue. While, nine points (○), directly read out from B-DNA simulations, are overlaid with $f(d)$ in dark red. Note that $f(d)$ roots at d_0 with 0 pN.

For intact DNA, the force is 0 pN at d_0, which is expected because it is the energy minimal state observed from free energy difference profile. The force shoots up rapidly when $d < d_0$, and in a small range near d_0, the stress-stain response is quite linear. Upon further compression at $4.8 > d > 4.6$ nm, the slope

reduced significantly under corresponding force range of $70 - 85$ pN. Then, for even shorter d, $f(d)$ becomes nearly flat, and the slope stabilizes again. Based on the linearity of force-extension curve, the persistence length is estimated based on Equation 1.14, as,

$$\tilde{A} = \beta Y I = \beta \frac{\Delta f}{\Delta d} \frac{LI}{S} = 57.0 \text{ nm} \tag{4.2}$$

, where $L \approx d_0$ is the contour length, $S = \pi R^2$ is the area of DNA cross section, $I = \frac{\pi R^4}{4}$ is area moments of inertia, $R \approx 1$ nm is the radius, and Δf, Δd are estimated using the two averaging points, $f\left(\left\langle d_{\{k\}} \right\rangle\right)$, for $k = 1, 2$. The persistence length obtain is in good agreement with $A = 53.4 \pm 2.3$ nm measured in DNA stretching experiments [12]. Moreover, the slope altering behaviour follows a typical Euler instability response of an ideal elastic rod with $L \ll A$, which predicts a critical force for the onset of rod bending (*i.e.*, switch point between the steep linear region to flattened region). It is evaluated based on our estimated persistence length as,

$$\tilde{f}_{\text{c}} = \beta^{-1} \pi^2 \frac{\tilde{A}}{L^2} = 79.1 \text{ pN} \tag{4.3}$$

, which falls right in the middle of our observed slope-alternating force range from ~ 70 to 85 pN. These evidences indicate that our MD simulations achieved typical WLC behaviours for intact B-form DNA, which are consistent with both theoretical predictions and experimental measurements. This successfully reproduction of typical elastic rod bending responses further proves that MD methods with ParmBSC0 force filed are suitable to study equilibrated DNA large-scale conformational behaviours using ~ 100 ns timescales simulations.

For the defect containing DNA, the force for maintaining specific d suddenly drops off a magnitude of ~ 50 pN from the ~ 85 pN force plateau of B-DNA force profile, near transition coordinate at ~ 3.8 nm. This sudden change reveals the first order structure transition nature of the defect excitation in B-DNA. Then, upon further bending, $f(d)$ of defected DNA stays at similar level over a wide range of d, till ~ 3 nm. This result suggests that DNA becomes much more bendable compared with intact DNA, once defected.

4.5 Discussion

In this chapter, using conformational sampling processes with restraining potentials on DNA end-to-end distances, d, we obtained the free energy difference profile $\Delta\mathcal{A}(d)$ and the force-extension curve $f(d)$, for intact B-form DNA, as well as, defect-excited DNA.

$\Delta\mathcal{A}(d)$ monotonically increases as d decreases for both intact and defected DNA. The effect of defect is manifested by a flatter profile compared to that of B-DNA. The difference between the B-DNA and the defected DNA is more prominent when we look at their $f(d)$ profiles. $f(d)$ of B-DNA increases sharply when d decreases, and then flatten out which marks the Euler instability of ideal rod. Using the linear region right after d_0 and before the onset of bending, the Young's modules was calculated, and subsequently a bending persistence length of $\sim$ 57 nm was estimated. This value perfectly agrees with that determined by single-DNA molecule stretching experiments [12, 1]. Using this persistence length, the critical force for buckling transition to occur was predicted to be $\sim$ 79 pN, which is consistent with obtained $f(d)$ profile. These observations validate the force field and the theoretical sampling analysis. In sharp contrast with $f(d)$ obtained for B-DNA, the defected DNA has an overall flat $f(d)$ with much smaller magnitude compared to the plateau region of the B-DNA $f(d)$. This result clearly indicates that the flexible defect excited in the DNA fragment costs much less force to maintain DNA in server-bending state.

Overall, for the first time, we obtained the short-scale DNA elastic responses under wide range of bending levels, which characterized by expected elastic rod behaviours below moderate level of bending, and defect-induced softening under strong bending. It verifies the WLC elasticities of B-DNA of our 20 bp DNA with persistence length $A \approx 57$ nm. It also shows that, under sharp enough bending constraints, DNA drops its homogeneities in form of localized basepair disruptions to reduce overall bending energy.

Chapter 5

Micromechanical properties of DNA with nicks and mismatches

5.1 Introduction

Here, we examined the effects of naturally occurring structural damages on DNA behaviours through investigating their micromechanical properties. Although DNA damages appear in different forms and degrees, basically, they can be categorized into backbone breakages (*i.e.*, nicks) and base alternations (*i.e.*, mismatches, where deletions and insertions are considered as their more extreme versions). Regarding nicks, the same procedures as nick-free DNA were applied to obtain corresponding $\Delta\mathcal{A}(d)$ and $f(d)$ for nicked DNA. We found that it behaves similar to normal B-DNA under weak bending, but promotes defect generations, progressions right at the nick position under strong bending. Moreover, lowing temperature strongly suppresses this defect excitation at nicked site. This nick-dependent bending responses turn out to be critical in interpreting experimental evidences, such as j-factor. In the case of mismatches, the mismatched basepairs are permanent "defects" by default, and the distinctive bending behaviours for various types of mismatched DNA were revealed even by their unconstrained dynamics. In general, they disrupt normal non-covalent interactions in various degrees, leading to localized softening or intrinsic kinking.

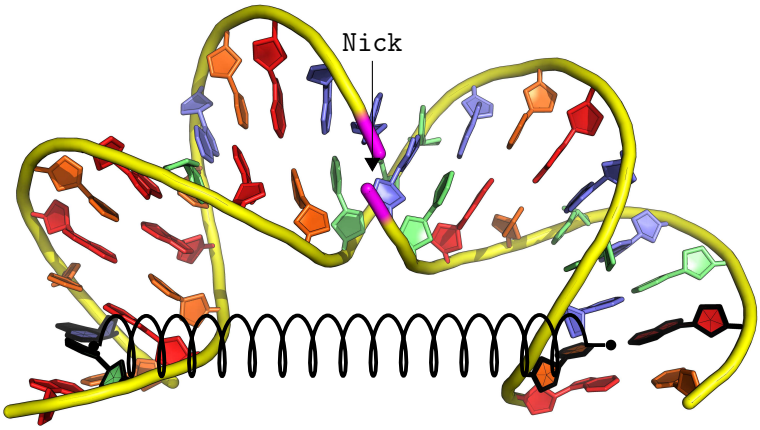
5.2 Effects of nicks on DNA bending

DNA nicks, which are single-strand discontinuities in DNA backbone, are the most commonly occurring DNA natural damages *in vivo*. They are generated under assorted physical stresses, such as UV lights, radiations or ultrasounds; as well as by chemical attacks, for instance, enzyme cleavages by restriction endonuclease and topoisomerase. More importantly, these nicks are biologically functional. For instance, during replication, torsion constraints in supercoiled DNA are released through nicks creased by topoisomerase, and in turn enhance the associations of polymerase to DNA. Recent evidences revealed positive correlations between nicks and protein-DNA interactions. The nicked sites may function as binding precursors and facilitate protein reorganizations, for example, recruiting reparation complex during DNA damages repair. On the other hand, as mentioned in Subsection 1.4.2, nicks may help to induce defects under sharp bending, which lead to the breakdown of Ω boundary condition and explain the anomalously high looping probabilities obtained in DNA looping assays. As a result, it is our interest to investigate the effects of nicks on DNA mechanical behaviours, through observing its deformation morphologies and sampling its configuration space during bending using MD simulations.

5.2.1 Introducing nicks in initials

Nicks are commonly generated during natural damages and enzyme actions by breaking the weakest covalent bonds in DNA backbones. In order to investigate the effects of nicks, we disrupted the phosphodiester bonds of phosphate group (PO_4^{3-}) between nucleotides to induce the discontinuity in a single strand. More specifically, starting from our uniformly bent initial in Subsection 3.3.1, the phosphate group between i^{th} and $(i+1)^{\text{th}}$ nucleotides on Strand I was completely deleted to remove phosphodiester bonds on both sides, while leaving O3' on i^{th} and O5' on $(i+1)^{\text{th}}$ deoxyriboses hydrolyzed to form hydroxyl groups. This is referred as nick after i^{th} basepair in the rest of this chapter. And the DNA initial with nick after 11^{th} basepair is shown in Figure 5.1, with the nicked site zoomed and highlighted using magenta colour.

(a) Nicked DNA initial with nick after 11th basepair



(b) Atomic stucture of nicked site

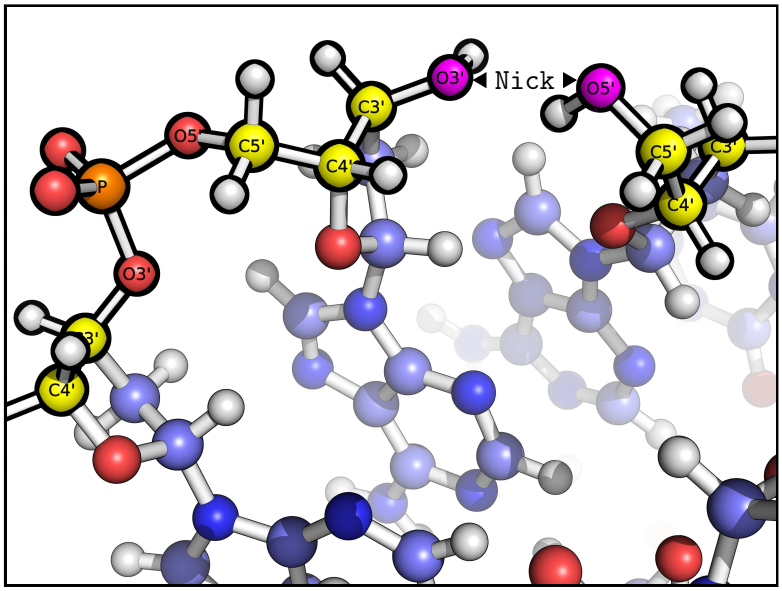


Figure 5.1: Nicked DNA constructs with nick after 11th basepair. (a) An smoothly bent DNA initial containing a nick between 11th and 12th basepairs in Strand I, highlighted by magenta. (b) Zoom in at the nicked site, where the phosphate group has been removed, leaving the O3' and O5' atoms (magenta) hydrolyzed. The backbone carbon atoms are highlighted by yellow.

In order to directly observe the structure evolutions of nicked DNA under bending, similar simulations as those in Chapter 3 were performed. Nicks were introduced into different locations along DNA, and compressional forces were exerted by attaching springs to their second and last-second basepairs. So, following the same setup procedures described in Subsection 3.3.1, different initials with

nick after i^{th} basepair, for $i = 6, 8, 11, 13$, were centered at the same rhombic dodecahedron unit cell used before, constrained by the zero length contractile spring with harmonic potential of $\mathcal{V} = \frac{\kappa}{2}d^2$. Here, $\kappa = 28.2$ pN/nm is used. All four MD simulations were carried out under NTV ensemble at constant temperature of 300 K and volume of ~ 1170 nm^3.

Later, in order to further explore the micromechanical properties of nicked DNA, sampling process was conducted on the particular nicked DNA with nick after 11^{th} basepair using similar umbrella sampling introduced in last chapter. Again, twelve pre-bend DNA initials with different end-to-end distances, in the range of ~ 2.6 to 5.4 nm, were selected from its previous structural bending simulations, and denoted by corresponding end-to-end distance l_m in decreasing order, for $m = 1, 2, \cdots, 12$. They were centered at the same unit cell, constrained by the finite length contractile springs with harmonic potential of $\mathcal{V}_m = \frac{\kappa_{\text{u}}}{2}(d - l_m)^2$. Here, $\kappa_{\text{u}} = 248.9$ pN/nm is used. Then, twelve parallel MD simulations were executed under same NTV ensemble at 300 K. Moreover, 290 K NTV ensemble umbrella sampling was conducted through seven l_n-constrained simulations to investigate the temperature effects on nicked DNA micromechanics. Biased potentials $\mathcal{V}_n = \frac{\kappa_{\text{u}}}{2}(d - l_n)^2$ were applied, where l_n gradually reduces in a smaller range of ~ 4.2 to 5.4 nm, for $n = 1, 2, \cdots, 7$.

5.2.2 Nicks direct defect excitations

Total of four destructive conformational evolutions have been obtained from simulations for nicked DNA with nicks located at different positions alone the polymer, which are between the 6^{th} and 7^{th}, 8^{th} and 9^{th}, 11^{th} and 12^{th}, and 13^{th} and 14^{th} basepairs, explicitly. These MD simulations were executed for 70 ns individually. Based on their end-to-end distance dynamics, they all developed defects, resulting in equilibrated $\langle d \rangle < 1.7$ nm, under $\kappa = 28.2$ pN/nm within nanosecond timescale. Their equilibrated conformational representatives from last 20 ns were extracted for further analysis.

Obviously, the non-covalent interactions at nicked sites are generally easier to break than intact form, due to the lack of particular covalent linkages in backbone. By looking at selected snapshots, we confirmed that the defects are

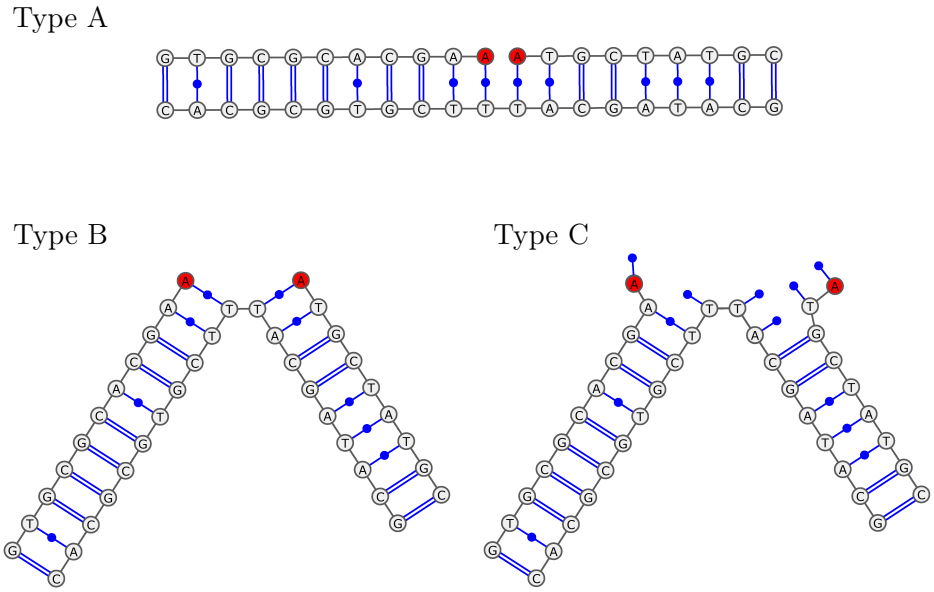
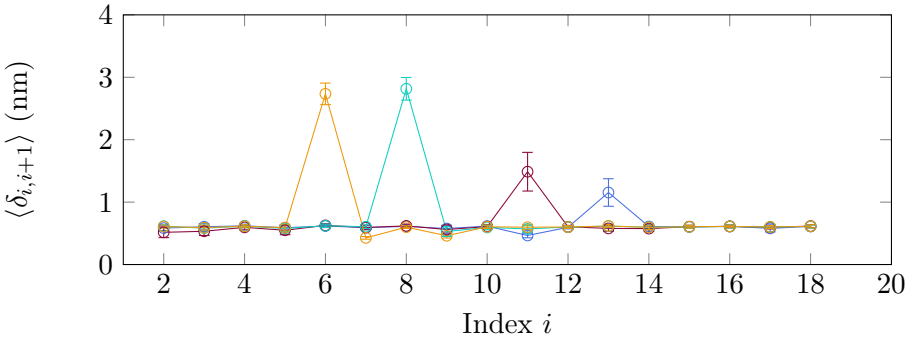


Figure 5.2: Illustrations for nicked DNA with different categories of non-covalent topologies. Type A shows the intact nicked DNA with both intact hydrogen bonding and basepair stacking. Type B represents the unstacked nicked DNA with disrupted basepair stacking only at nicked position. Type C indicates a particular case of the peeled nicked DNA with both nicked ends split, resulting in both disrupted base-stackings and base-pairings around nicked site. This figure uses planner structures with nick after 11^{th} basepair as demonstrations.

generated right at those nicked sites. But, this kind of defects are more complex than the basepair disruptions observed in Chapter 3. As a reminder, the constitutive failures occurred during nick-free DNA bending are in form of localized and correlated hydrogen bonds and basepair stacks disruptions, refer to Figure 3.6, 3.7 for more details. On the contrary, for the defects at nicks, the localized interruptions of those interactions are not necessarily coupled, and can even lead to severer forms through strand separations. So, the hydrogen bonding and basepair stacking profile analysis used before is not sufficient here. The observed topologies of nicked DNA can be categorized into three types, as shown in Figure 5.2 using nick after 11^{th} basepair as example. Type A represents intact nicked DNA without any unstacked nor unpaired non-covalent interactions. While type B, C are nicked DNA with defects, where type B is unstacked at the nicked site alone with all hydrogen bonds remained, and type C is peeled from the nicked ends. Note that the strands in type C can split into two with various

degrees from both ends, and the flapping *ss*DNA overhangs can adhere to itself randomly. As we can see, hydrogen bonding profiles cannot distinguish between type A and type B, and type C complicates both hydrogen bonding and basepair stacking profiles, especially for the latter one.

(a) Basepair distance profiles for nicked DNA



(b) Hydrogen bonding profiles for nicked DNA

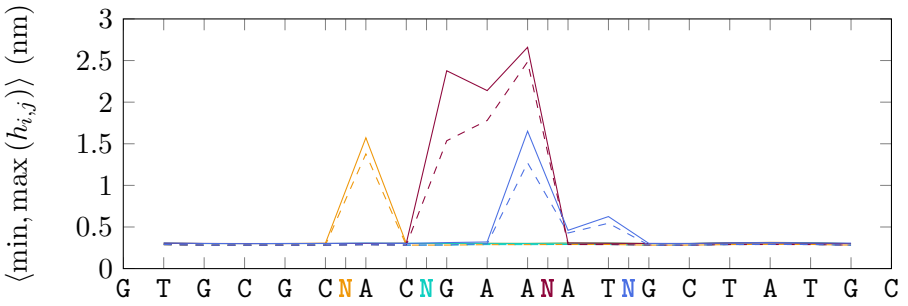


Figure 5.3: (a) The basepair distance profiles, $\langle \delta_{i,i+1} \rangle$, which measure the equilibrated distances between adjacent C4' atoms of i^{th} and $(i+1)^{\text{th}}$ basepairs on top strand scanned through entire DNA, for simulations with nick after i^{th} basepair, where $i = 6, 8, 11, 13$. The dramatic increased $\langle \delta_{i,i+1} \rangle$ in corresponding nick-containing simulations reveal that the disruptions of basepair occurred at nicked sites. Note that C4' atoms of deoxyriboses are part of the DNA backbone. See Figure 5.1(b) for their exact locations. (b) The Hydrogen bonding profile, $\langle \min, \max(h_{i,j}) \rangle$ *vs.* $i = 2, 3, \cdots, 19$ averaged over the last 20 out of 70 ns trajectories for four independent simulations with nick right after the 6^{th}, 8^{th}, 11^{th} and 13^{th} basepair steps. These hydrogen bonding profiles further reveal the existence of two distinctive types of disruptions: clean unstacking at nicked site in the case of nick after 8^{th} basepair step, and unstacking accompanied by peeling from nicked sites in the rest cases.

Then, we used an additional approach to quantify the locations of defects in equilibrated states for nicked DNA under bending. Besides previously detailed analysis, the distances between adjacent basepairs were indirectly measured. For each simulation, the inter-base distance between the adjacent C4' atoms along the backbone of the nicked strand, $\delta_{i,i+1}$, was monitored. Here i indexes the

position of the C4' atoms counted from the 5′ end of the nicked strand. For all the four nicked DNA, sharp bending led to significantly increased $\langle \delta_{i,i+1} \rangle$ that straddle the nick, indicating separation of the two nick-straddling C4' atoms and their associated bases [Figure 5.3(a)]. Among them, only the case with nick after 8^{th} basepairs kept all its hydrogen bonds intact, as shown in Figure 5.3(b).

Combining these basepair distance profiles with hydrogen bonding profiles, we can conclude that the defects in nick-containing DNA is caused by either unstacked basepairs straddling the nick without hydrogen bond disruptions (in the case of simulation with nick between 8^{th} and 9^{th} basepair; type B) or by strand separation involving a few melted basepairs near the nick (in the cases of the rest three simulations; type C). The selection between the two types of defects depends on the sequence of the two nick-straddling basepairs, with `GC` basepairs prone to unstack whereas `AT` basepairs prone to peel.

To conclude, the $\kappa = 28.2$ pN/nm contractile spring excites defection inside nicked DNA. The defects are directed by fragile nicks regardless of nick's location along DNA (*i.e.*, nicks after i^{th} basepair, $i = 6, 8, 11, 13$). The defects around nicked sites differs those observed in nick-free DNA, and are in forms of unstacking or peeling.

5.2.3 Nicks promote localized sharp bends

Further analysis shows that the separation of the two nick-straddling C4' atoms is accompanied with a large bending angle developed at the nicked position that relaxed the rest of DNA into a less bent *B*-form conformation. This is demonstrated in Figure 5.4(a) using the nick located between the 8^{th} and 9^{th} basepairs as an example. In the sharply bent conformation, the 8^{th} and 9^{th} basepairs are unstacked, causing the increase in $\delta_{8,9}$. The bending angle between the 7^{th} and 10^{th} basepairs, $\theta_{7,10}$, rapidly increased from the initial value of $\sim 30^\circ$ to $\sim 150^\circ$ in 2 ns after simulation began, synchronized with the increase in $\delta_{8,9}$. It is also synchronized with relaxations of the three-basepair-step bending angles in the rest of DNA to more straight conformations, as shown by $\theta_{4,7}$ and $\theta_{10,13}$ dynamics.

For another example, similar nick promoted localized sharp bend was also observed for the case of peeling around the nick [Figure 5.4(b)], using the nick

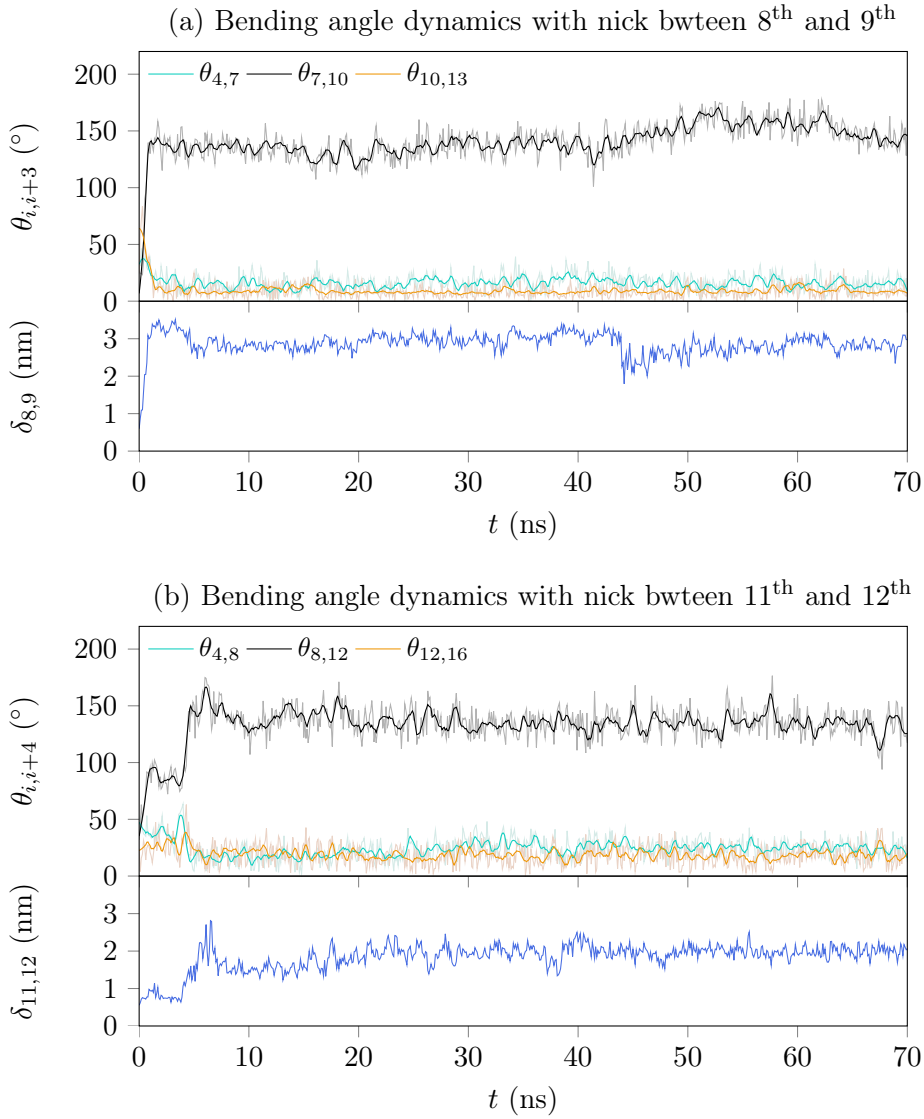


Figure 5.4: 70 ns dynamics of local bending deformations and basepair separations at nicked sites under $\kappa = 28.2$ pN/nm. (a) Row 1: time evolution of $\theta_{7,10}$ (black) enclosing nicked site between 8^{th} and 9^{th} basepairs, which shows the kink development around unstacked region. The bending angle evolutions of two intact regions with same length, $\theta_{4,7}$ (cyan) and $\theta_{10,13}$ (orange), are shown for comparison. Rows 2: time evolutions of $\delta_{8,9}$ (dark blue) indicates basepair separation at nicked sites. (b) Similar dynamics of kink development ($\theta_{8,12}$, black), bending relaxation ($\theta_{4,8}$, cyan; $\theta_{12,16}$, orange) and basepair separation ($\delta_{11,12}$, dark blue) for the peeled DNA with nick between 11^{th} and 12^{th} basepair.

located between the 11^{th} and 12^{th} basepairs that caused by disruptions of hydrogen bonds in adjacent 11^{th}, 10^{th}, 9^{th}, and 8^{th} basepairs. The development of a large bending angle around the nicked position is synchronized with the relaxation of the rest of DNA to a less bent *B*-form conformation as well. Overall, these nick-dependent excited defects induce global shape changes to reduce DNA total bending energy, in the same manner as intrinsic defects in nick-free DNA.

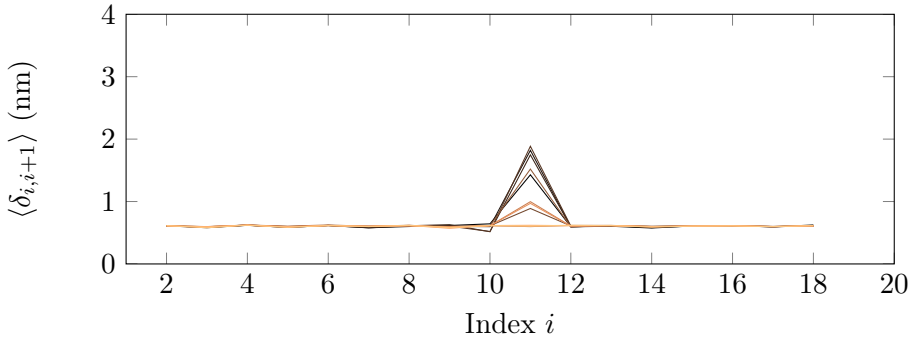
5.2.4 Micromechanical properties of nicked DNA

The re-weighting sampling for nicked DNA was attained through umbrella sampling using twelve independent MD simulations, which were initiated from selected DNA conformations with nick after 11^{th} basepair, confined using l_m length springs with spring constant of κ_{u}, and run for 50 ns each. Their corresponding configuration samples in last 20 ns were collected for further analysis. Then, they were assembled together to obtain the free energy difference profile and force-extension curve for nicked DNA.

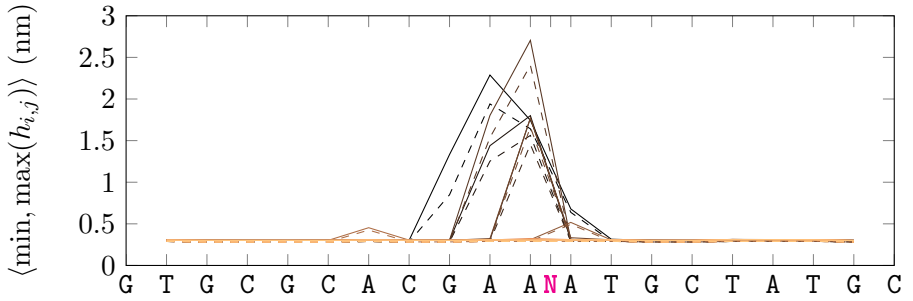
Figure 5.5(a) plots the basepair distance profiles for these twelve simulations. Nine of them developed basepair dissociations at their nicked sites, indicated by enlarged $\langle \delta_{11,12} \rangle$ averaged over final 20 ns ($m = 4, 5, \cdots, 12$). Figure 5.5(b) shows the hydrogen bonding profiles for these twelve simulations. Seven of them remained all hydrogen bonds in most of their last 20 ns dynamics ($m = 1, 2, \cdots, 7$). Together, the three under weak bending stay in B-form ($m = 1, 2, 3$, *i.e.*, type A); while next four under moderate bending contain unstacks after 11^{th} basepairs, but preserve their hydrogen bonding ($m = 4, 5, \cdots, 7$, *i.e.*, type B). And the other five under severe bending further accommodate disrupted hydrogen bonding at $9^{\text{th}} - 12^{\text{th}}$ basepairs around the nicked site ($m = 8, 9, \cdots, 12$), indicating ends peeling started from the nick (*i.e.*, type C). Compared against those for normal DNA [*i.e.*, cases for $k = 11, 12, 13$ in Figure 4.1(a)], the region with interrupted pairing shifted ~ 2 bp left due to the presence of nick.

The biased distributions of end-to-end distances, $\rho_{\{m\}}(d)$, for nicked DNA cover the region between ~ 2.3 to 5.7 nm in Figure 5.5(c). It is worth mentioning that the insufficient sampling at gap near ~ 4 nm can be overcame by simply conducting more simulations from suitable initials, because the energy barriers are relative small for peeling from nicks *vs.* defect excitations for B-DNA. But this additional information is irrelevant to subsequent results and discussions. Then, based on these statistics, the unbiased probabilities were evaluated at 200 values of d using `g_wham` in the regions of ~ 2.2 to 4.0, ~ 3.9 to 5.4 and ~ 4.8 to 5.7 nm, respectively for type C, B, A nick-containing DNA.

(a) Basepair distance profiles for nicked DNA



(b) Hydrogen bonding profiles for umbrella sampling



(c) Direct probability density functions for umbrella sampling

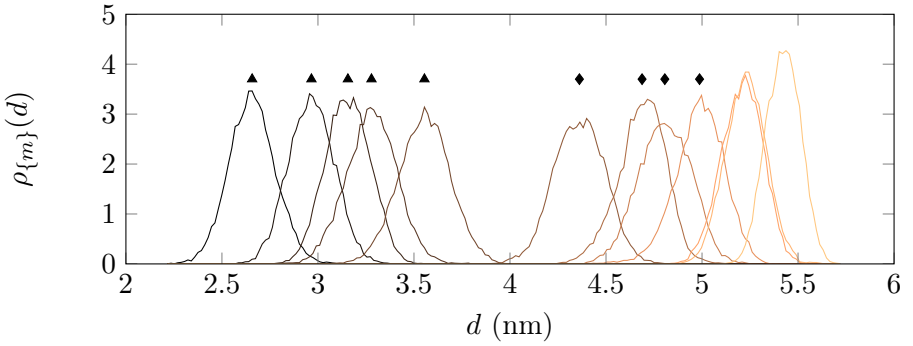


Figure 5.5: 300 K, umbrella sampling for nicked DNA with nick after 11$^{\text{th}}$ basepair, which consists of twelve independent simulations constrained by different potentials (a) Basepair distance profiles plot adjacent C4' $30-50$ ns averaging distances, $\langle \delta_{i,i+1} \rangle$, along DNA. Nine most constrained DNA generate basepair dissociations around nick, while the other three under weak bending behaves as B-DNA at 11$^{\text{th}}$ and 12$^{\text{th}}$ basepair. (b) Hydrogen bonding profiles of their equilibrated conformations averaged over the last 20 ns. Solid lines show $\langle \max(h_{i,j}) \rangle$ and the dashed lines show $\langle \min(h_{i,j}) \rangle$. Five most constrained DNA contains peeled ends from nicked site, while the other seven are nicked DNA with intact hydrogen bonding. Note that N points out the location of nick. (c) Direct probability density functions against end-to-end distances for the twelve simulations based on corresponding 10000 samples. The distributions indicated with ▲ are from the five peeled nicked DNA; with ♦ are from the four unstacked nicked DNA, while the rest are obtained from the three intact nicked DNA. $\rho_{\{m\}}(d)$ overlap with each other in the entire d range from 2.3 to 5.7 nm. Lines are coloured from light to dark as intrinsic contractile spring lengths l_m decreases (*i.e.*, $l_m =$ 5.42, 5.10, 5.04, 4.83, 4.64, 4.42, 4.12, 3.45, 3.23, 3.01, 2.81, 2.62 nm respectively) in both (a), (b) and (c).

Based on Equation 4.1, the three discretized free energy difference profiles for type C, B and A nick-containing DNA were obtained in the same manner as before, through taking logarithm on unbiased probabilities, zeroing at global minimum state and shifting to match overlaps. Then their smoothed continuous versions were generated using cubic-spline interpolations at low variance ranges of $\sim$ 2.4to 3.8, $\sim$ 4 to 5.2, and $\sim$ 5.2 to 5.6 nm correspondingly. Together with the free energy difference profiles for DNA without nick, they were plotted in inset of Figure 5.6 for comparison. We noticed these $\Delta\mathcal{A}(d)$ profiles for nicked and nick-free DNA superimpose with each other near the same equilibrated length $d_0 \approx 5.43$ nm. As bending increases, the two profiles quickly deviate from one another. Generally speaking, $\Delta\mathcal{A}(d)$ for nicked DNA is lower, globally steady, but locally much frustrated, while $\Delta\mathcal{A}(d)$ for nick-free DNA is higher, smoother, but with a sudden slope drop at its transition coordinate $\sim$ 3.8 nm.

The force distance relationships for nicked DNA were worked out in two ways as before. The continuous force-extension curves were differentiated from continuous corresponding $\Delta\mathcal{A}(d)$, and the discretized $f\left(\left\langle d_{\{m\}}\right\rangle\right)$, which are equilibrated maintaining forces *vs.* end-to-end distances, were directly read out from individual l_m-restrained simulations. As we can see, the results from two approaches coincide with each other in main of Figure 5.6, which indicates that the sampling is statistically sufficient for entire d region. By comparing against the force-extension curves for normal DNA, for $d_0 > d > 5.2$ nm, $f(d)$ for type A nicked DNA (orange) exactly overlaps with that for intact nick-free DNA (dark red). As d further decreases, it transits to type B nicked DNA (cyan), and immediately deviates from the nick-free DNA force responses. Type B requests much smaller force to maintain certain end-to-end distances over $5.2 > d > 4$ nm against normal B-DNA. With $d < 3.8$ nm, the force profile for type C nicked DNA (pink) continuously drops to the level even lower than that for defected nick-free DNA (dark blue).

Based on further quantitative and structural analysis, under weak bending constraints, intact nicked DNA retains all of its non-covalent interactions and resembles normal B-DNA behaviours with estimated persistence length $\tilde{A} = 54.6$ nm (*i.e.*, based on Equation 4.2, using averaging points $f\left(\left\langle d_{\{m\}}\right\rangle\right)$, for $m = 1, 2$).

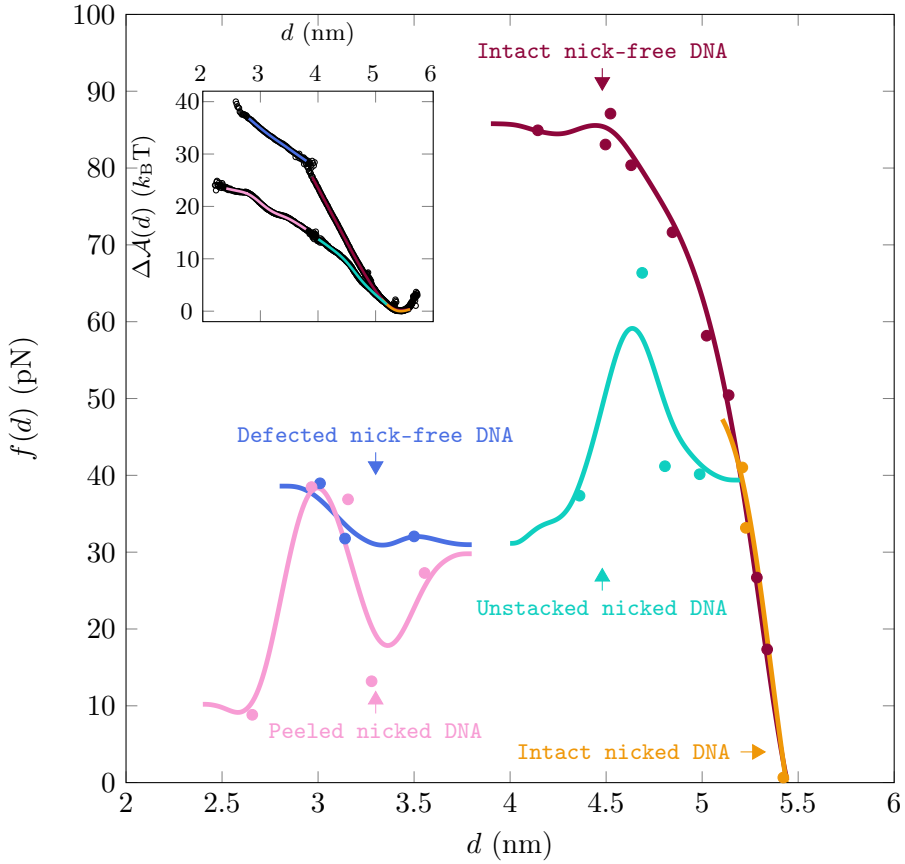


Figure 5.6: Free energy difference profile and force-extension curve for nicked DNA with defect excitations. Upper-left inset displays the discretized $\Delta\mathcal{A}(d)$, reference to global minimum state, separately obtained for intact (type A), unstacked (type B) and peeled (type C) nicked DNA are plotted against end-to-end distances. Smoothed and continuous versions for these three curves are achieved in smaller and more confident regions using interpolation (orange line for type A, cyan line for type B and pink line for type C). The main figure shows continuous $f(d)$ for type A, type B and type C in corresponding colours, which were calculated from differentiations of continuous $\Delta\mathcal{A}(d)$. The discrete points of $\langle f_{\{m\}}\rangle$ against $\langle d_{\{m\}}\rangle$ ($\bullet$ in corresponding colours) were directly read out from 12 simulations ($m = 1, 2, 3$ for type A, $m = 4, 5, \cdots, 7$ for type B and $m = 8, 9, \cdots, 12$ for type C). They generally overlaid with their $f(d)$ obtained using statistical approach. The profiles for normal DNA were plotted in both inset and main for comparison (dark red for intact, dark blue for defected nick-free DNA). Note that free energy difference profile for nicked DNA has the global energy minimal at $d_0 \approx 5.43$ nm as well. And its force-extension curve, which zeros at d_0, superimposes with that of intact nick-free DNA over a small region with $d < d_0$.

Then, under moderate bending, our intact nicked DNA can hold up to ~ 40 pN until unstacking occurs at the nicked site. It revealed a stacking Helmholtz free energy of $\Delta\mathcal{A}_{\mathrm{stack}} \approx -1.5\ k_{\mathrm{B}}T$, which is close to the experimental measurements using stacking-unstacking equilibration [108]. This event results in earlier defect excitation and leads to more bending adaptive unstacking nicked DNA, through

developing localized sharp bends. In contrast, intact nick-free DNA can hold up $\sim$ 90 pN (*i.e.*, even beyond buckling transition at $\sim$ 80 pN), and remains its non-covalent integrity till breaks near 3.8 nm. Under even stronger bending, the open ends at nicked site start to peel off, which leads to peeled nicked DNA. Although its force magnitude is overall similar to that of defected nick-free DNA, more flexible structures with severer forms of defects can be induced (*i.e.*, for instance, those structures with maintaining forces less than 20 pN).

5.3 Nicks and DNA looping experiments

In order to build linkages between our MD results with exist experimental evidences, we continue to investigate the effects of nicks on DNA bending micromechanics from the experimental point of view. Despite of practical challenges, many approaches have been applied to probe DNA behaviours under sharp bending conditions. Among them, DNA looping experiments are relatively quantitative and systematic to acquire DNA responses under bending sharper than random coil. For example, the anomalous bending elasticity of $\sim$ 94 bp DNA reported by Cloutier *at el.* was obtained through such ligase-based DNA looping measurements. However, the interpretations for these evidences are not definite, and even sometimes contradictory, especially at the bending level of interest, $\sim$ 100 bp minicircles. Here, we utilize our obtained MD results on DNA sharp bending responses to better elucidate those evidences. Moreover, we show that nicks play a non-negligible role in DNA looping experiments, and effects of nicks may dominate observed softening of sharply bent DNA.

5.3.1 Details and interpretations on j-factor measurements

Considering a DNA molecule with two ends denoted by "A" and "B", respectively. The DNA looping probability density, $\rho(\mathbf{0})$, is the probability density when the two ends of the same DNA molecule meet. In other words, $P_{\text{loop}} = \rho(\mathbf{0}) \cdot \delta V$ is the probability to find the end "B" in an infinitesimal volume δV in the vicinity of the end "A" of the same DNA. In the presence of the N identical DNA molecules in a total volume of V, the probability of finding an end "B" from another molecule

in the same vicinity around the end "A" of the target DNA is $\frac{N}{V}\delta V = c \cdot \delta V$. In order to determine $\rho(\mathbf{0})$, a cyclization approach has been proposed to chemically trap (such as ligation) the "A-B" ends in δV, which results in either looped DNA or dimerized DNA molecules with reaction rates of K_{loop} and $K_{\text{dimer}} = c \cdot K^0_{\text{dimer}}$, respective, where K^0_{dimer} denotes the dimerization rate per unit concentration of DNA. Theoretically, $\frac{K_{\text{loop}}}{K_{\text{dimer}}} = \frac{\rho(\mathbf{0})}{c}$, which results in experimental determination of looping probability density as: $\rho^{\text{E}}(\mathbf{0}) = \frac{K_{\text{loop}}}{K^0_{\text{dimer}}}$. The ratio $\frac{K_{\text{loop}}}{K^0_{\text{dimer}}}$ is often referred to as the "j-factor" [24, 3, 25].

A typical j-factor measurement approach is based on using ligase to covalently link the DNA ends. This method requires two short complementary *ss*DNA overhangs at the two DNA ends, which transiently hybridize the "A-B" ends into a conformation suitable for a subsequent ligation reaction. A prerequisite for such measurement is to achieve pre-equilibration between looped and unlooped molecule before the ligation reaction [24, 26]. This requires the hybridization interaction to be relatively weak, allowing reversible unlooping or undimerzation. It also requires the ligase concentration to be relatively low, ensuring end-sealing by ligase to be time limiting step. What measured in such experiments are rates of covalent closure of hybridized loop K'_{loop} and hybridized dimer K'_{dimer}. With an additional assumption that $\frac{K'_{\text{loop}}}{K'_{\text{dimer}}} = \frac{K_{\text{loop}}}{K_{\text{dimer}}}$, the looping probability density $\rho^{\text{E}}(\mathbf{0})$ can be determined.

Further, hybridization also imposes a constraint on the orientations of the hybridized "A-B" ends. In the case of dimerization, which does not involve any DNA bending or twisting, the hybridized DNA assumes a straight B-form conformation, as suggested in our MD simulation. This results in two requirements on termini orientations: (i) the hybridized ends "A" and "B" should be parallel to each other; (ii) the ends must be helically phase matching which constraints the axial twist degree of freedom of one molecule to the other (Figure 5.7). Together, these requirements imply the probability density of the hybridizable end "B" from another molecule is $\frac{c}{4\pi \times 2\pi}$. As a result, $\rho^{\text{E}}(\mathbf{0}) = \left(8\pi^2\right)^{-1} \frac{K_{\text{loop}}}{K^0_{\text{dimer}}}$. In Chapter 1, we have referred such boundary constraint as the "Ω" boundary condition.

Hybridization may also impose certain constraint on the orientations of the hybridized "A-B" ends in the hybridized looped DNA. In the case of large loop

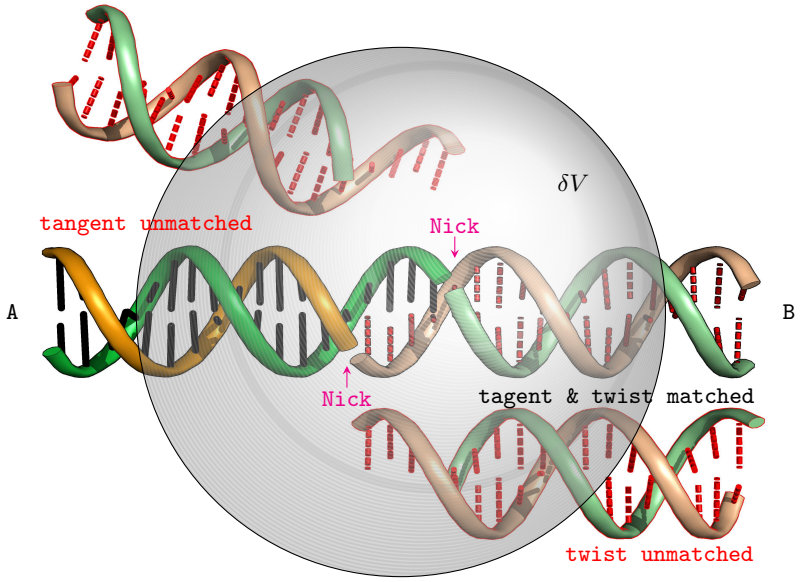


Figure 5.7: Ω boundary condition: In ligase based DNA looping experiments, within the infinitesimal volume, δV, around reference "A" end (with black solid basepairing), only a subset of entered complimentary "B" ends (with red dashed basepairing) can assemble into transiently stabilized hybridized "A-B" ends, and chemically trapped by a subsequent ligation reaction. Under the Ω boundary condition defined in the main text, it entails a $(4\pi \times 2\pi)^{-1}$ factor. Tangent unmatched (at top) and twist unmatched (at bottom) "B" ends are shown for comparison. Note that two pre-existing nicks (magenta arrows) are formed right after hybridization, which may cause violation of Ω boundary condition when DNA is sharply bent.

($L > A_{\mathrm{Tw}}$, where $A_{\mathrm{Tw}} \sim 100$ nm is the twist persistence length), the same Ω boundary condition should apply due to the decreased twist energy in large DNA loops. In order to extract the DNA micromechanical properties, such as persistence lengths, one should calculate the looping probability density based on the WLC model $\rho^{\mathrm{WLC}}(\mathbf{0})$ under the Ω boundary condition treating the DNA bending and twisting persistence lengths as free parameters, and compare it with the experimentally measured value $\rho^{\mathrm{E}}(\mathbf{0})$. However, for small DNA loops, the Ω boundary condition may not hold, due to the increased bending energy in the loops which may cause defect excitation at the nicks (as suggested in our simulations). In this case, $\rho^{\mathrm{WLC}}(\mathbf{0})$ should be calculated under a different boundary condition $\xi \neq \Omega$, which has been ignored in previous discussions.

As shown in Section 1.4, large-scale DNA looping probability densities from j-factor measurements (> 200 bp) well fit theoretical calculations based on WLC

model with A around 50 nm, which justify aforementioned Ω boundary condition imposed on looped ends. On the other hand, for shorter DNA fragments around 100 bp, the j-factor measurements reported a DNA looping probability density that are several orders of magnitude larger than that predicted by WLC model with $A \approx 50$ nm [3]. This disagreement between experiments and the WLC model prediction has led to a decade of confusion on its nature. There are two alternative possibilities that may cause such anomaly: (i) It is an intrinsic elastic response of *ds*DNA under sharp bending condition, which might be caused by bending induced flexible defect excited inside the DNA as proposed by several groups [3, 30, 4, 5]. (ii) The Ω boundary condition assumption is no longer valid for the hybridized looped DNA when DNA is sharply bent. In the latter case, Equation 1.13 used to extract the mechanical properties of DNA is invalid.

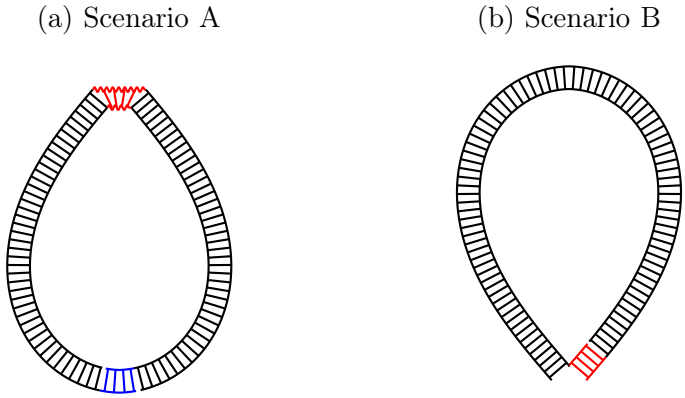


Figure 5.8: Two possibilities to interpret the observed abnormal DNA bending rigidities under sharp bending conditions in both ligase-based and FRET-based DNA looping experiments. (a) scenario A: DNA intrinsic basepair disruptions, and (b) scenario B: nick-dependent unstacking. In both cases, defects are excited with certain excitation energy consumptions (*i.e.*, $\mu_{\rm A}$ and $\mu_{\rm B}$, respectively), which lead to localized flexible structures with softer apparent persistence lengths (*i.e.*, $A'_{\rm A}$ and $A'_{\rm B}$, correspondingly). After these, kinks develop around defects, results in teardrop shapes with overall lower free energy. Note that $\mu_{\rm B} < \mu_{\rm A}$ and $A'_{\rm A} \sim A'_{\rm B}$, thus scenario B is much more probable.

Based on our MD results, we showed that B-form DNA is possible to develop $1-3$ bp correlated hydrogen bonding and basepair stacking disruptions upon sharp enough bending. This localized defect creates a hinge, which may in turn absorb bending to release overall bending energy [*i.e.*, scenario A, Figure 5.8(a)]. In last section, we demonstrated that a pre-existing nick may also lead to flexible

defect at the nicked location when the DNA is sharply bent, by either unstacking the two adjacent basepairs straddling the nick, or further strand peeling near the nick. This result suggests that the two pre-existing nicks in the hybridized loop may cause DNA kink in sharply bent DNA minicircles; thereby violating the Ω boundary condition [*i.e.*, scenario B, Figure 5.8(a)]. As we can see, two scenarios are both possible, and physically equivalent in DNA loops: they are flexible defect excitations in nature, which direct kink formations, and result in teardrop shapes to lower looping energy. However, more importantly, they are quantitatively different: the defect at nicked site requires much less excitation energy. Thus, in the presence of nick, scenario B will always occur prior to scenario A, and further suppress scenario A once kinks developed.

In summary, for ligase-based j-factor measurements where nicks cannot be avoid, defects generate at nicked site under sharp bending condition; and the looping probability density should be calculated based on a different boundary condition ξ. As shown in previous theoretical predictions [23, 4, 109], if the two ends of the same DNA can meet in a kinked conformation, the looping probability density is greatly increased compared to that under the Ω boundary condition (*i.e.*, $\rho_{\xi}(\mathbf{0}) > \rho_{\Omega}(\mathbf{0})$). In the extreme case when the two ends can meet with arbitrary angle, the looping probability density can be at least three orders of magnitude larger than that under the Ω boundary condition for < 100 bp DNA fragments. So, this nick-decedent mechanism naturally explained the j-factor measurement results for ~ 100 bp short DNA fragments reported by Cloutier *et al.* using ligation approach [3, 110].

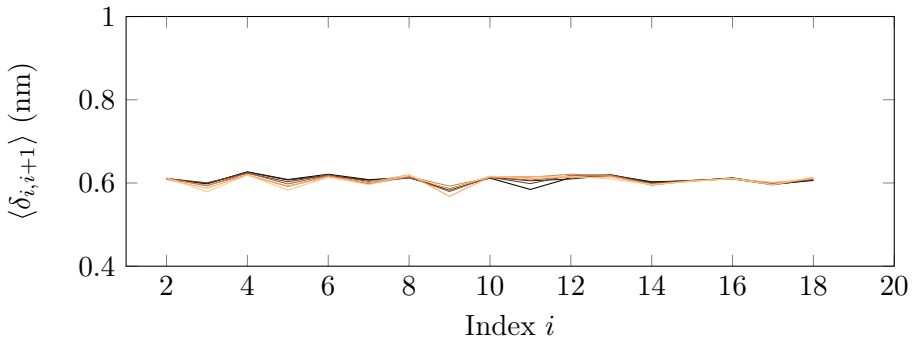
Besides ligase-based j-factor measurements, elastic anomaly of DNA was also revealed by analyzing the unlooping rate of a hybridized looped DNA or looping rate of an originally unlooped DNA using smFRET approach. Both assays suggested that there exist a critical contour length of DNA (< 100 bp), below which the dependency of unlooping (or looping) rate on DNA size indicates a softer DNA backbone than that predicted by the WLC model with 50 nm persistence length [29, 104]. Again, scenario B is not avoidable: the pre-existing nicks of the hybridized DNA loop may develop kinks at the nicked location, resulting in the observed anomaly.

5.3.2 Temperature sensitivities of unstacking at nicked site

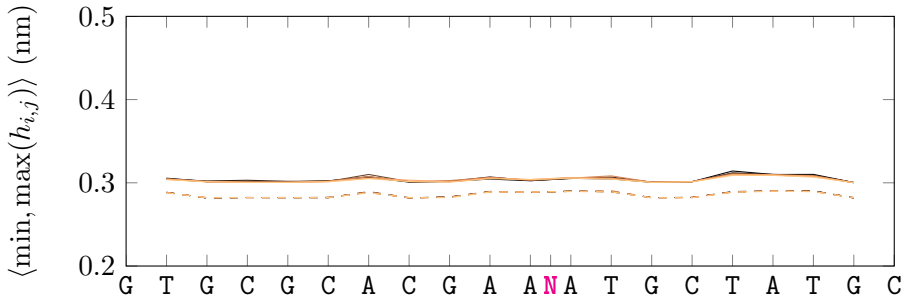
Right after Cloutier *et al.* reported abnormally high looping probability density on 94 bp DNA [3] in 2004, Du *et al.* [25] performed similar j-factor experiments on $\sim$ 100 bp DNA in 2005, which revealed expected $\rho^{\mathrm{E}}(\mathbf{0})$. As mentioned in Chapter 1, these contradictory j-factor measurements were likely caused by their different experimental temperature. It is well known that total-disruption free energy, $\Delta\mathcal{G}$, in case of DNA peeling roughly linear-dependent on both temperature and logarithm of salt concentrations [2, 11]: more specifically, increasing kinetic energy enhances basepair denaturation, while decreasing ionic strength induces strand separation (see Section 1.5). The same trends are also true for the $\Delta\mathcal{G}$ involved in unstacking at nicked sites [111, 112], where temperature dependency is slightly stronger than peeling (*i.e.*, larger $|\Delta S|$), due to the lack of additional backbone constraints. However, in both situations, the 9°C discrepancy in T only leads to less than 0.5 $k_{\mathrm{B}}T$ free energy change. As a result, the key question in understanding the temperature sensitive of DNA looping becomes: how such small difference in free energy induces drastic mechanical response switching?

To answer this question, we recall previous results, that homogeneous B-form DNA follows classical WLC behaviours; its $f(d)$ contains a steep linear region near equilibrium length with force less than 80 pN, and a much flattened region after buckling transition with force larger than 80 pN. Furthermore, at 300 K, basepair disruption inside nick-free DNA occurs in the flattened region; on the contrary, defect excitation at nicked site happens in the steep linear region (Figure 5.6). So, for unstacking at nick, small perturbation in free energy leads to huge jump in required bending constraints for defect excitation, because of small magnitude of force, as well as the large Young's module $\sim$ 300 pN/nm. On the other hand, for intrinsic defect, nine degrees temperature difference hardly shifts its critical failure load. In other words, stacking-unstacking switching at nicked site is much more sensitive to T, in term of its mechanical response. Thus, the unstacking around nick is a more possible hypothesis to interpret the observed transformation between abnormal and normal bending rigidities upon small temperature change.

(a) Basepair distance profiles for nicked DNA



(b) Hydrogen bonding profiles for umbrella sampling



(c) Direct probability density functions for umbrella sampling

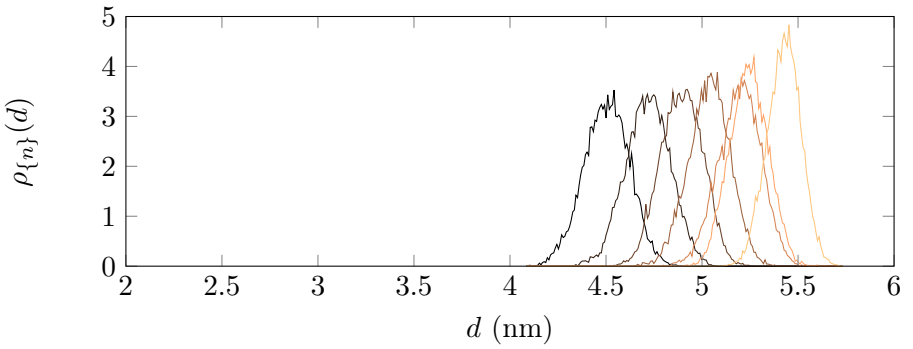


Figure 5.9: 290 K, umbrella sampling for nicked DNA with nick after 11$^{\text{th}}$ basepair, which consists of twelve independent simulations constrained by different potentials (a) Basepair distance profiles plot adjacent C4' $80-100$ ns averaging distances, $\langle \delta_{i,i+1} \rangle$, along DNA. All seven constraint simulations under weak to moderate bending have normal basepair distances between 11$^{\text{th}}$ and 12$^{\text{th}}$ basepair. (b) Hydrogen bonding profiles of their equilibrated conformations averaged over the last 20 ns. Solid lines show $\langle \max(h_{i,j}) \rangle$ and the dashed lines show $\langle \min(h_{i,j}) \rangle$. All seven simulations maintained their Waston-Crick basepairing. Note that **N** points out the location of nick. (c) Direct probability density functions against end-to-end distances for the twelve simulations based on corresponding 10000 samples. $\rho_{\{n\}}(d)$ overlap with each other in the entire d range from 4.1 to 5.7 nm. Lines are coloured from light to dark as intrinsic contractile spring lengths l_n decreases (*i.e.*, l_n = 5.42, 5.12, 5.05, 4.83, 4.64, 4.42, 4.19 nm respectively) in both (a), (b) and (c).

To investigate aforementioned temperature sensitivities of nicked DNA bending responses, the similar umbrella sampling for nicked DNA as Subsection 5.2.4 was conducted, but at 290 K, ten degrees lower than before. Independent MD simulations were started from same initial DNA atomic structures with nick after 11^{th} basepair, and constrained by same springs with same κ_{u}. As we are interested in the bending responses near previous nick disruption transition coordinate ~ 5.2 nm, only seven constraint sampling with l_n gradually decreases from $\sim d_0$ to 4.2 nm were conducted, for $n = 1, 2, \cdots, 7$. To compensate the slowness of evolutions at lower temperature, each confined simulations was simulated 50 ns longer, and corresponding data in 80 to 100 ns were collected for analysis.

The resultant basepair distance profiles, $\langle \delta_{i,i+1} \rangle$, [Figure 5.9(a)] and hydrogen bonding profiles, $\langle \min, \max (h_{i,j}) \rangle$, [Figure 5.9(b)] for these seven independent simulations reveal intact basepair associations between $11^{\text{th}} - 12^{\text{th}}$ basepairs, as well as intact hydrogen bondings. It means that nicked DNA remains in type A at 290 K under a wide range of bending, even keeps its integrity at $d \approx 4$ nm. This is in sharp contrast to its bending responses at 300 K, where transition from type A to type B occurs at only $d \approx 5.2$ nm within 50 ns. Apparently, the ten degree temperature reduction significantly altered the stacking-unstacking switching behaviours. The biased distributions of end-to-end distances, $\rho_{\{n\}}(d)$, for intact nicked DNA overlap with each other and cover the region between ~ 4 to 5.7 nm in Figure 5.9(c). The unbiased probabilities were evaluated at 200 values of d using `g_wham` over entire region.

Subsequently, discretized $\Delta\mathcal{A}(d)$ for intact nicked DNA at 290 K was calculated through negative logarithm based on Boltzmann distribution, and continuous $\Delta\mathcal{A}(d)$ was achieved by spline interpolation at a more statistically confident region from 4.2 to 5.6 nm. Then, its $f(d)$ at 290 K were obtained using derivative and direct read-out (Figure 5.10, dark blue solid lines and $\bullet$, correspondingly). The resulting force-distance curve nearly overlaps with that obtained for normal B-form DNA (dark red dotted line and $\bullet$) including the region after the buckling transition. This suggests nicked DNA resembles nick-free DNA over wider d at lower temperature. The $f(d)$ of intact nicked DNA at 290 K is slightly below that of intact nick-free DNA at 300 K; this is due to the temperature dependency of

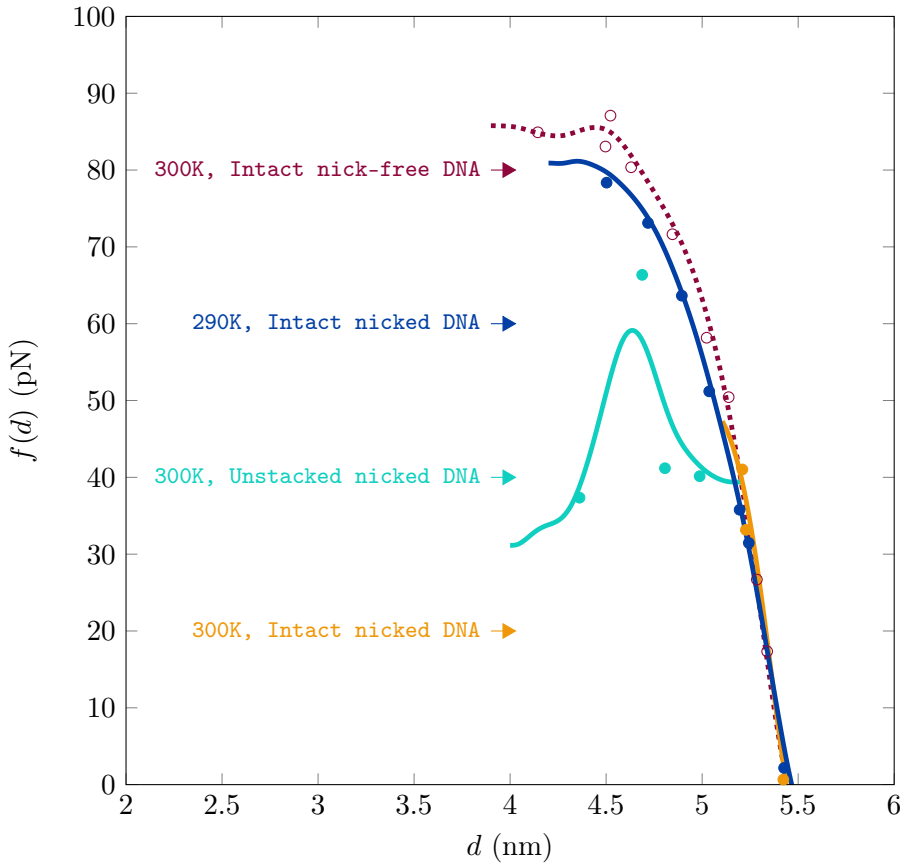


Figure 5.10: Temperature sensitivities of free energy difference profile and force-extension curve for nicked DNA. Continuous $f(d)$ for intact nicked DNA at 290 K (dark blue solid line), which were calculated from differentiations of its continuous $\Delta\mathcal{A}(d)$. The seven discrete points of $\langle f_{\{n\}} \rangle$ against $\langle d_{\{n\}} \rangle$ (dark blue •), for $n = 1, 2, \cdots, 7$, were directly read out from springs. The force responses of nicked DNA at 300 K over the same d region were plotted for direct comparison, where $f(d)$ for intact and unstacked nicked DNA at 300 K were coloured in orange and cyan, correspondingly. Their compelling differences reveal high temperature sensitivity in nick-dependent flexible defect excitations, especially in form of critical compressional load (*i.e.*, altered from $\sim$ 40 pN at 300 K to more than 80 pN at 290 K). The profile for intact nick-free DNA at 300 K was plotted in dark red for further comparison, which nearly overlaps with that of nicked DNA at 290 K. It double confirms the B-DNA like elasticity of intact nicked DNA, even after Euler instability at low T.

the buckling transition force given by $f_c = k_B T \pi^2 A / L^2$. Compared against $f(d)$ of nicked DNA at 300 K (orange and cyan solid lines and •), we finally quantified the mechanical responses of stacking-unstacking switching: intact nicked DNA can resist up to $\sim$ 40 pN compressional load at 300 K; on the other hand remains unstacked even under $>$ 80 pN at 290 K, due to the large Young's modulus $Y \approx 300$ pN/nm^2, and relative low failure loads. This result indicates that the lowered temperature strongly suppresses the nick-dependent flexible defect

excitations, which supports our hypothesis and explains the contradictory experimental results. Note that this is another strong evidence to justify scenario B, instead of scenario A.

5.4 Effects of mismatches on DNA bending

DNA mismatches are base alternations, which in general include insertions, deletions and misincorporations of basepairing. They are generated all the time during cellular mechanisms, such as replications or recombinations, as well as induced by environmental factors, such as chemical or physical stresses. For instance, during replication by DNA polymerases, mismatches are produced in the rate of $\sim 10^{-4} - 10^{-5}$ errors per basepair, while subsequent proofreading by exonuclease associated with polymerases achieve a final rate of $\sim 10^{-7}$ [113, 114]. Then, remaining mismatches usually result in either reparations or mutations, whose ratio is crucial to the balance between genetic maintenances and evolutions. Obviously, such misincorporations naturally disrupt the *B*-form secondary structures, and give rise to abnormal DNA elasticities and/or morphologies. Here we are going to investigate their effects on DNA micromechanical properties, which are believed to play a critical role in their functional processes, such as initiation of mismatch repairing by MutS in *E. coli* [115].

In order to study the effects of mismatches, several kinds of mismatching were introduced into the middle of short DNA fragments. Because the misincorporations of basepairing are permanent defects by themselves, their rigidities are intrinsically different, which lead to distinctive apparent angular movements even under normal conditions. As a result, we omitted the previously used bending-constraint excitation processes, directly simulated their non-constraint trajectories, and monitored corresponding bending angle fluctuations.

To speed up the simulations and minimize the likelihood of peeling, a $9 - 10$ bp GC-rich short DNA template as below,

$$\begin{array}{c} 5' - \texttt{CGCGXGCGC} - 3' \\ 3' - \texttt{GCGCXCGCG} - 5' \\ \uparrow\ \uparrow \\ \texttt{L}\ \ \texttt{R} \end{array} \tag{5.1}$$

, were used, where **X/X** position was artificially filled by various versions of mismatches **G/G**, **C/C**, **E/E**, double **E/E** and by normal basepairing **C/G** as control. Here, **E** represents the abasic site, which is a kind of spontaneously occurring lesions in DNA [116], whose nitrogen bases of nucleotide is completely deleted, leaving with flexible backbone linkage and disturbed non-covalent interactions. In the case of **E/E**, two "pairing" abasic sites were introduced to **X/X** position and formed an "empty" basepair. For double **E/E**, two consecutive such "empty" basepairs were put at the center of our DNA sequence.

Five DNA molecules in straight shape, based on above designed sequences, were produced using `X3DNA` with some manual editing. They were centered and aligned at corresponding minimal enclosing cuboid boxes. Then, each simulation unit was properly prepared in a similar way as our previous unconstrained 20 bp DNA according to the descriptions in Section 3.2. Note that the force field parameters for α-anomer abasic site, including structures, partial charges and atom types [117, 118], were integrated into `GROMACS` Parm99 for topology generations. Finally, for every sequence, a 100 ns free dynamic trajectory was simulated under NVT ensemble, with constant temperature of 300 K and sequence specific constant volume ranging from $\sim 200 - 260$ nm^3.

The mismatch-related bending angle fluctuations were extracted from each trajectory for further analysis. The two 4 bp `GC`-tails beside the "defects" remained their integrities during 100 ns simulations, due to their relative strong basepairing. So, we were able to attach the right-handed basepair reference frames to the intact neighbouring basepairs at the left and right of mismatches (*i.e.*, denoted by L, R in Equation 5.1). The instant bending angle was calculated from their deflected z-axes, using $\theta_{\mathrm{L,R}} = \cos^{-1}(\hat{z}_{\mathrm{L}} \cdot \hat{z}_{\mathrm{R}})$. Intuitively, the equilibrated mean of $\theta_{\mathrm{L,R}}$ fluctuations, if other than zero, reports the preferred kinking, while its variance reflects the rigidities of enclosing `X/X`. For each sequence, we derived their bending energy profiles against bending angle, by eliminating the entropic contributions from lateral movements, as,

$$E_{\mathrm{L,R}}(\theta) = -\ln\left(\frac{\rho_{\mathrm{L,R}}(\theta)}{\sin\theta}\right) + E_{\mathrm{offset}} \tag{5.2}$$

, where E_{offset} constant zeros the corresponding energy minimum for individual cases, as shown in Figure 5.11.

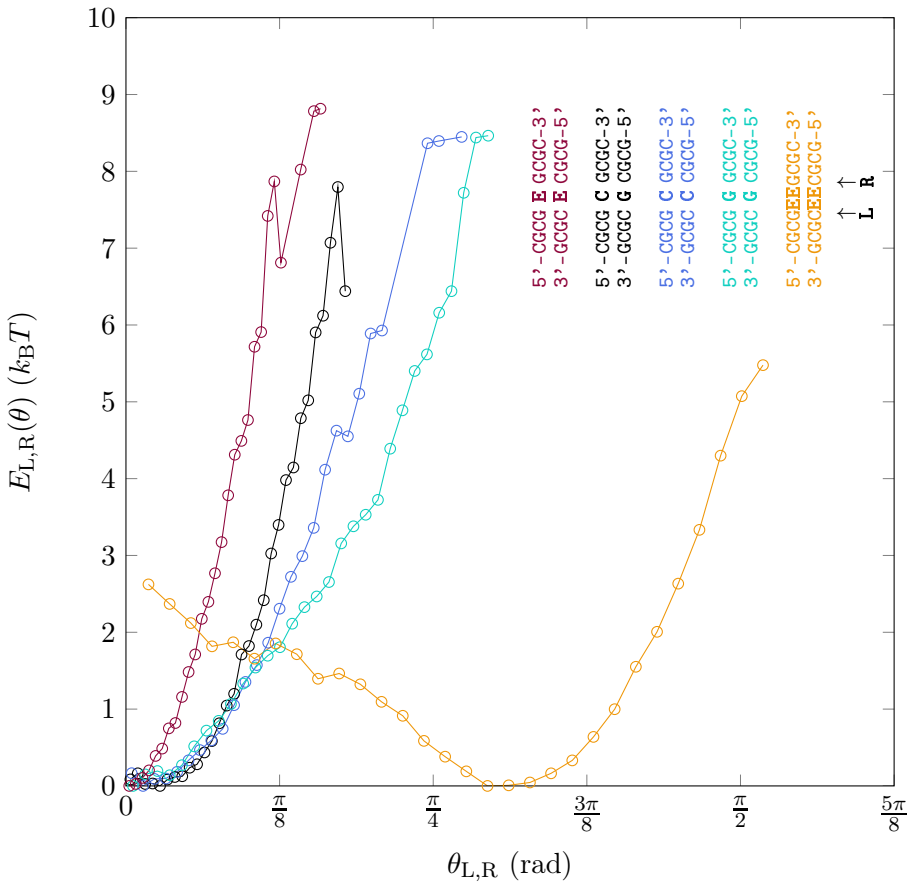


Figure 5.11: Bending energy profiles against the bending angles of most adjacent basepairs at the two sides of the "defect" of interest. Local bending energy curves were calculated form the equilibrated bending angle distributions, which were sampled using $70-100$ ns deflection angle ($\theta_{\text{L,R}}$) fluctuations. These profiles for enclosing `X/X` substituted by `E/E`, `C/G`, `C/C`, `G/G` and double `E/E` are coloured in dark red, black, dark blue, cyan and orange, respectively. The bending energies follow the typical WLC quadratic relationships, $E_{\text{L,R}}(\theta) \propto \theta^2_{\text{L,R}}$, ranging from rigid to flexible, as shifting from left to right, except for the case of double `E/E`. The two consecutive abasic sites lead to a intrinsic kinking of $\sim 60°$.

The mismatching bending energy profiles were hard to be precisely quantified at such small length scales. The misincorporations of basepairs permit more degrees of freedom for the relative arrangements of nearby basepairs, which result in more frustrated energy landscapes. And this give rise to much harder quantifications by MD simulation method, because of larger fluctuations, transient lock in at local energy minimal states, and significant increases of required simulation timescales. Furthermore, the short length scales also cause several troubles. Firstly, these local behaviours are much more diverge, while the large-scale quan-

tities are more reliable, such as end-to-end distances over 20 bp measured before, due to averaging effects. Secondly, as the distances are close to the DNA discretization limit (*i.e.*, basepair step), the contour length itself is hard to define, and the effects of other factors, like radius, are no longer negligible.

Nevertheless, these resultant profiles qualitatively preserve their shapes and relative orders among different runs, as well as under different recent force fields. So, we can still get some qualitative descriptions from these consistent trends. Compared against normal `C/G` pairing, `C/C` pairing is more flexible, due to the lost of hydrogen bonding, as well as some stacking. The bulky `G/G` pairing further disturbs the non-covalent interactions, and enhances the softening effect. In the case of `E/E`, the paired bases are gone, leaving the small sugar rings and adaptable phosphate backbones, and one may expect it is a very weak point. Actually, the most adjacent L, R basepairs are very near and orient with one another, which have a great change to adhere with each other. This was confirmed by further structural analysis, which explains the most rigid bending elasticity observed. But then, for the case of double `E/E`, the stacking distances have increased, and these long and flexible linkages hardly maintain the correct orientations between two adhesive basepairs. This usually leads to a permanent kinking, rather than stacking, which was shown by its 60° intrinsic kinking bending energy profile.

5.5 Discussion

In this chapter, $\Delta\mathcal{A}(d)$ and $f(d)$ of DNA with a pre-existing nick at 300 K, revealed interesting features of nicked DNA during bending. At low bending level when d is close to its contour length, the profiles of the nicked DNA overlap with those of nick-free DNA. This result indicates that under weak bending condition, nicked DNA maintains in B-form, and nick does not affect DNA elasticity, which is consistent with that reported in [119, 26]. However, at further decreased d when force reaches around 40 pN, which is still less than the predicted critical load from Euler instability, $\Delta\mathcal{A}(d)$ and $f(d)$ begin to deviate from nick-free DNA profiles. This marks the transition of nicked DNA from type A to type B, which is induced by unstacking of the two adjacent basepairs at the nicked site. Then,

at a much smaller d, melting from nicked site occurs, resulting in transition from type B to type C. As d further decreases, the level of peeling gradually increases, while $f(d)$ gradually decrease.

Furthermore, the unstacking at nicks is very sensitive to temperature change. $\Delta\mathcal{A}(d)$ and $f(d)$ of nicked DNA at 290 K significantly differ from those at 300 K. The observed unstacking at ~ 40 pN does not happen even beyond buckling transition, while nicked DNA bends like a nick-free DNA.

Overall, a pre-existing nick may facilitate excitation of flexible defects at the nicked location when the DNA is sharply bent. Moreover, the nick-dependent defect excitation is highly temperature sensitive. In DNA looping experiments, this result suggests that the two pre-existing nicks in the hybridized loop may cause DNA kink in sharply bent DNA minicircles, thereby violating the Ω boundary condition and boosting the probability of looping. This naturally explains the anomalously elasticity of < 110 bp observed in ligase-based DNA looping probability density experiments [3, 29], as well as in FRET-based looping (unlooping) rate measurements [29, 104]. It also clarifies the contradictory results for ~ 100 bp DNA looping, where $\rho^{\mathrm{E}}(\mathbf{0}) \gg \rho^{\mathrm{WLC}}(\mathbf{0})$ at 30°C, $\rho^{\mathrm{E}}(\mathbf{0}) = \rho^{\mathrm{WLC}}(\mathbf{0})$ at 21°C.

Importantly, the results obtained from nicked DNA draw cautions in using nick-containing DNA in experiments to infer micromechanics of sharply bent nick-free DNA. In addition, these results might have a physiological relevance, as DNA nicks are frequently produced *in vivo*, for examples, during DNA replication and DNA damages by UV or chemical attacks, which might facilitate local DNA bending for DNA damage recognition and repair [120, 121].

Chapter 6

Polymer model with defect excitations

6.1 Introduction

In order to check whether the simulated full atomic DNA behaviours can be described by modified WLC polymer model, such as generalized defect excitation model introduced in Chapter 1, we obtained the same free energy difference profiles and force-extension curves for nick-free and nicked DNA using Equation 1.24, and compared against those $\Delta\mathcal{A}(d)$ and $f(d)$ described in Chapters 4 and 5. Here, coarse-grained Monte Carlo (MC) methods were adapted to extrapolate the polymer models and approach analytically difficult or intractable properties, such as end-to-end distance distributions. In overall, we are going to briefly introduce the MC methods, and then, present the analogous MC results for normal and nick-containing DNA bending responses.

6.2 Monte Carlo simulations on DNA

MC simulations are computational strategies, which utilize sequences of random numbers to generate time-evolutions of models stochastically. On one hand, it breaks away from the high frequency physical laws, such as Newton's equations of motion used in classical MD simulations, and aggressively samples the phase space following some designed random movements. On the other hand,

it produces next sample by updating previous histories, while, obeying rigorous probability guidances. Thus, MC simulations achieve very high efficiencies, as well as, establish certain reliabilities, at least statistically. Usually, this stochastic process is further simplified and mathematically described by the first-order discrete-time Markov process, which is a collection of random variables in discrete and memoryless manner, as,

$$p\left(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_2 = x_2, X_1 = x_1\right) = p\left(X_i = x_i | X_{i-1} = x_{i-1}\right) \tag{6.1}$$

, where X_i is a random variable and takes a state x_i at discretized time point i, that only relates to its immediate preceding sample.

Now, we are going to illustrate the generations of such Markov chains by Metropolis MC simulations [122], using our thin elastic rod DNA polymer as an example system. Given any DNA discretized chain conformation, the next sampling point in configuration space is obtained by random number driven alternations of current configuration. There are various kinds of off-lattice displacements that particularly designed for such chain movements, among which reptation, crankshaft and pivot moves were frequently used in our MC simulations. In the reptation move, either the head or tail segment of the chain is removed, and replaced by a newly added segment with same length, which may evenly points to any direction in 3D [123, 124]. In the crankshaft move, two vertices are randomly selected, and their connecting line defines a rotating hinge. Then, the subchain in between is rotated about this hinge by an arbitrary angle in the range of $[-\phi_{\mathrm{c}}, \phi_{\mathrm{c}}]$ [125]. In the pivot move, a vertex is randomly picked up as rotating center, while, the shorter tail is rotated by a randomly generated angle in the range of $[-\phi_{\mathrm{p}}, \phi_{\mathrm{p}}]$ about an uniformly selected arbitrary axis [126].

After the new trial conformation is produced, its bending energy difference against previous conformation, $\Delta E = E_{\mathrm{new}} - E_{\mathrm{old}}$, is evaluated based on underlying polymer model Hamiltonians, for example, Equation 1.3, 1.24. Finally, this trial conformation is accepted with certain chances, $p = \min\left(1, \exp\left(-\beta \Delta E\right)\right)$, otherwise, discarded. So, starting from some initial chain conformation, recursively executions of above procedures lead to targeted Markov chains, which

randomly sample our DNA configuration space. Note that, the parameters, ϕ_{c} and ϕ_{p} in crankshaft and pivot moves, are adaptively adjusted during simulations to yield about half acceptance ratios, respectively.

It is worth the efforts to deeper understand the reasons behind above Metropolis MC simulations, which ensure the both "real" and "fast" sampling processes. The generated Markov chains are "real", because they are eventually equilibrated at Boltzmann distribution. Firstly, these Markov chains are reversible, due to $p(n)\,W\,(n \to m) = p(m)\,W\,(m \to n)$, where W is the transition rates, and, for the case of Metropolis algorithm, it is set to be previous acceptance probability divided by a scalable waiting time τ_0. This detailed balance is a sufficient condition for the following relation to hold,

$$\frac{\partial p(n)}{\partial t} = \sum_{n \neq m} p(m)\,W\,(m \to n) - \sum_{n \neq m} p(n)\,W\,(n \to m) \equiv 0. \tag{6.2}$$

The master equation above equaling zero indicates the existence of stationary distributions. Secondly, the chain movements are carefully designed to assure the ergodicity of such Markov chains, in other words, it is possible to get from every state to every other state with positive probabilities. Thus, this guarantees the uniqueness of stationary distributions. Taking together, Metropolis MC simulations are stochastically populating lots of DNA conformations, whose equilibrated subsets satisfy the Boltzmann distribution. Regarding about "fast", the embedded importance sampling prefers the evolutions towards lower free energy states to provoke more statistically impact samples (*i.e.*, with larger weighted factors in calculating ensemble averages). This further boosts the efficiency of our simulations.

Before proceeding to more complicated cases, we utilized MC simulations to reproduce the typical WLC polymer behaviours as justifications. Thorough comparisons have been conducted between simulated outcomes and analytic solutions in various aspects. During MC simulations, we performed aforementioned three kinds of movements repeatedly, till acceptance once for each. And this is considered as one MC cycle. For each simulation, the Markov chain initialized from a spiral configuration, and underwent a trial phase of $\sim 10^5$ cycles to determine

the adaptive parameters, ϕ_c and ϕ_p. Then, the polymer chain was brought to its stationary distribution through an equilibration phase of $\sim 10^6$ cycles. After that, a production phase of $\sim 10^7$ cycles were conducted, while, DNA conformations were recorded with a frequency of $\sim 10^{-2}$ for further analysis. In short, our simulated results perfectly agree with WLC analytic predictions from different perspectives, ranging from macroscopic end-to-end distances, to bending correlations, then, to microscopic bending angle distributions, as shown in Figure 6.1.

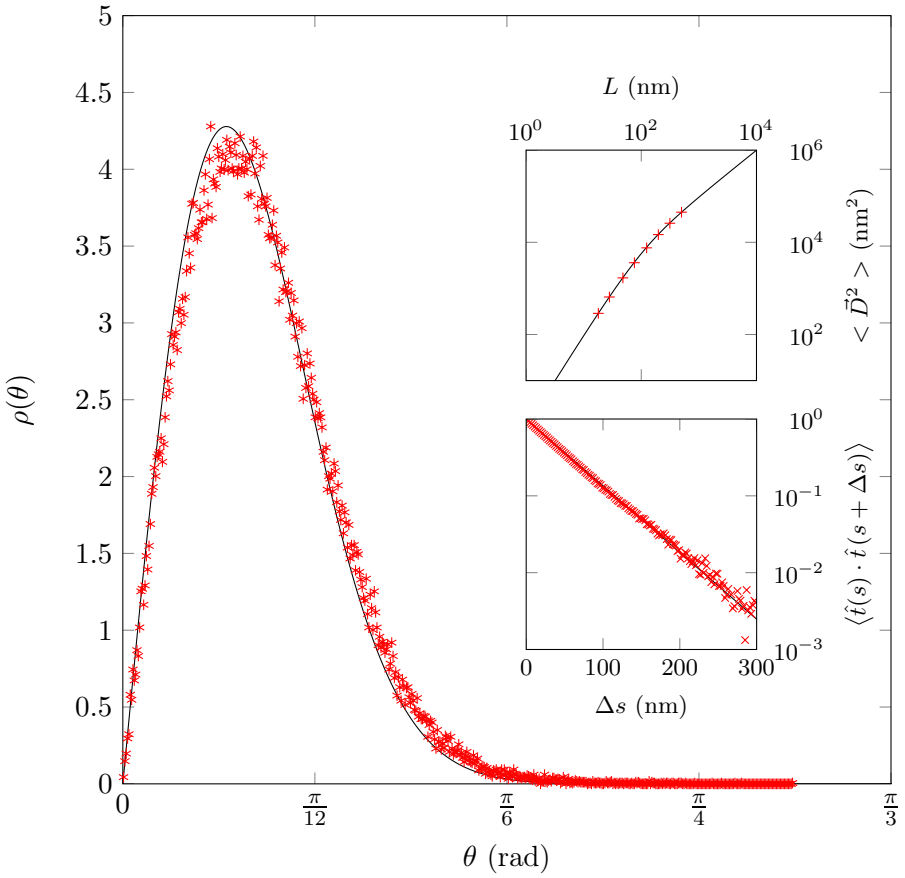


Figure 6.1: MC simulated WLC polymer behaviours compared against various WLC analytic predictions. The upper inset shows the simulated mean-square end-to-end distances for several chains with different contour lengths against Equation 1.8. The lower inset presents observed bending correlations *vs.* contour separations for the long chain with $L = 500$ nm against Equation 1.7. The main figure plots the equilibrated bending angle distributions over a discretized segment sampled evenly across the same chain, against $\rho(\theta) \propto \sin\theta \exp(a\cos\theta)$. Note that the red data points are MC simulated results, while black curves are analytic solutions for WLC model. And the persistence length $A = 50$ nm are used for both approaches.

6.3 Effects of defect excitations on the DNA mechanics

Based on our MD observations, DNA under sharp bending conditions develops local flexible defects, and breaks away from typical WLC behaviours. As introduced in Section 1.7, these excitations of mechanical defects can be naturally described by the generalized polymer model with nonlinear bending elastic responses. In this section, we intend to compare this flexible defect excitation model against computational expensive MD results, and hopefully to achieve more elegant quantifications, through obtaining similar $\Delta\mathcal{A}(d)$ and $f(d)$ using MC methods.

In General, each of MC simulations started from a random configuration. During each MC cycle, the conformation is updated by one crankshaft, one reptation and one pivot move each in random order. After initial 10^6 equilibration steps, data in the subsequent 10^9 steps were collected per hundred steps for analysis. For parameter tuning, we utilized Wang-Landau algorithm [127] to iteratively flatten the overall energy landscape with biased potential, and rapidly approach $\Delta\mathcal{A}(d)$ over entire bending range for preview. Once the tunable parameters have been finalized, more accurate $\Delta\mathcal{A}(d)$ is obtained through multiple production MC simulations using Metropolis algorithm and umbrella sampling, from which $f(d)$ is achieved through differentiation and smoothing.

For direct comparison, we need to model DNA using a particular discretized chain having the same stiffness and equilibrium length with our 20 bp full atomic DNA. To investigate the non-constraint behaviours, DNA was firstly modelled as a non-defectable, thickless chain of 7 beads connected by 6 bonds, which results in 5 vertices. At each vertex, the WLC harmonic potential with $A = 57$ nm was applied to penalize bending. In addition, its tunable bond length l was determined to be 0.9134 nm, which leads to targeted equilibrated end-to-end distance, $\langle d_{\{0\}} \rangle \approx 5.43$ nm. After its more accurate chain statistics were obtained using umbrella sampling.

As expected, this chain follows ideal WLC behaviours even at extreme bending, whose $\Delta\mathcal{A}(d)$ and $f(d)$ are plotted using orange lines in main and inset of

Figure 6.2(a) respectively. More characteristically, its force-distance curve have two force rising regimes: the steep linear regime upon bending and flattened regime after buckling transition with $f_c \approx 80$ pN. Although the slopes of these force regimes differ from those observed in MD simulation (due to the thickless assumption of polymer model), their general trends are identical at weak bending condition. On the other hand, under sharp bending, the additional force reclining regime after first order phase transition (*i.e.*, flexible defect excitation) is lacking based on non-defectable classic WLC model. In turn, a modified WLC model with nonlinear elasticity is required to capture the non-monotonic force responses.

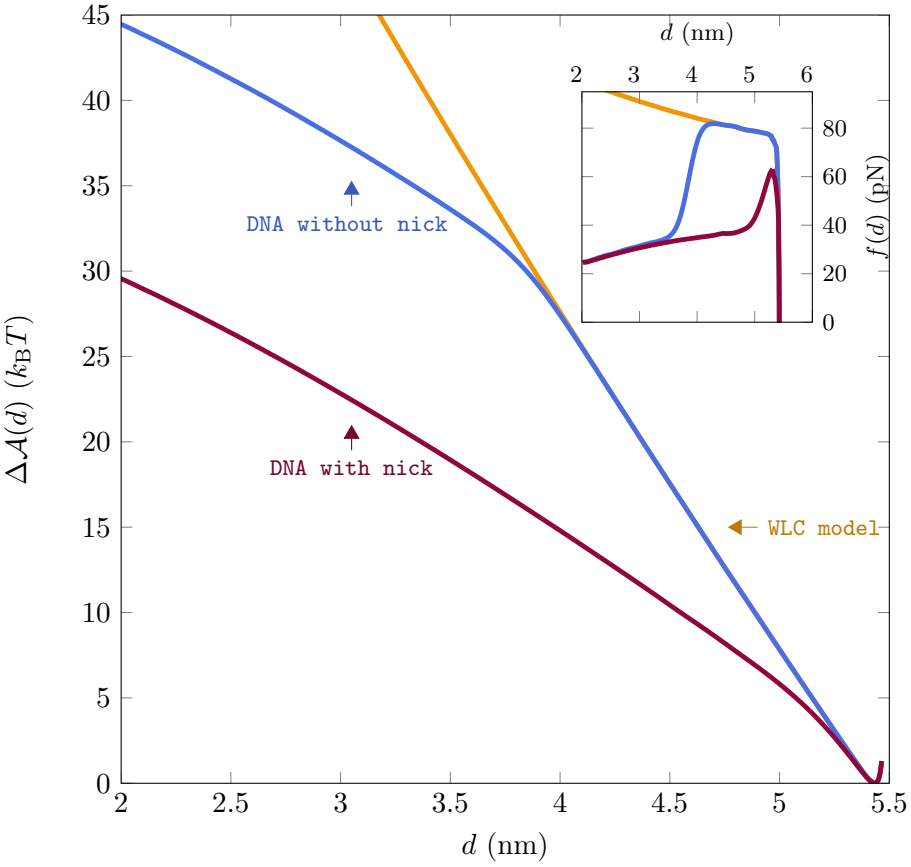
Thus, the original sum of harmonic potentials was replaced by following sum of non-harmonic potentials to incorporate the possibilities of defections in form of hinges,

$$\mathcal{H}_{\mathrm{MC}} = \beta^{-1} \sum_{i=1}^{5} \left(-\ln \left(e^{-\frac{a}{2}\left(\hat{t}_{i+1}-\hat{t}_i\right)^2} + e^{-\frac{a'_i}{2}\left(\hat{t}_{i+1}-\hat{t}_i\right)^2 - \beta\mu_i} \right) \right) \tag{6.3}$$

, where each effective vertex bending energy follows Equation 1.24 with fixed $a = \frac{A}{l} = \frac{57}{0.9134}$ and tunable $a'_i = \frac{A'_i}{l}$ along with μ_i for individual vertex i. Physically, $A'_i < A$ denotes the effective persistence length of hinged site, while $\mu_i > 0\ k_{\mathrm{B}}T$ denotes the energy cost for defect excitation. Intuitively, regarding their effects on $f(d)$, A'_i controls the height of reclining regime, while μ_i controls the transition coordinate, assuming the i^{th} vertex is defected under sharp bending constraint.

For nick-free DNA, based on the knowledge from previous MD simulation, the localized, cooperative, flexible, sequence-independent basepair disruptions can occur within any vertices, but only within one vertex (*i.e.* $1-3$ out of 20 bp). As a result, all vertices are modelled by same $A'_i = A'_B < A$ and $\mu_i = \mu_B > 0\ k_{\mathrm{B}}T$ for $i = 1, 2, \cdots, 5$. Under sharp bending, at least one vertex should be excited to form kink. Through preview $f(d)$, indeed, we produced the third reclining force regime after transition coordinate, upon introducing another available flexible vertex state, which only reveals in sharply bent B-DNA. Meanwhile, before transition coordinate, the $f(d)$ remains linear rising

(a) $\Delta\mathcal{A}(d)$ and $f(d)$ for DNA without and with nick



(b) Probability of defect occurrences

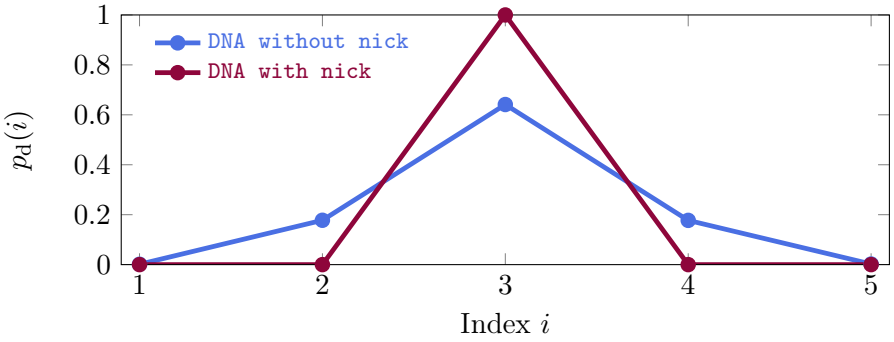


Figure 6.2: MC simulations on DNA without and with nick, using flexible defect excitation model. (a) $\Delta\mathcal{A}(d)$ and $f(d)$ for DNA without nick (dark blue) and DNA with nick (dark red), simulated using parameters ($A'_B = 15$ nm, $\mu_B = 18$ k_BT) for normal site and ($A'_N = 15$ nm, $\mu_N = 3$ k_BT) for nicked site. (b) The locational probabilities of defect occurrences along the 5 vertices in MC simulations, which shows that defect is always excited at the center of DNA. A vertex is determined as defected if its bending angle exceeds a threshold value $\theta_c = \cos^{-1}\left(1 - \frac{\mu_i}{a - a'_i}\right)$. Note that the orange coloured lines show the MC results based on WLC model with $A = 57$ nm as control.

and flattened force regimes, overlapping with that of WLC case. After extensive tuning, we found that $\mu_B = 18$ k_BT, $A'_B = 15$ nm were able to fit MD force-

distance relationship, yielding transition coordinate near 3.8 nm and $\sim$ 25 to 40 pN holding force with $d < 3.8$ nm (dark blue lines in Figure 6.2(a)).

For DNA with nick, the defection at nicked site requires much less excitation energy, due to the discontinuity in backbone. Here, nick's position is fixed after 11^{th} basepair, while the rest remains as B-DNA. To mimic the central localization of nick in MD, we assigned parameters for the third vertex with $\mu_3 = \mu_{\text{N}} < \mu_B$, $A'_3 = A'_{\text{N}} = A'_B$, while using the B-DNA parameters for the rest vertices. This leads to early defect at the nicked site, thus, transition into reclining force regime occurs at larger d. After extensive tuning, the dark red $f(d)$ were obtained with $\mu_{\text{N}} = 3\ k_{\text{B}}T$, which defects prior to buckling transition and generally fits MD force-distance curve of nicked DNA.

Figure 6.2(b) plots the probabilities of defect occurrence (*i.e.*, $p_{\text{D}}(i)$, for $i = 1, 2, \cdots, 5$) along the DNA for nick-free (dark blue) constraint near its transition coordinates $\sim$ 3.8 mm, which reveals the central localization of defects similar with the conclusion drew from MD simulations. In the case of nicked DNA, its intact-defect switching dynamics were obtained through constraint to $\sim$ 5.2 nm; and it $\sim$ 100% defects at the third vertex (dark red), indicating that nicks direct defect excitations.

The overall agreement on $\Delta\mathcal{A}(d)$ and $f(d)$ obtained between MD and MC simulations strongly suggest that the DNA bending elasticity from weak to sharp condition can be extracted using modified non-harmonic WLC polymer model through incorporating hinge excitation.

Chapter 7

Conclusion

In conclusion, in this thesis I describe my studies on the micromechanics of DNA under sharp bending conditions. Through it, I used full-atom MD simulations to induce DNA defects under bending constraints, umbrella sampling technique for accelerated configuration space sampling, statistical and mechanical analysis to quantify free energy difference profiles and force-extension curves for various types of DNA, as well as MC simulations to link observed defect excitations to generalized DNA polymer model. Several important results have been obtained.

One of the main findings from this research is that mechanical defects indeed can be excited when DNA is sharply bent, and these defects cause adsorptions of the DNA bend to defected sites; thereby relax the rest of DNA to maintain in their B-form. These defects typically involve disruptions of $1-3$ bp of DNA. Existence of such mechanically excited defects strongly suggests that the traditional homogeneous thin rod polymer model of DNA should breakdown under sufficiently sharp bending constraint. However, whether such bending condition was met in the currently reported experiments, where anomalous DNA elasticity was observed, remain unclear. As we have demonstrated in Chapter 5, in most of these experiments, there were at least one pre-existing nick, which likely might promote defect excitation at a less bending constraint.

To understand the mechanical impacts of aforementioned defect excitations on the DNA overall bending elastic properties, I performed umbrella sampling and WHAM to reconstruct the free energy difference profile as a function of the end-to-end distance of DNA, $\Delta\mathcal{A}(d)$, from which I also derived the force-

extension curve, $f(d)$, for DNA without and with excited defects. Prior to DNA defect excitation, $f(d)$ obtained for B-DNA is consistent with that expected from the classic WLC polymer model, which treats DNA as a thin elastic rod with homogeneous harmonic bending potential. Baesd on its Young's modulus estimated from my simulations, a bending persistence length of $\sim$ 57 nm was achieved, in good agreement with the value measured in experiments. I also observed the buckling transition at a force range of $70-85$ pN for the 20 bp DNA fragment, which is consistent with the prediction for thin rod by Euler instability based on our estimated A. These results gave me confidence on the validity of force field and the sampling methods to explore such large-scale properties of DNA at near equilibrium conditions.

In sharp contrast to the profiles of intact B-DNA, $\Delta\mathcal{A}(d)$ and $f(d)$ for defects excited DNA reveal totally different mechanical behaviours. $\Delta\mathcal{A}(d)$ becomes much flatter, which results in a significantly reduced $f(d)$. These results demonstrate that in the presence of defects, the overall required bending energy of DNA drastically decreases. In other words, the defects are mechanically flexible. Furthermore, these results also imply that it needs much less compressional force to maintain DNA in sharply bent conformation.

Although I could not access the equilibrium transitions between the B-DNA and defected DNA due to the limited timescales of MD simulation, the near equilibrium $\Delta\mathcal{A}(d)$ and $f(d)$ obtained for intact B-DNA and defect excited DNA already provide a clear physical picture of the overall $\Delta\mathcal{A}(d)$ and $f(d)$ for our 20 bp DNA during the whole bending process. $\Delta\mathcal{A}(d)$ is anticipated to be a monotonically increasing function as d decreases, which consists of three distinct regimes indicated by their slopes – a steeply rising region corresponding to B-DNA near its free equilibrium length, a flattened rising region corresponding to B-DNA after the buckling transition, and finally a reclining region after defect excitation. Correspondingly, $f(d)$ also exhibits distinct force responses in these regimes – a steeply rising region for B-DNA near its free equilibrium length, a flattened region for B-DNA after the buckling transition, and a significantly dropped flat region after defect excitation, resulting in a non-monotonic $f(d)$ profile.

The mechanistic insights obtained from the MD simulations are further tested in a generalized polymer model that permits excitation of flexible defects associated with an energy cost (*i.e.*, the flexible defect excitation model, Equation 1.24 with hinged E^1), using MC simulations. The $\Delta\mathcal{A}(d)$ and $f(d)$ obtained from the flexible defect excitation model successfully demonstrate the three regimes. Besides providing an understanding of the mechanical impact of flexible defects on the overall elasticity of DNA, these MC simulations also quantified the rigidity for the defects with effective persistence length ~ 15 nm.

Another important result is that pre-existing nicks on DNA have interesting mechanical effects on DNA overall elastic responses. Under weak bending condition, the nicked DNA fragment has identical $\Delta\mathcal{A}(d)$ and $f(d)$ profiles to B-DNA, indicating that nick-containing DNA can be treated with traditional WLC model when it is weakly bent. However, further bending of nick-containing DNA leads to defect excitation at the nicked site, which occurs at a level of bending much weaker than that needed to disrupt B-DNA basepairs. In fact, it is even weaker than the buckling transition of B-DNA at our contour length. It suggests that using WLC model quickly becomes invalid, due to defects at nicked site when DNA is sharply bent. Furthermore, this nick-dependent defect excitations, in form of unstacking or peeling, can be strongly suppressed by reducing temperature. In summary, our MD results reveal that pre-existing nicks in a sharply bent DNA are hotspots to adsorb the bending through developing localized kinks which relaxes the rest of nick-free DNA regions in a temperature dependent manner. This finding provides a natural explanation to the sometimes contradictory DNA elastic responses reported in nearly all previous DNA bending experiments, where the DNA necessarily contained nicks by experimental design.

Thus, these results put a doubt on previous interpretations of a series of recent experiments reporting anomalously flexible DNA fragments under sharp bending conditions. Instead of an intrinsic elastic response of *ds*DNA, the nick-dependent defect excitation is more plausible underlying mechanisms in those experiments, which cannot avoid nicks. Whether flexible defects can be excited in intact DNA at the level of bending in ~ 100 bp minicircles remains unknown and new rigorously designed experiments are needed to fully resolve these debates.

Bibliography

[1] J. F. Marko and E. D. Siggia, "Stretching DNA," *Macromolecules*, vol. 28, pp. 8759–8770, Dec. 1995.

[2] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 1460–1465, Feb. 1998.

[3] T. E. Cloutier and J. Widom, "Spontaneous sharp bending of double-stranded DNA," *Molecular Cell*, vol. 14, pp. 355–362, May 2004.

[4] J. Yan, R. Kawamura, and J. F. Marko, "Statistics of loop formation along double helix DNAs," *Physical Review E*, vol. 71, p. 061905, Jan. 2005.

[5] P. A. Wiggins, R. Phillips, and P. C. Nelson, "Exact theory of kinkable elastic polymers," *Physical Review E*, vol. 71, p. 021909, Jan. 2005.

[6] F. H. C. Crick and A. Klug, "Kinky helix," *Nature*, vol. 255, pp. 530–533, June 1975.

[7] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*, vol. 171, pp. 737–738, Apr. 1953.

[8] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich, "Molecular structure of a left-handed double helical DNA fragment at atomic resolution," *Nature*, vol. 282, pp. 680–686, Dec. 1979.

[9] A. Rich and S. Zhang, "Timeline: Z-DNA: the long road to biological function," *Nature Reviews Genetics*, vol. 4, pp. 566–572, July 2003.

[10] H. Fu, H. Chen, X. Zhang, Y. Qu, J. F. Marko, and J. Yan, "Transition dynamics and selection of the distinct S-DNA and strand unpeeling modes of double helix overstretching," *Nucleic acids research*, vol. 39, pp. 3473–3481, Jan. 2011.

[11] X. Zhang, H. Chen, H. Fu, P. S. Doyle, and J. Yan, "Two distinct overstretched DNA structures revealed by single-molecule thermodynamics measurements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 8103–8108, May 2012.

[12] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, "Entropic elasticity of λ-phage DNA," *Science*, vol. 265, pp. 1599–1600, Jan. 1994.

[13] S. B. Smith, L. Finzi, and C. Bustamante, "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads," *Science*, vol. 258, pp. 1122–1126, Jan. 1992.

[14] S. Cocco, J. Yan, J.-F. Léger, D. Chatenay, and J. F. Marko, "Overstretching and force-driven strand separation of double-helix DNA," *Physical Review E*, vol. 70, p. 011910, Jan. 2004.

[15] S. B. Smith, Y. Cui, and C. Bustamante, "Overstretching *B*-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules," *Science*, vol. 271, pp. 795–799, Jan. 1996.

[16] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, "DNA: an extensible molecule," *Science*, vol. 271, pp. 792–794, Feb. 1996.

[17] T. Odijk, "Stiff chains and filaments under tension," *Macromolecules*, vol. 28, pp. 7016–7018, Sept. 1995.

[18] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching DNA with optical tweezers," *Biophysical Journal*, vol. 72, pp. 1335–1346, Mar. 1997.

[19] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, "Effect of pH on the overstretching transition of double-stranded DNA: evidence of force-induced DNA melting," *Biophysical Journal*, vol. 80, pp. 874–881, Feb. 2001.

[20] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, "Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching," *Biophysical Journal*, vol. 80, pp. 1932–1939, Apr. 2001.

[21] C. Rivetti, M. Guthold, and C. Bustamante, "Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis," *Journal Of Molecular Biology*, vol. 264, pp. 919–932, Dec. 1996.

[22] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, "High flexibility of DNA on short length scales probed by atomic force microscopy," *Nature Nanotechnology*, vol. 1, pp. 137–141, Nov. 2006.

[23] J. Shimada and H. Yamakawa, "Ring-closure probabilities for twisted wormlike chains. Application to DNA," *Macromolecules*, vol. 17, pp. 689–698, July 1984.

[24] D. Shore, J. Langowski, and R. L. Baldwin, "DNA flexibility studied by covalent closure of short fragments into circles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, pp. 4833–4837, Aug. 1981.

[25] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, "Cyclization of short DNA fragments and bending fluctuations of the double helix," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 5397–5402, Jan. 2005.

[26] D. Shore and R. L. Baldwin, "Energetics of DNA twisting: I. relation between twist and cyclization probability," *Journal Of Molecular Biology*, vol. 170, pp. 957–981, Nov. 1983.

[27] W. H. Taylor and P. J. Hagerman, "Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure: II. NaCl-dependence of DNA flexibility and helical repeat," *Journal Of Molecular Biology*, vol. 212, pp. 363–376, Mar. 1990.

[28] R. A. Forties, R. A. Forties, R. Bundschuh, R. Bundschuh, M. G. Poirier, and M. G. Poirier, "The flexibility of locally melted DNA," *Nucleic acids research*, vol. 37, pp. 4580–4586, Aug. 2009.

[29] R. Vafabakhsh and T. Ha, "Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization," *Science*, vol. 337, pp. 1097–1101, Jan. 2012.

[30] J. Yan and J. F. Marko, "Localized single-stranded bubble mechanism for cyclization of short double helix DNA," *Physical Review Letters*, vol. 93, p. 108108, Jan. 2004.

[31] H. Shroff, B. M. Reinhard, M. Siu, H. Agarwal, A. Spakowitz, and J. Liphardt, "Biocompatible force sensor with optical readout and dimensions of 6 nm^3," *Nano Letters*, vol. 5, pp. 1509–1514, Aug. 2005.

[32] H. Shroff, D. Sivak, J. J. Siegel, A. L. McEvoy, M. Siu, A. Spakowitz, P. L. Geissler, and J. Liphardt, "Optical measurement of mechanical forces inside short DNA loops," *Biophysical Journal*, vol. 94, pp. 2179–2186, Mar. 2008.

[33] H. Q. Zocchi, C.-Y. Tseng, Y. Wang, A. J. Levine, and Giovanni, "The elastic energy of sharply bent nicked DNA," *Europhysics Letters*, vol. 90, p. 18003, Apr. 2010.

[34] H. Qu, Y. Wang, C.-Y. Tseng, and G. Zocchi, "Critical torque for kink formation in double-stranded DNA," *Physical Review X*, vol. 1, p. 021008, Jan. 2011.

[35] D. S. Sanchez, H. Qu, D. Bulla, and G. Zocchi, "DNA kinks and bubbles: temperature dependence of the elastic energy of sharply bent 10-nm-size DNA molecules," *Physical Review E*, vol. 87, p. 022710, Feb. 2013.

[36] Q. Du, A. Kotlyar, and A. Vologodskii, "Kinking the double helix by bending deformation," *Nucleic acids research*, vol. 36, pp. 1120–1128, Mar. 2008.

[37] A. Vologodskii, Q. Du, and M. D. Frank-Kamenetskii, "Bending of short DNA helices," *Artificial DNA: PNA & XNA*, vol. 4, pp. 1–3, Jan. 2013.

[38] J. Sponer, J. Leszczyński, and P. Hobza, "Nature of nucleic acid–base stacking: nonempirical *ab initio* and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and h-bonded base pairs," *The Journal of Physical Chemistry*, vol. 100, pp. 5590–5596, Jan. 1996.

[39] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, "The nature of the hydrogen bond in DNA base pairs: the role of charge transfer and resonance assistance," *Chemistry - A European Journal*, vol. 5, pp. 3581–3594, Dec. 1999.

[40] R. F. Goldstein and A. S. Benight, "How many numbers are required to specify sequence-dependent properties of polynucleotides?," *Biopolymers*, vol. 32, pp. 1679–1693, Dec. 1992.

[41] A. V. Vologodskii, B. R. Amirikyan, Y. L. Lyubchenko, and M. D. Frank-Kamenetskii, "Allowance for heterogeneous stacking in the DNA helix-coil transition theory," *Journal of Biomolecular Structure and Dynamics*, vol. 2, pp. 131–148, Aug. 1984.

[42] R. D. Blake, J. W. Bizzaro, J. D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx, and J. J SantaLucia, "Statistical mechanical simulation of polymeric DNA melting with `MELTSIM`," *Bioinformatics*, vol. 15, pp. 370–375, May 1999.

[43] D. Poland, "Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations," *Biopolymers*, vol. 13, pp. 1859–1871, Sept. 1974.

[44] T. Garel and H. Orland, "Generalized Poland-Scheraga model for DNA hybridization," *Biopolymers*, vol. 75, pp. 453–467, Dec. 2004.

[45] R. D. Blake, "`MELTSIM`: The DNA melting simulator since 1980." `http://bioinformatics.org/meltsim/wiki/`, 1999.

[46] M. Fixman and J. J. Freire, "Theory of DNA melting curves," *Biopolymers*, vol. 16, pp. 2693–2704, Dec. 1977.

[47] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 15431–15436, Aug. 2010.

[48] P. Gross, N. Laurens, L. B. Oddershede, U. Bockelmann, E. J. G. Peterman, and G. J. L. Wuite, "Quantifying how DNA stretches, melts and changes twist under tension," *Nature Physics*, vol. 7, pp. 731–736, May 2011.

[49] B. Bohrmann, M. Haider, and E. Kellenberger, "Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in STEM," *Ultramicroscopy*, vol. 49, pp. 235–251, Feb. 1993.

[50] R. J. Ellis, "Macromolecular crowding: obvious but underappreciated," *Trends in Biochemical Sciences*, vol. 26, pp. 597–604, Oct. 2001.

[51] V. Vogel and M. Sheetz, "Local force and geometry sensing regulate cell functions," *Nature Reviews Molecular Cell Biology*, vol. 7, pp. 265–275, Feb. 2006.

[52] G. V. Shivashankar, "Mechanosignaling to the cell nucleus and gene regulation," *Annual Review of Biophysics*, vol. 40, pp. 361–378, June 2011.

[53] L. Bai, T. J. Santangelo, and M. D. Wang, "Single-molecule analysis of RNA polymerase transcription," *Annual Review of Biophysics and Biomolecular Structure*, vol. 35, pp. 343–360, June 2006.

[54] A. J. Courey and J. C. Wang, "Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions," *Cell*, vol. 33, pp. 817–829, July 1983.

[55] J. J. Champoux, "DNA topoisomerases: structure, function, and mechanism," *Annual Review Of Biochemistry*, vol. 70, pp. 369–413, June 2001.

[56] R. Kavenoff and B. Bowen, "Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*," *Chromosoma*, vol. 59, pp. 89–101, Dec. 1976.

[57] X. Wang, P. M. Llopis, and D. Z. Rudner, "Organization and segregation of bacterial chromosomes," *Nature Reviews Genetics*, vol. 14, pp. 191–203, Feb. 2013.

[58] N. V. Hud and K. H. Downing, "Cryoelectron microscopy of λ-phage DNA condensates in vitreous ice: The fine structure of DNA toroids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, pp. 14925–14930, Dec. 2001.

[59] M. Emanuel, G. Lanzani, and H. Schiessel, "Multiplectoneme phase of double-stranded DNA under torsion," *Physical Review E*, vol. 88, p. 022706, Aug. 2013.

[60] J. R. C. van der Maarel, S. S. Zakharova, W. Jesse, C. Backendorf, S. U. Egelhaaf, and A. Lapp, "Supercoiled DNA; plectonemic structure and liquid crystal formation," *Journal of Physics: Condensed Matter*, vol. 15, pp. S183–S189, Jan. 2003.

[61] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, pp. 251–260, Sept. 1997.

[62] L. R. Comolli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante, and K. H. Downing, "Three-dimensional architecture of the bacteriophage ϕ29 packaged genome and elucidation of its packaging process," *Virology*, vol. 371, pp. 267–277, Feb. 2008.

[63] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer," *Science*, vol. 271, pp. 1247–1254, Mar. 1996.

[64] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution," *Journal Of Molecular Biology*, vol. 319, pp. 1097–1113, June 2002.

[65] T. J. Richmond and C. A. Davey, "The structure of DNA in the nucleosome core," *Nature*, vol. 423, pp. 145–150, May 2003.

[66] P. A. Rice, S.-w. Yang, K. Mizuuchi, and H. A. Nash, "Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn," *Cell*, vol. 87, pp. 1295–1306, Dec. 1996.

[67] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, "The bacteriophage ϕ29 portal motor can package DNA against a large internal force," *Nature*, vol. 413, pp. 748–752, Oct. 2001.

[68] J. Tang, N. Olson, P. J. Jardine, S. Grimes, D. L. Anderson, and T. S. Baker, "DNA poised for release in bacteriophage ϕ29," *Structure*, vol. 16, pp. 935–943, June 2008.

[69] S. Oehler, E. R. Eismann, H. Krämer, and B. Müller-Hill, "The three operators of the *lac* operon cooperate in repression," *The EMBO Journal*, vol. 9, pp. 973–979, Apr. 1990.

[70] L. Finzi and J. Gelles, "Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules," *Science*, vol. 267, pp. 378–380, Jan. 1995.

[71] P. Wiggins and P. Nelson, "Generalized theory of semiflexible polymers," *Physical Review E*, vol. 73, p. 031906, Mar. 2006.

[72] L. Verlet, "Computer "experiments" on classical fluids. I. Thermodynamical properties of lennard-jones molecules," *Physical Review*, vol. 159, pp. 98–103, July 1967.

[73] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters," *The Journal of Chemical Physics*, vol. 76, pp. 637–649, Jan. 1982.

[74] P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. New York, N.Y.: Clarendon Press, 1987.

[75] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, pp. 253–287, Jan. 1921.

[76] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, pp. 10089–10092, Jan. 1993.

[77] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, pp. 5179–5197, May 1995.

[78] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A new force field for molecular mechanical simulation of nucleic acids and proteins," *Journal of the American Chemical Society*, vol. 106, pp. 765–784, Feb. 1984.

[79] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, "An all atom force field for simulations of proteins and nucleic acids," *Journal of computational chemistry*, vol. 7, pp. 230–252, Apr. 1986.

[80] T. E. Cheatham, P. Cieplak, and P. A. Kollman, "A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat," *Journal of Biomolecular Structure and Dynamics*, vol. 16, pp. 845–862, May 1999.

[81] 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, Jan. 2000.

[82] P. Várnai and K. Zakrzewska, "DNA and its counterions: a molecular dynamics study," *Nucleic acids research*, vol. 32, pp. 4269–4280, Jan. 2004.

[83] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco, "Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers," *Biophysical Journal*, vol. 92, pp. 3817–3829, June 2007.

[84] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, "A standard reference frame for the description of nucleic acid base-pair geometry," *Journal Of Molecular Biology*, vol. 313, pp. 229–237, Oct. 2001.

[85] S. Diekmann, "Definitions and nomenclature of nucleic acid structure parameters," *Journal Of Molecular Biology*, vol. 205, pp. 787–791, Feb. 1989.

[86] M. A. El Hassan and C. R. Calladine, "The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme," *Journal Of Molecular Biology*, vol. 251, pp. 648–664, Aug. 1995.

[87] M. S. Babcock, E. P. D. Pednault, and W. K. Olson, "Nucleic acid structure analysis: mathematics for local cartesian and helical structure parameters that are truly comparable between structures," *Journal Of Molecular Biology*, vol. 237, pp. 125–156, Mar. 1994.

[88] S. Arnott and D. W. L. Hukins, "Refinement of the structure of *B*-DNA and implications for the analysis of X-ray diffraction data from fibers of biopolymers," *Journal Of Molecular Biology*, vol. 81, pp. 93–105, Dec. 1973.

[89] R. Chandrasekaran and S. Arnott, "2.4.4.1 Molecular and crystal structures," in *Crystallographic and Structural Data II*, pp. 36–37, Berlin/Heidelberg: Springer-Verlag, 1989.

[90] L. Clowney, S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman, "Geometric parameters in nucleic acids: nitrogenous bases," *Journal of the American Chemical Society*, vol. 118, pp. 509–518, Jan. 1996.

[91] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, pp. 629–642, Apr. 1987.

[92] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling," *Journal of computational physics*, vol. 23, pp. 187–199, Feb. 1977.

[93] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *Journal of computational chemistry*, vol. 13, pp. 1011–1021, Jan. 1992.

[94] M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations," *Computer Physics Communications*, vol. 135, pp. 40–57, Mar. 2001.

[95] 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. 1701–1718, Jan. 2005.

[96] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "`GROMACS` 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, pp. 845–854, Jan. 2013.

[97] X.-J. Lu and W. K. Olson, "`3DNA`: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures," *Nature protocols*, vol. 3, pp. 1213–1227, July 2008.

[98] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, pp. 926–935, Jan. 1983.

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

[100] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: a new molecular dynamics method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, Dec. 1980.

[101] R. Lavery and H. Sklenar, "The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids," *Journal of Biomolecular Structure and Dynamics*, vol. 6, pp. 63–91, Nov. 1988.

[102] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, "Conformational analysis of nucleic acids revisited: `Curves+`," *Nucleic acids research*, vol. 37, pp. 5917–5929, Jan. 2009.

[103] H. Yamakawa, "Statistical mechanics of wormlike chains. II. Excluded volume effects," *The Journal of Chemical Physics*, vol. 57, pp. 2843–2854, Oct. 1972.

[104] T. T. Le and H. D. Kim, "Probing the elastic limit of DNA bending," *Nucleic acids research*, vol. 42, pp. 10786–10794, Dec. 2014.

[105] G. H. Michler, *Micromechanical Properties*, vol. 1 of *Properties and Behavior of Polymers*. Hoboken, N.J.: Wiley, Apr. 2011.

[106] A. Grossfield, "An implementation of weighted histogram analysis method." `http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html`, 2013.

[107] J. S. Hub, B. L. de Groot, and D. van der Spoel, "`g_wham`—a free weighted histogram analysis implementation including robust error and autocorrelation estimates," *Journal of Chemical Theory and Computation*, vol. 6, pp. 3713–3720, Jan. 2010.

[108] E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, "Stacked–unstacked equilibrium at the nick site of DNA," *Journal Of Molecular Biology*, vol. 342, pp. 775–785, Sept. 2004.

[109] H. Chen and J. Yan, "Effects of kink and flexible hinge defects on mechanical responses of short double-stranded DNA molecules," *Physical Review E*, vol. 77, p. 041907, Apr. 2008.

[110] T. E. Cloutier and J. Widom, "DNA twisting flexibility and the formation of sharply looped protein–DNA complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 3645–3650, Mar. 2005.

[111] D. V. Pyshnyi and E. M. Ivanova, "Thermodynamic parameters of coaxial stacking on stacking hybridization of oligodeoxyribonucleotides," *Russian Chemical Bulletin*, vol. 51, no. 7, pp. 1145–1155, 2002.

[112] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix," *Nucleic Acids Research*, vol. 34, no. 2, pp. 564–574, 2006.

[113] T. A. Kunkel, "DNA replication fidelity," *Journal of Biological Chemistry*, vol. 279, pp. 16895–16898, Apr. 2004.

[114] R. R. Iyer, A. Pluciennik, V. Burdett, and P. L. Modrich, "DNA mismatch repair: functions and mechanisms," *Chemical Reviews*, vol. 106, pp. 302–323, Feb. 2006.

[115] B. O. Parker and M. G. Marinus, "Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 1730–1734, Mar. 1992.

[116] J. Lhomme, J.-F. Constant, and M. Demeunynck, "Abasic DNA structure, reactivity, and recognition," *Biopolymers*, vol. 52, pp. 65–83, Jan. 1999.

[117] J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner, and J. Stubbe, "Nuclear magnetic resonance structural studies and molecular modeling of duplex DNA containing normal and 4′-oxidized abasic sites," *Biochemistry*, vol. 46, pp. 3096–3107, Feb. 2007.

[118] J. Chen, F.-Y. Dupradeau, D. A. Case, C. J. Turner, and J. Stubbe, "DNA oligonucleotides with `A`, `T`, `G` or `C` opposite an abasic site: structure and dynamics," *Nucleic acids research*, vol. 36, pp. 253–262, Jan. 2008.

[119] J. B. Hays and B. H. Zimm, "Flexibility and stiffness in nicked DNA," *Journal Of Molecular Biology*, vol. 48, pp. 297–317, Mar. 1970.

[120] D. Branzei and M. Foiani, "Regulation of DNA repair throughout the cell cycle," *Nature Reviews Molecular Cell Biology*, vol. 9, pp. 297–308, Feb. 2008.

[121] E. C. Friedberg, "DNA damage and repair," *Nature*, vol. 421, pp. 436–440, Jan. 2003.

[122] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092, June 1953.

[123] S. K. Kumar, M. Vacatello, and D. Y. Yoon, "Off-lattice Monte Carlo simulations of polymer melts confined between two plates," *The Journal of Chemical Physics*, vol. 89, pp. 5206–5215, Oct. 1988.

[124] L. Dai, J. J. Jones, J. R. C. van der Maarel, and P. S. Doyle, "A systematic study of DNA conformation in slitlike confinement," *Soft Matter*, vol. 8, pp. 2972–2982, Feb. 2012.

[125] K. V. Klenin, A. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii, "Computer simulation of DNA supercoiling," *Journal Of Molecular Biology*, vol. 217, pp. 413–419, Feb. 1991.

[126] N. Madras and A. Sokal, "The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk," *Journal of Statistical Physics*, vol. 50, pp. 109–186, Jan. 1988.

[127] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 86, no. 10, pp. 2050–2053, 2001.

[128] A. D. McLachlan, "Gene duplications in the structural evolution of chymotrypsin," *Journal Of Molecular Biology*, vol. 128, pp. 49–79, Feb. 1979.

[129] W. R. Hamilton, "On quaternions; or on a new system of imaginaries in algebra," *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, vol. xxv, pp. 10–13, July 1844.

Appendix A

DNA local correlations among varies models

The long DNA molecules under weak constrains form ball-shape random coil, and well described by Gaussian chain model, where their end-to-end distances, as well as any subsections ($L \gg A$), approximate a characteristic distribution,

$$\rho\Big(\vec{D}\Big) = \left(\frac{2}{3}\pi\left\langle\vec{D}^2\right\rangle\right)^{-\frac{3}{2}} \exp\left(-\frac{3\vec{D}^2}{2\left\langle\vec{D}^2\right\rangle}\right). \tag{A.1}$$

All the common models, such as WLC, freely rotating chain (FRC), FJC models, follow this behaviour as large-scale and weak constrains. They can actually be treated as coarse-grained models with different levels of simplifications, where FJC, FRC, WLC rank from simple to complex. They can be unified through the second moments of end-to-end distance distributions, which is also the summations of total local correlations over contour length,

$$\left\langle\vec{D}^2\right\rangle = \int_{s=0}^{L}\int_{s'=0}^{L}\left\langle\hat{t}(s)\cdot\hat{t}\big(s'\big)\right\rangle ds'\,ds \tag{A.2}$$

In long homogeneous DNA, the correlations decay very fast, and boundary effect is negligible. As a result, the mean square of end-to-end distance relates to the signature constant of total correlations as,

$$\left\langle\vec{D}^2\right\rangle = \int_{s=0}^{L}\int_{\Delta s=-\infty}^{\infty}\left\langle\hat{t}(s)\cdot\hat{t}\,(s+\Delta s)\right\rangle d\Delta s\,ds = LC_{\text{corr}}. \tag{A.3}$$

This signature constant among different models are the same, which is $C_{\text{corr}} \approx$ 100 nm for DNA. We can prove this by integrating the correlations *vs.* s curves

plotted in Figure A.1, using different models,

$$C_{\text{corr}} = b = \frac{1+\cos\left(\theta_{\text{eff}}\right)}{1-\cos\left(\theta_{\text{eff}}\right)} l = \frac{1+\mathcal{L}(a)}{1-\mathcal{L}(a)} l = \int_{\Delta s=-\infty}^{\infty} \exp\left(-\frac{\Delta s}{A}\right) d\Delta s = 2A \tag{A.4}$$

, where b is the Kuhn length from FJC model, in which the correlations beyond A is suddenly lost. θ_{eff} is the effective bending angle over segment l in FRC model, in which only the projections of correlations propagate. $\mathcal{L}(a) = \langle \cos\theta \rangle$ is the segmental correlations, which is a function of discretized bending elastic constant $a = \frac{A}{l}$ in discretized WLC model, and $\mathcal{L}$ the Langevin function (its stepwise correlation function is exactly the same as FRC model in Figure A.1). And, finally, $\exp\left(-\frac{\Delta s}{A}\right)$ is the explicit bending correlations in continuous WLC model.

So, at large length scale, the long DNA molecules under weak constrains described by varies models, end up with same second moments of end-to-end distance distribution. The chain movements are purely entropic in all models. Its entropy is,

$$S = k_{\text{B}} \ln \Omega = k_{\text{B}} \ln \rho + S_0 \tag{A.5}$$

, and its Helmholtz free energy is,

$$\mathcal{A}\left(\vec{D}\right) = U - TS\left(\vec{D}\right) = \frac{3k_{\text{B}}T\vec{D}^2}{2\left\langle \vec{D}^2 \right\rangle} + \mathcal{A}_0. \tag{A.6}$$

By taking the derivative, we obtained the force following Hooke's law, with a temperature dependent elastic constant, same as the result obtained in Equation 1.11, at low force regime during stretching,

$$f = \frac{\partial \mathcal{A}\left(\vec{D}\right)}{\partial \vec{D}} = \frac{3k_{\text{B}}T}{2AL}\vec{D}. \tag{A.7}$$

On the other hand, as the force increases, the discrepancies among different models start to reveal, where WLC is more realistic in describing DNA. The polymer thermal fluctuations can be considered as superimpositions of waves with different wavelengths. The wavelength shorter than Kuhn length, b, is absent in FJC models, while is present in WLC models. Additional force is

needed to overcome the contributions from high frequency fluctuations in WLC model. This is the cause of shorter end-to-end distances of WLC, compared against FJC predictions under large extension force in Figure 1.3.

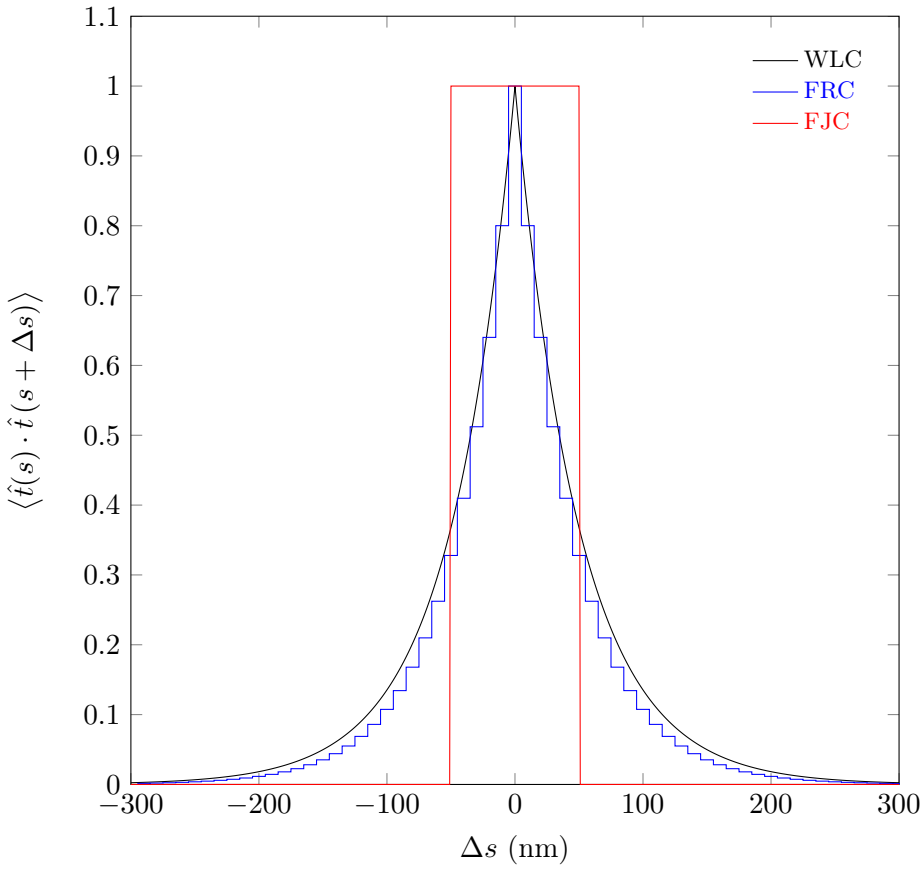


Figure A.1: The local correlation function, $\langle \hat{t}(s) \cdot \hat{t}\,(s+\Delta s) \rangle$, at any position s against another position Δs away for varies models with the same mean square of end-to-end distance, $\left\langle \vec{D}^2 \right\rangle$. The red line is that for FJC model with Kuhn length $b = 100$ nm, the blue line is that for FRC model with effective bond length $l = 10$ mm; effective bending angle $\theta_{\text{eff}} = 36.86°$, and the black line is that for WLC model with persistence length $A = 50$ nm. Further more, for discretized WLC model with segment length $l = 10$ nm, its local correlation function is the same as the curve for FRC model. For all these models, the area below corresponding curves are the same, $C_{\text{corr}} = 2A$. This is due to the relation of $\left\langle \vec{D}^2 \right\rangle = C_{\text{corr}} L$ for long polymer chains with contour length of L. Note that the local correlation functions are plotted for s locate right at the center of segments in FJC, FRC, discretized WLC models; for other s the curves are exactly the same in shapes, but with a shift along x-axis.

Appendix B

DNA orientation parameters illustration

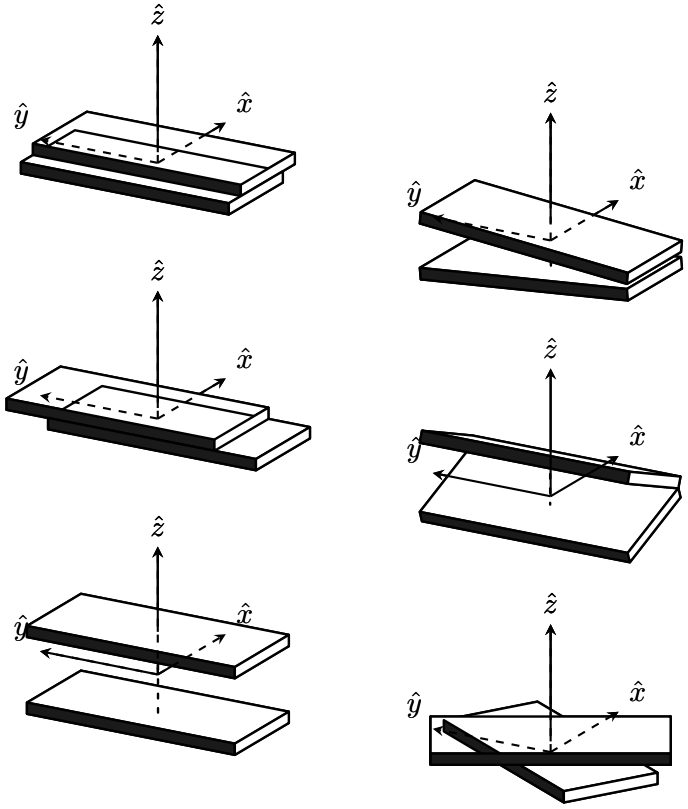


Figure B.1: Sequential basepair parameter illustrations. The first column lists three translational orientations for adjacent basepairs, **shift** (D_x), **slide** (D_y), and **rise** (D_z) along x, y and z-axis from top to bottom. While, the second column lists three rotational orientations, which are **tilt** (τ), **roll** (ρ), and **twist** (Ω) along x, y and z-axis respectively. Note that the two basepairs move in reflection symmetry about the triads, which is indeed the middle frames.

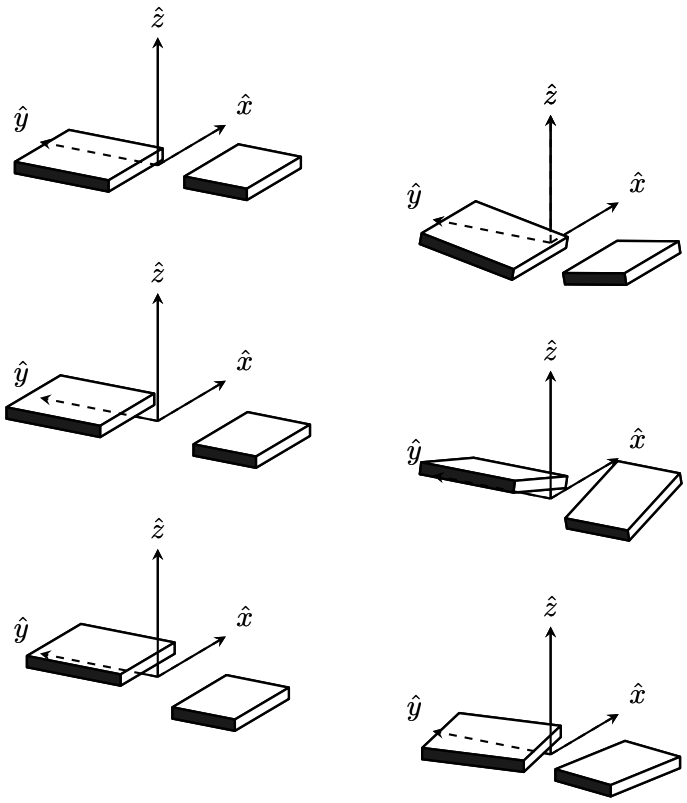


Figure B.2: Complementary base parameter illustrations. The first column lists three translational orientations for complementary bases, **shear** (S_x), **stretch** (S_y), and **stagger** (S_z) along x, y and z-axis from top to bottom. While, the second column lists three rotational orientations, which are **buckle** (κ), **propeller twist** (ω), and **opening** (σ) along x, y and z-axis respectively. Note that the movements for complementary bases are mirrored through triads as well.

Appendix C

Closed-form solution for absolute orientation

As mentioned in 2.5.1, in order to assign a base reference frame, we need to fit the standard base to an instant base atomic arrangement. This classical problem of finding absolute spatial orientation between two sets of points is usually approached by minimizing the sum of their residual squares, and is commonly refer as least-square fitting. The least-square fitting problem has a wide range of applications in variety of field, which is also frequently encountered in other forms through this thesis, *e.g.* structure superpositions. It has been implemented in different ways, such as a conventional solution given by McLachlan using Euler angles [128], and a closed-form solution provided by Horn using quaternion [91]. Here, we focus on the second method, which has been adopted by `X3DNA` and used in our base reference frame attachment processes.

C.1 Least-square fitting

Given two sets of points $\{\mathbf{r}_{\mathrm{A},i}\}$ and $\{\mathbf{r}_{\mathrm{B},i}\}$, $i = 1, 2, \cdots, N$, the absolute relative orientation can be characterized using a rotational operation $\mathcal{R}$, a scaling factor s and a translational vector $\mathbf{r}_0$. It brings the first set of points to the vicinity of the second by,

$$\mathbf{r}_{\mathrm{B},i} = s\mathcal{R}(\mathbf{r}_{\mathrm{A},i}) + \mathbf{r}_0. \tag{C.1}$$

While the "best" fitting is achieved by minimizing the residual sum of squares (RSS),

$$\mathrm{RSS} = \sum_{i=1}^{N} \|e_i\|^2 \tag{C.2}$$

, where the residual is $e_i = \mathbf{r}_{\mathrm{B},i} - s\mathcal{R}(\mathbf{r}_{\mathrm{A},i}) - \mathbf{r}_0$.

The problem is simplified by resetting their local coordinates to their corresponding centroids using,

$$\mathbf{r}'_i = \mathbf{r}_i - < \mathbf{r}_i > . \tag{C.3}$$

And RSS is expanded into following,

$$\begin{aligned} \mathrm{RSS} &= \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{B},i} - s\mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big) - \mathbf{r}'_0\|^2 \\ &= \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{B},i} - s\mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big)\|^2 - 2\mathbf{r}'_0 \cdot \sum_{i=1}^{N} \big[\mathbf{r}'_{\mathrm{B},i} - s\mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big)\big] + N\|\mathbf{r}'_0\|^2 \end{aligned} \tag{C.4}$$

, which has a zero second term due to the centroid locations, and is obviously minimized with $\mathbf{r}'_0 = 0$. Then, we further expand the remaining first term to find a proper scaling factor,

$$\begin{aligned} \mathrm{RSS} &= \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{B},i} - s\mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big)\|^2 \\ &= \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{B},i}\|^2 - 2s\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big) + s^2 \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{A},i}\|^2 \end{aligned} \tag{C.5}$$

, where $\|\mathcal{R}\Big(\mathbf{r}'_{\mathrm{A},i}\Big)\|^2 = \|\mathbf{r}'_{\mathrm{A},i}\|^2$, because translational operation preserves the vector length. Equation C.5 is expressed as,

$$\mathrm{RSS} = \left(s\sqrt{\sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{A},i}\|^2} - \frac{\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\Big(\mathbf{r}'_{\mathrm{A},i}\Big)}{\sqrt{\sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{A},i}\|^2}} \right)^2 + \sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{B},i}\|^2 - \frac{\Big(\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\Big(\mathbf{r}'_{\mathrm{A},i}\Big)\Big)^2}{\sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{A},i}\|^2} \tag{C.6}$$

, who is minimized, when first term is zero with the scaling factor,

$$s = \frac{\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\Big(\mathbf{r}'_{\mathrm{A},i}\Big)}{\sum_{i=1}^{N} \|\mathbf{r}'_{\mathrm{A},i}\|^2} \tag{C.7}$$

, and when cross term $\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\Big(\mathbf{r}'_{\mathrm{A},i}\Big)$ is as large as possible.

As a result, we have found the solutions for translational vector and scaling factor, which minimize RSS. While, the remaining errors is only relate to the rotational operation, and RSS is minimized when the following term is maximized,

$$\sum_{i=1}^{N} \mathbf{r}'_{\mathrm{B},i} \cdot \mathcal{R}\big(\mathbf{r}'_{\mathrm{A},i}\big) \tag{C.8}$$

C.2 Solving rotation with quaternion

The quaternion is a numbering system (*i.e.*, hypercomplex number), in form of $\mathring{q} = q_0 + \mathrm{i}q_x + \mathrm{j}q_y + \mathrm{k}q_z$, which was proposed by Hamilton in 1843 [129]. It can be used to describe the rigid body orientation in 3D Euclidean space through representing its rotation matrix. Compared with conventional Euler angles, it avoids the singularity issue, when pitch equals $\pm\frac{\pi}{2}$ (*i.e.*, when the first axis coincides with the second, also well-known as gimbal lock). Moreover, it is computational efficient, due to minimal parameterizations, free of trigonometric functions and less multiplications. Here, it is applied to simplify the deviation of the closed-form solution for optimized rotation.

We start the deviation from introducing some basic properties of quaternion. The imaginary units can be treat as right-handed triad, and their products are,

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1 \tag{C.9}$$

, and,

$$\begin{aligned} \mathrm{ij} &= \ \ \mathrm{k}, \quad \mathrm{jk} = \ \ \mathrm{i}, \quad \mathrm{ki} = \ \ \mathrm{j}; \\ \mathrm{ji} &= -\mathrm{k}, \quad \mathrm{kj} = -\mathrm{i}, \quad \mathrm{ik} = -\mathrm{j}. \end{aligned} \tag{C.10}$$

As a result, the multiplication rule of quaternion is noncommunicative, $\mathring{q}\mathring{r} \neq \mathring{r}\mathring{q}$, which can be expressed by the product of an orthogonal 4×4 matrix and a vector, as follow,

$$\mathring{q}\mathring{r} = \begin{bmatrix} q_0 & -q_x & -q_y & -q_z \\ q_x & q_0 & -q_z & q_y \\ q_y & q_z & q_0 & -q_x \\ q_z & -q_y & q_x & q_0 \end{bmatrix} \mathring{r} = \mathbb{Q}\mathring{r} \tag{C.11}$$

, while, shifting their order yields,

$$\mathring{r}\mathring{q} = \begin{bmatrix} q_0 & -q_x & -q_y & -q_z \\ q_x & q_0 & q_z & -q_y \\ q_y & -q_z & q_0 & q_x \\ q_z & q_y & -q_x & q_0 \end{bmatrix} \mathring{r} = \bar{\mathbb{Q}}\mathring{r} \tag{C.12}$$

, where $\mathring{r} = r_0 + \mathrm{i}r_x + \mathrm{j}r_y + \mathrm{k}r_z$.

The dot product of quaternion is defined as,

$$\mathring{q} \cdot \mathring{r} = q_0 r_0 + q_x r_x + q_y r_y + q_z r_z. \tag{C.13}$$

Then, the dot product of quaternion with itself is the square magnitude of quaternion,

$$\mathring{q} \cdot \mathring{q} = \|\mathring{q}\|^2 = q_0^2 + q_x^2 + q_y^2 + q_z^2 \tag{C.14}$$

, and if it equals 1, $\mathring{q}$ is called a unit quaternion. Based on Equation C.11, the product of quaternion with its conjugate, $\mathring{q}^* = q_0 - \mathrm{i}q_x - \mathrm{j}q_y - \mathrm{k}q_z$, equates to this quantity as well,

$$\mathring{q}\mathring{q}^* = \bar{\mathbb{Q}}^T \mathring{q} = q_0^2 + q_x^2 + q_y^2 + q_z^2 = \mathbb{Q}^T \mathring{q} = \mathring{q}^* \mathring{q}. \tag{C.15}$$

With above basic equations, we can define the rotation using simplest unit quaternion operations as,

$$\mathring{r}' = \mathring{q}\mathring{r}\mathring{q}^* \tag{C.16}$$

, where both dot products and cross products are preserved under this operations, refer to Hore *et al.* [91] for more details. Then, it can be written into a more familiar form as, $\mathring{r}' = \bar{\mathbb{Q}}^T \mathbb{Q} \mathring{r}$, which is the targeting rotational operation $\mathbf{r}_i' = \mathcal{R}(\mathbf{r}_i)$ by selecting 3×3 matrix at the bottom right of $\bar{\mathbb{Q}}^T\mathbb{Q}$ as $\mathcal{R}$, while eliminating real part of $\mathring{r}$. The explicit form of $\bar{\mathbb{Q}}^T\mathbb{Q}$ expressed using unit quaternion $\mathring{q}$ is,

$$\bar{\mathbb{Q}}^T\mathbb{Q} = \begin{bmatrix} \mathring{q} \cdot \mathring{q} & 0 & 0 & 0 \\ 0 & q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2\left(q_x q_y - q_0 q_z\right) & 2\left(q_x q_z + q_0 q_y\right) \\ 0 & 2\left(q_x q_y + q_0 q_z\right) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2\left(q_y q_z - q_0 q_x\right) \\ 0 & 2\left(q_x q_z - q_0 q_y\right) & 2\left(q_y q_z + q_0 q_x\right) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix}. \tag{C.17}$$

With mastering the quaternion basics, it is relatively straight forward to obtain the optimized rotation by finding the unit quaternion $\mathring{q}$ that minimizes

the term below,

$$\begin{aligned}
\sum_{i=1}^{N} \left(\mathring{q}\mathring{r}'_{\mathrm{A},i}\mathring{q}^{*}\right) \cdot \mathring{r}'_{\mathrm{B},i} &= \sum_{i=1}^{N} \left(\mathring{q}\mathring{r}'_{\mathrm{A},i}\mathring{q}^{*}\mathring{q}\right) \cdot \left(\mathring{r}'_{\mathrm{B},i}\mathring{q}\right) \\
&= \sum_{i=1}^{N} \left(\mathring{q}\mathring{r}'_{\mathrm{A},i}\right) \cdot \left(\mathring{r}'_{\mathrm{B},i}\mathring{q}\right) \\
&= \sum_{i=1}^{N} \left(\bar{\mathbb{R}}_{\mathrm{A},i}\mathring{q}\right) \cdot \left(\mathbb{R}_{\mathrm{B},i}\mathring{q}\right) \\
&= \sum_{i=1}^{N} \mathring{q}^{T}\bar{\mathbb{R}}_{\mathrm{A},i}^{T}\mathbb{R}_{\mathrm{B},i}\mathring{q} \\
&= \mathring{q}^{T}\left[\sum_{i=1}^{N} \bar{\mathbb{R}}_{\mathrm{A},i}^{T}\mathbb{R}_{\mathrm{B},i}\right]\mathring{q} \\
&= \mathring{q}^{T}\mathcal{N}\mathring{q}
\end{aligned} \tag{C.18}$$

, where $\mathring{r}'_i = \mathrm{i}x'_i + \mathrm{j}y'_i + \mathrm{k}z'_i$ is the quaternion to represent a point, and $\mathbb{R}$, $\bar{\mathbb{R}}$ is its corresponding orthogonal 4×4 matrices as shown in Equation C.11, C.12. By setting sum of elementary products between two point sets, $\mathcal{M}$, as

$$\mathcal{M} = \sum_{i=1}^{N} \mathbf{r}'_{\mathrm{A},i}\mathbf{r}'^{T}_{\mathrm{B},i} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \tag{C.19}$$

, the $\mathcal{N}$ can be calculated using Equation C.20, which contains all the information needed to determine the optimized rotation matrix.

$$\begin{bmatrix} S_{xx}+S_{yy}+S_{zz} & S_{yz}-S_{zy} & S_{zx}-S_{xz} & S_{xy}-S_{yx} \\ S_{yz}-S_{zy} & S_{xx}-S_{yy}-S_{zz} & S_{xy}+S_{yx} & S_{zx}+S_{xz} \\ S_{zx}-S_{xz} & S_{xy}+S_{yx} & -S_{zx}+S_{yy}-S_{xz} & S_{xy}-S_{yx} \\ S_{xy}-S_{yx} & S_{zx}+S_{xz} & S_{yz}+S_{zy} & -S_{xx}-S_{yy}+S_{zz} \end{bmatrix}. \tag{C.20}$$

Finally, $\mathring{q}^{T}\mathcal{N}\mathring{q} \leq \lambda_{\max}$ is maximized with $\mathring{q} = \mathring{e}_{\max}$, where $\mathring{e}_{\max}$ is the eigenvector of $\mathcal{N}$ pairing with the most positive eigenvalue $\lambda_{\max}$. A closed-form solution of $\lambda_{\max}$ can be routinely obtained through solving the fourth-order polynomial, $\det\left[\mathcal{N} - \lambda\mathbb{I}\right] = 0$, while $\mathring{e}_{\max}$ obeys the homogeneous equation, $\left[\mathcal{N} - \lambda_{\max}\mathbb{I}\right]\mathring{e}_{\max} = 0$; here $\mathbb{I}$ is the identity matrix. Then, the rotational matrix $\mathcal{R}$ can be constructed using $\mathring{e}_{\max}$ following Equation C.17. In summary, we acquired the optimized absolute orientations, which brings $\{\mathbf{r}_{\mathrm{A},i}\}$ to $\{\mathbf{r}_{\mathrm{B},i}\}$.

PUBLICATION LIST

1. M. Yao, W. Qiu, R. Liu, A. K. Efremov, **P. Cong**, R. Seddiki, M. Payre, C. T. Lim, B. Ladoux, R.-M. Mège, and J. Yan, "Force-dependent conformational switch of α-catenin controls vinculin binding," *Nat. Comms.*, vol. 5, p. 4525. Jul. 2014.

2. M. Yao, B. T. Goult, H. Chen, **P. Cong**, M. P. Sheetz, and J. Yan, "Mechanical activation of vinculin binding to talin locks talin in an unfolded conformation," *Sci. Rep.*, vol. 4, p. 4610, Apr. 2014.

3. S. Le, H. Chen, **P. Cong**, J. Lin, P. Droge, and J. Yan, "Mechanosensing of DNA bending in a single specific protein-DNA complex," *Sci. Rep.*, vol. 3, p. 3508, Dec. 2013.

4. B. Kundukad, **P. Cong**, J. R. C. van der Maarel, and P. S. Doyle, "Time-dependent bending rigidity and helical twist of DNA by rearrangement of bound HU protein," *Nucleic Acids Res.*, vol 41, pp. 8280-8288, Jul. 2013.

5. H. Chen, X. Zhu, **P. Cong**, M. P. Sheetz, F. Nakamura, and J. Yan, "Differential mechanical stability of filamin A rod segments," *Biophys. J.*, vol. 101, pp. 1231-1237, Sep. 2011.

6. H. Chen, H. Fu, X. Zhu, **P. Cong**, F. Nakamura, and J. Yan, "Improved high force magnetic tweezers for stretching and refolding of proteins and short DNA," *Biophys. J.*, vol. 100, pp. 517-523, Jan. 2011.