Molecular Dynamics Study of Supercoiled DNA Minicircles

Tightly Bent and Supercoiled DNA in Atomistic Resolution

Thana Sutthibutpong

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Physics and Astronomy

March, 2015.

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

All the atomistic MD simulation protocols both in implicit and explicit solvents were provided by Dr. Sarah Harris. Simulations in chapter 3 and 4 were performed along with the team members, including Agnes Noy (chapter 3 and 4), Gabriel Slade and Charlie Laughton (chapter 3). Simulations and analysis of the protein-DNA complex in chapter 5 were performed in company with Ilda D'Annessa, while the cryo-ET density maps and computational traced were provided by Rossitza Irobalieva.

Acknowledgements

This thesis, with all the research work it carries, cannot be completed without the help of these people:

First of all, I would like to thank Sarah Harris, who provided me this intriguing research topic, and provided me chances to participated in many of the great research communities, and Nicola Stonehouse, who helped in reading through the manuscript and gave a lot of useful comments.

I thank all our collaborators: Charlie Laughton (Nottingham), Christian Matek (Oxford), Ilda D'Annessa (Rome), Rossitza Irobalieva and Jonathan Fogg (Baylor College, Texas) for all the great jobs we have done together.

Also, very big thanks to Agnes Noy, who helped me through most of the works in chapter 4, and Gabriel Slade, a genuinely great collaborator and friend for his DESIGNED sequence in chapter 3 and a valuable experience in Brazil.

Finally, I would like to thank my parents back in Thailand for their love and lifelong support, my Soft Matter Physics colleagues and all my friends in Leeds for always keeping my spirit high, and special thanks to Kritsadi Thetpraphi for always staying by my side even if being 6000 miles away.

Thana Sutthibutpong

March, 2015.

Abstract

Towards the complete understanding of the DNA response to superhelical stress, sequence dependence structural disruptions on the ~100 base pairs supercoiled DNA minicircles were examined through a series of atomistic MD simulations. The results showed the effects from some subtle structural characteristics of DNA on defect formation, including flexibility at base pair step level and anisotropy, whose dynamic information are available only from atomistic MD simulations. For longer supercoiled DNA minicircles (240-340 bp), the molecules adapt into their writhed conformations. Writhe can be calculated by a Gauss' integral performed along the DNA central axis path. A new mathematical definition for the DNA central axis path was developed for the more exact writhe calculation. Finally, atomistic representation of supercoiled 336 base pairs minicircles was provided by fitting the DNA structure obtained by explicitly solvated MD simulations into the density maps from electron cryo-tomography. Structural data were analysed and provided a decent explanation for the mechanism of the sequence specific binding of the enzyme topoisomerase 1B onto the negatively supercoiled DNA.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Chemical compounds

A adenine

C cytosine

DNA deoxy ribonucleic acid

dsDNA double stranded DNA

H-bond hydrogen bond

FBP FUSE binding protein

FIR FBP-interacting repressor

FUSE far upstream sequence element

G glycosine

NaCl sodium chloride

NCP nucleosome core particle

R purine

siRNA small interference RNA

ssDNA single stranded DNA

T thymine

TADS topologically associated domains

TF transcription factors

Y pyrimidine

Experimental techniques

AFM atomic force microscopy

Cryo-EM electron cryo-microscopy

Cryo-ET electron cryo-tomography

FRET fluorescent resonance energy transfer

IR infrared radiation

NMR nuclear magnetic resonance

SAXS small angle X-ray scattering

TPM tethered particle motion

Units and quantities

Å Angstrom

bp base pair

K Kelvin

kb kilobases

Lk linking number

M molar

ns nanosecond

R_g radius of gyration

Tw global twist

Wr writhe

σ superhelical density

Computational and numerical terms

ff forcefield

GB generalised Born approximation

GB/SA generalised Born surface approximation

MD molecular dynamics

MM molecular mechanics

MPI message passing interface

NPT particle numbers – pressure – temperature

PDB protein data bank

QM quantum mechanics

RMSD root mean square deviation

SD standard deviation

SIDD stress-induced duplex destabilisation

Chapter 1
Introduction to DNA Structure and Topology

1.1) Motivation

Gene expression starts when the information contained within DNA bases is transcribed by RNA polymerase into a messenger RNA (mRNA). The mRNA template then carries the genetic code in the form of base sequences to be translated into the building blocks for functional proteins by the ribosome complexes, containing proteins and ribosomal RNA (rRNA). However, it is not only the chemical sequence of DNA bases and the digital genetic code that is essential for the regulation of transcription, and therefore cellular processes in general. The 3D structure of the compacted genetic material, how this 3D structure reacts to bending and torsional stress, and the effect this ultimately has on the interaction between DNA and proteins, which is also central to genetic control (1). The motivation for this thesis is the development of theoretical and computational methods that consider DNA not only as a passive object to be manipulated by proteins, or as a simple linear matrix of protein binding sites (2), but also as a major contributor to DNA-protein interactions, as the affinity of protein binding to DNA is affected by both the DNA geometry and topology (3).

DNA can be classified as a heterogeneous polymer that contains combinations of four different types of monomer subunits called 'nucleotides'. The first order DNA structure is determined purely from the chemical component of this linear sequence of nucleotides. Second order DNA structure takes into account the precise spatial alignment of nucleobases with their complementary and neighbouring bases, and their sugar-phosphate backbone within the double helical structure of DNA, as described by base step and base pair parameters (discussed in section 1.3). Moreover, DNA is highly polymorphic and capable of forming non-duplex structures under high mechanical stress. The third level in the hierarchy of DNA structure considers topologically restrained geometries, such as DNA in a closed loop or attached to the surface of the protein. These more complex DNA topologies can be 'supercoiled', which means that they are over- or underwound from a relaxed state, which can cause either the helical twisting to accumulate or the DNA helix to become entangled about its own axis (4) (see further discussion in section 1.4).

In order to understand the enhancement or silencing of genes through higher order DNA processes, one requires a complete understanding of DNA mechanics. The advancement of *in vivo*, *in vitro* and *in silico* techniques has provided new evidence of the importance of higher order DNA structure in genomic control, and has shown that placing DNA under superhelical stress can be used to 'fine tune' gene expression to optimise its key cellular processes (4). This thesis focuses on the structure and dynamics of bent and supercoiled DNA at a fully atomistic level and on a nano-second timescale. Small close-looped DNA of different sizes, topological states, and sequences have been used as case studies. Molecular dynamics (MD) simulations, along with the complementary conformational analysis techniques have provided new insight into how DNA reacts to mechanical stress and how this influences the way it interacts with other molecules in atomistic detail.

1.2) Overview

This thesis presents a series of MD simulations of supercoiled DNA minicircles performed during my time as a PhD student at the University of Leeds. The latter parts of Chapter 1 begin with the introduction of the double helical structure of DNA (5). The topology of closed loop DNA will also be discussed (6). Experimental observations of DNA supercoiling and bending in prokaryotes and eukaryotes from the genomic scale (7) down to the atomistic scale is discussed subsequently. Furthermore, due to the high stability of DNA minicircles (8) and their ability to permeate through the cell membrane (9), small supercoiled DNA loops have been proposed as a novel siRNA vectors for cancer gene therapy. The final section of this chapter concerns other *in vitro* studies of the mechanical properties of DNA molecules using emerging optical technologies and single molecule manipulation experiments. Chapter 2 provides a methodological background to computer simulations of supercoiled DNA in full atomistic detail along with examples of the atomistic simulations of mechanically stressed DNA.

Atomistic MD simulations of supercoiled DNA minicircles containing defects were previously performed by Jonathan Mitchell (10) in order to reproduce the cyclisation experiments of Du *et al.* (11). In chapter 3, to investigate the sequence dependence of DNA denaturation under torsional and bending stress, a new series of atomistic MD simulations of these circles and minicircles with modified DNA sequences were performed in close

collaboration with Agnes Noy, Gabriel Slade and Charlie Laughton, and under the supervision of Sarah Harris. In these new simulations, we utilised the modified forcefield parameters provided by Jiri Sponer (12). Results showing the sequence dependence of the DNA response to bending and torsional stress were later compared to those of the statistical mechanical model performed by Craig Benham (13) and those of the 'oxDNA' coarse-grained simulations (14) performed by Christian Matek, under the supervision of Jonathan Doye and Ard Louis.

In Chapter 4, I present a new method for calculating DNA writhing, or superhelicity so that the most exact writhe values can be extracted from atomistic MD data. This work has been done in collaboration with Agnes Noy. WrLINE, the python script to extract the coordinates representing DNA helical axis was developed and then tested with a series of implicitly solvated MD simulations of the writhed DNA (15).

Finally, in Chapter 5, simulations of the writhed 336 bp minicircle structures were performed in parallel with the cryo-ET experiments carried out by Rossitza Irobalieva (supervised by Mike Schmidt and Wah Chiu) and gel electrophoresis carried out by Jonathan Fogg and Jamie Catanese in Lynn Zechiedrich's group at Baylor College of Medicine, Texas. These combined results (16) provide the most exact atomistic representation of supercoiled DNA to date. From the data, structural parameters were extracted and provided a generic perception of the mechanisms behind the deformation of different DNA sequences under superhelical stress. These general principles were also observable in simulations of the topoisomerase 1B protein-DNA complex (performed by Ilda D'Annessa (17)).

1.3) DNA structure

1.3.1) Chemical Components of DNA Double Helix

In the double helical structure of DNA (Figure 1.1a), two chains of nucleotide monomer units are held together by hydrogen bonds between the double-ringed purine (A or G) and the single-ringed pyrimidine (C or T) bases, as shown in Figure 1.1b. Each nucleotide unit consists of a heterocyclic base, a 2'-deoxyribose sugar and a negatively charged phosphate group at the backbone. The sugar and base segments can also be termed as 'nucleoside'. In canonical DNA structures, adenine (A) is paired

with thymine (T) by two hydrogen bonds, and guanine (G) pairs with cytosine (C) through three hydrogen bonds.

Combinations of base sequences encode genetic information for the synthesis of different functional proteins within cells. This first order of structural determination is termed the 'primary structure' of DNA, in analogy with that of amino acid sequences in proteins. The base sequence is conventionally read from the 5′ to 3′ direction. As an example, the first two base pairs in Figure 1.1b can be read as either AC or GT.

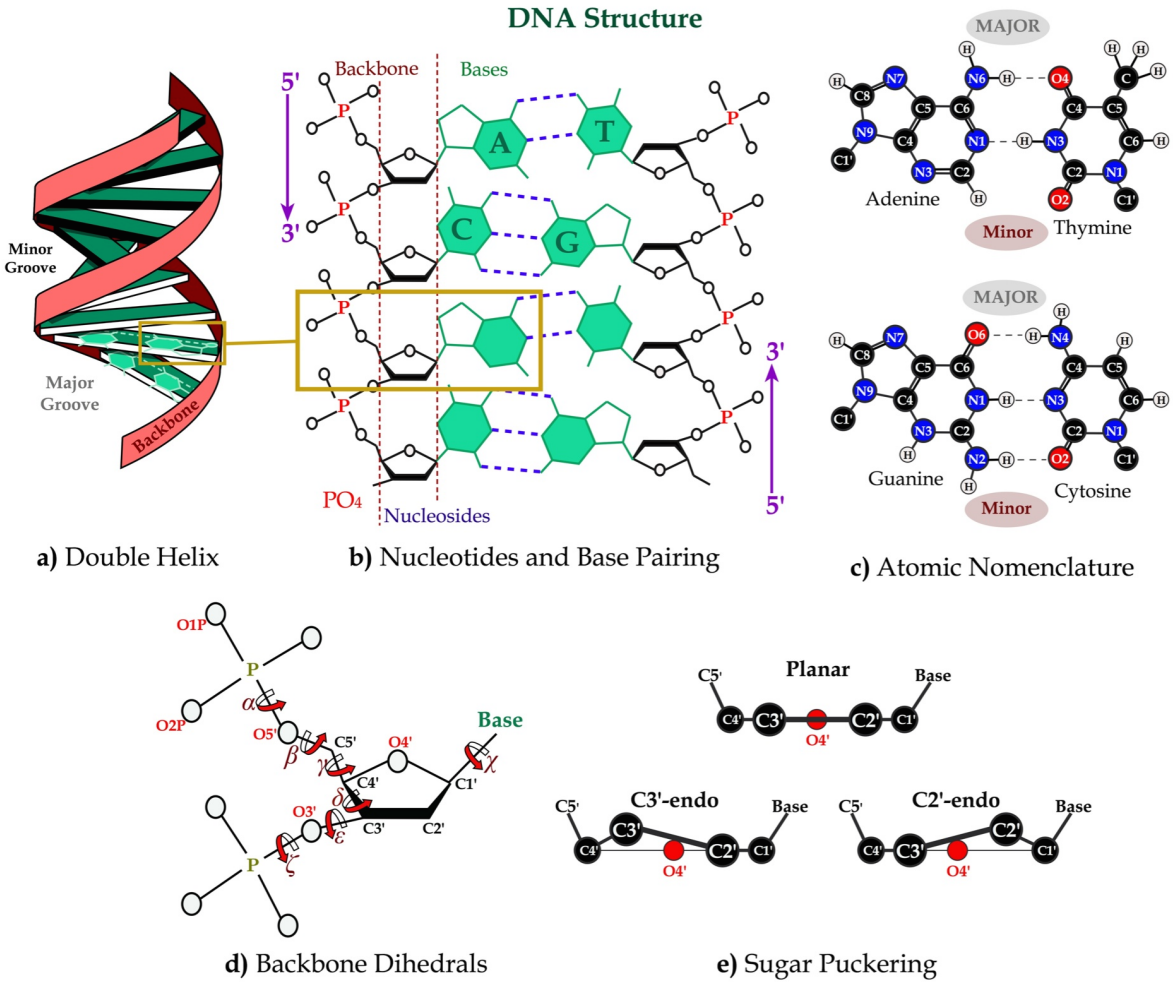


Figure 1.1 (adapted from (18)) Molecular structure of DNA: **a)** The double helix structure proposed by Watson and Crick (adapted from (5)) showing the hydrogen bonded base pairs (green) as a spiral staircase. Two backbones (pink) has formed an anisotropic double helix, showing the major and minor grooves. **b)** Schematic diagram of nucleotide units and base pairing by hydrogen bonds (blue), showing chemical components of the nucleotide chains: sugars and phosphate groups as backbones; sugars and bases as nucleosides (adapted from (6)), **c)** Atomic nomenclature of nucleobases used in atomistic studies of nucleic acids, **d)** Nomenclature for all seven backbone dihedrals of a nucleotide unit, and **e)** Common types of backbone sugar puckering deviated from the planar conformation: C3′-endo (commonly found in A-DNA) and C2′-endo (commonly found in B-DNA)

The next order of determination, or the 'secondary structure' of DNA, concerns the internal structure of each nucleotide monomer unit. Figure 1.1c shows the standard nomenclature for each base type of nucleotides (18). This nomenclature has been used for most of the studies involving atomistic structures of nucleic acids and is particularly useful for the analysis of helical parameters and Watson and Crick hydrogen bonding in this thesis. Each atom associated with a ringed-atom is numbered according to the heavy atom it is bound to; for example, a hydrogen covalently bonded to N1 nitrogen atom of guanine is named H1. The C1′ atoms are parts of the 2′-deoxyribose sugar connected to the backbone strands (19). The two backbone strands are closer together on one side of the helix, the minor groove, than on the other side, the major groove (see both Figures 1.1a and 1.1c). Therefore, DNA is considered as an anisotropic rod as bending into the major groove and the minor groove is not equivalent (1).

The nomenclature of the dihedral angles is also given for the phosphate backbone in Figure 1.1d. For each dinucleotide, six covalent bonds define the backbone, giving rise to the six dihedral angles. The dihedral angles α, β and γ are located in the 5′ direction of 2′-deoxyribose sugar, while the ε and ζ dihedrals are located in the 3′ direction. The δ dihedral angle is located within the sugar ring and influences sugar puckering. Another dihedral angle χ represents the rotation of glycosidic bond, connecting sugar and base. As covalent bonds possess greater torsional flexibility compared to bending and stretching flexibility, conformational diversity of DNA is mainly rooted in the variation of backbone dihedrals. In atomistic DNA modelling, parameterisation of the force constants associated with dihedral angles is important for the stability of DNA structures.

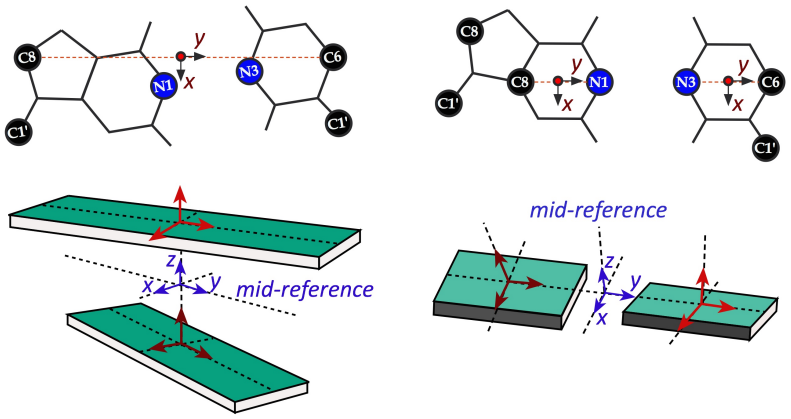
The atomic nomenclature of 2′-deoxyribose sugar rings is described in Figure 1.1d. A 2′-deoxyribose can be classified as a furanose ring, consisting of an oxygen atom and four carbon atoms in the ring. A 2′-deoxyribose sugar (and also ribose for RNA) has another chiral carbon atom covalently bonded to C4′, named C5′. The puckered furanose ring is more energetically favourable than its planar conformation. Ringed atoms can be placed above (referred as *endo*) or below (referred as *exo*) the ring plane. Figure 1.1e shows the C3′-endo and C2′-endo configurations commonly found in canonical A-DNA and B-DNA conformations, respectively.

1.3.2) Base pair and base pair step parameters

Information provided by the atomistic representation of DNA described in Figure 1.1c - 1.1e enables us to reconstruct the complete DNA structure. However, such a fine grained representation involving all the interatomic distances, bending angles and dihedral rotations, makes it difficult to monitor such phenomena as global DNA bending, twisting and defect formation under mechanical stress. Description of DNA atomistic structure can be simplified by treating the heterocyclic bases as rigidly planar objects, so that their relative motions can be tracked.

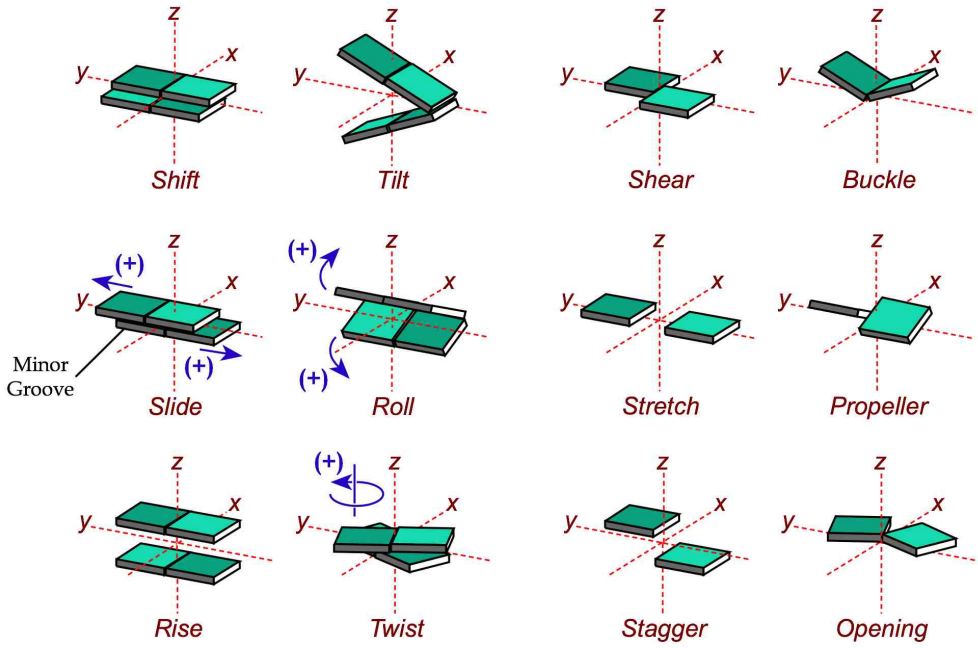
To define the helical parameters for DNA, a coordinate reference frame is assigned to each base pair along the helix and to each base from both strands, as shown in Figures 1.2a (top) and 1.2b (top). The x-axis, y-axis and z-axis lie parallel to the transversal, longitudinal and vertical axes of a regular base pair plane. Relative translation parameters can be directly measured through the displacement between two coordinate origins and can be decomposed into three directional components. However, the rotation vector is non-commutative (it provides different values when operations are performed in a different order) and cannot be decomposed. Therefore, mid-reference frames (Figure 1.2 (bottom)) are defined between each neighbouring base pairs, or base pair step, (Figure 1.2a) and between each complementary bases (Figure 1.2b), so that the rotation angle measurement can be made for each axis relative to the mid-reference frame without vector decomposition (20, 21).

The relative translation and rotation between neighbouring base pairs can be described by the six base pair step parameters shown in Figure 1.3a. The relative translation and rotation between complementary bases can be described by another six base pair parameters shown in Figure 1.3b. These parameters can be used to quantify the bending, over-/under-winding and structural disruptions of DNA, as well as the sequence dependent flexibility.



a) Basepair Step Coordinates b) Basepair Coordinates

Figure 1.2 (adapted from (20)) (***top***) Coordinate reference frames assigned to **a)** Neighbouring base pairs (between the purine C8 atom and the pyrimidine C6 atom of each base pair) for 'base pair step parameters', **b)**. Both complementary bases (purine: between C8 and N1 atoms, pyrimidine: between C6 and N3 atoms) for 'base pair parameters'. (***bottom***) Rigid-base block representation and mid-reference frames (blue) are given for each two neighbouring base pair step and base pair coordinates.



a) Base Pair Step Parameters **b)** Base Pair Parameters

Figure 1.3 (adapted from (20)) Schematic representation of **a)** Base pair step parameters describing the relative translation (*Shift*, *Slide* and *Rise*) and rotation (*Tilt*, *Roll* and *Twist*) between neighbouring base pairs. Sign conventions (blue arrows) are given for positive twist, roll, and slide, **b)** Base pair parameters describing the relative translation (*Shear*, *Stretch* and *Stagger*) and rotation (*Buckle*, *Propeller* and *Opening*) between complementary bases.

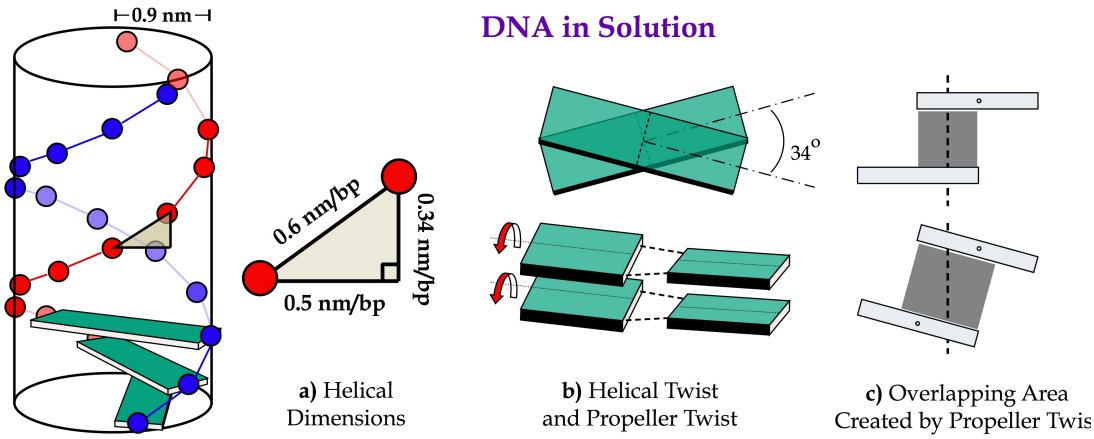
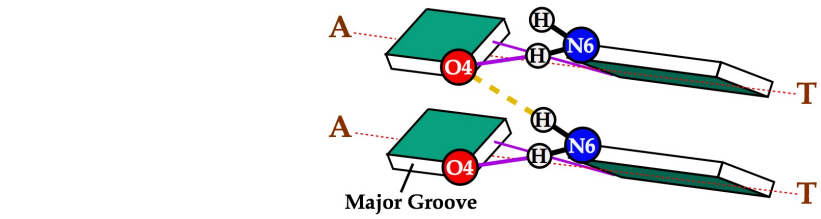


Figure 1.4 (adaped from (19)) Schematic diagrams describing DNA structure in a solution: **a)** Dimensions of a DNA double helix in a solution: the approximated helical radius (~0.9 nm), twisting of the DNA base pairs (represented by green blocks) created by the optimal distance of base stacking (~0.34 nm/bp) and the length of sugar phosphate backbone between neighbouring phosphate groups (~0.6 nm/bp) **b)** DNA helical twist and propeller twist (curved arrows) created to reduce the exposure of base pair to solution. **c)** Lateral view of two stacking bases showing the hydrophobic overlapping area created (top) without propeller twist and (bottom) with propeller twist.

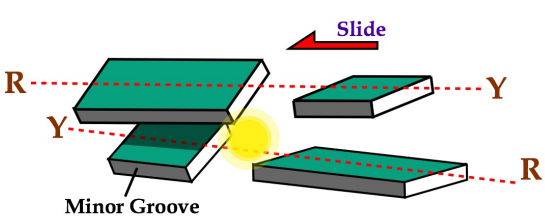
1.3.3) The Conformation of DNA in Aqueous Solution

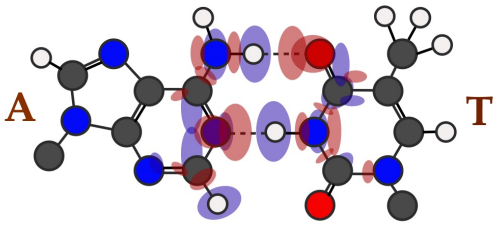
In an aqueous environment, the major influence responsible for the helicity of DNA originates in the electrostatic properties of the nucleotide monomer units. Uncharged nucleobases are hydrophobic in solution. However, the highly charged phosphate backbone is hydrophilic. Thus, phosphate groups are more likely to face towards the solution and wrap around their hydrophobic counterparts. To further reduce the surface area between bases and the external solution, the two DNA backbone strands are twisted into a double helical structure. Another influence on the helicity is the base stacking interactions. The optimal stacking distance between two base pairs is ~0.34 nm but the length of phosphate backbone is ~0.6 nm/bp (see Figure 1.4a). As a consequence, DNA is twisted by approximately ~34°/bp to create adequate space for the backbones connecting two optimally stacked base pairs (19). This helical twist effectively reduces the surface contact between bases and solution as the hydrophilic phosphate groups acting as a shield, preventing the hydrophobic bases inside.

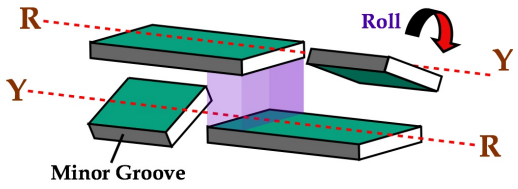
However, base pair twisting still creates a portion of base pair surface area exposed to the solution. To accommodate this, each base on the same strand is rotated clockwise about its longitudinal axis (y-axis) which allows a higher degree of overlap with its neighbours and reduces the surface area exposed to solution (see Figure 1.4b and 1.4c).

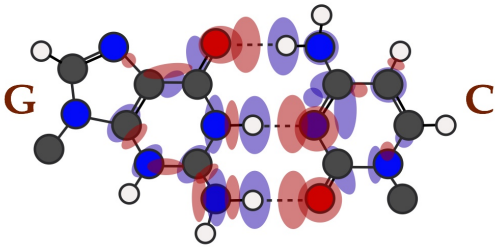


a) Bifurcated H-Bond at a TT/AA step









b) A Bistable Pyrimidine/Purine (Y/R) Step

c) Partial Charge Distribution of A-T and C-G

Figure 1.5 Block representation of **a)** Highly positive propeller twist in an AA step and an additional hydrogen bond (dashed line) formed between the hydrogen atom associated with the N6 atom of a thymine base and the O4 atom of the upper inter-stranded adenine base (adapted from (19)), **b)** Zero 'roll' and negative 'slide' at a pyrimidine-purine (YR) step due to a steric clash (illustrated by a yellow sphere) between purine bases, and positive 'roll' and zero 'slide' due to stacking between purine bases (illustrated by a purple slab) (adapted from (19)), and **c)** Schematic diagram of the electron density distribution of a G-C base pair compared to a relatively non-polar A-T base pair (adapted from a quantum mechanical calculation in (22)).

Propeller twist also results in the sequence-dependence of DNA structures, specifically at AA and YR (pyrimidine-purine) steps. This additional feature influence DNA flexibility as well as the structure. Figure 1.5a shows an AA dinucleotide base pair step, in which the propeller twist brings an oxygen atom O4 and an amino group containing a nitrogen N6

into close proximity. An additional hydrogen bond can be formed between a donor O4 and an acceptor H6-N6 which enhances the DNA stiffness. AA steps appear to have approximately zero 'roll' and 'slide' with small fluctuations (19). Figures 1.5b show the two alternate configurations of a YR step. In Figure 1.5b (top), the dinucleotides maintain their intra-strand stacking interactions so that there is no bending (zero 'roll'). The propeller twist leads to a steric clash between purine bases and the upper base pair needs to slide towards the left (negative 'slide'). Figure 1.5b (bottom), shows another form of the YR step, in which inter-strand stacking between purine bases is stabilised and keeps the approximately zero 'slide', while the intra-strand stacking is sacrificed, resulting in a positive 'roll' that widens the minor groove. These interchangeable configurations enhance the local flexibility of YR steps.

Another sequence-dependent effect arises from the different electrostatic properties of the base pairs. An A-T base pair has a relatively homogeneous electron distribution and therefore neutrally charged and non-polar compared to a G-C base pair, as illustrated in Figure 1.5c. For the base pair steps containing two G-C base pairs (CG, GC and GG/CC), especially the GG/CC steps, the two dipolar base pairs tend to avoid stacking directly on top of each other. This results in either positive or negative, non-zero 'slide' (19).

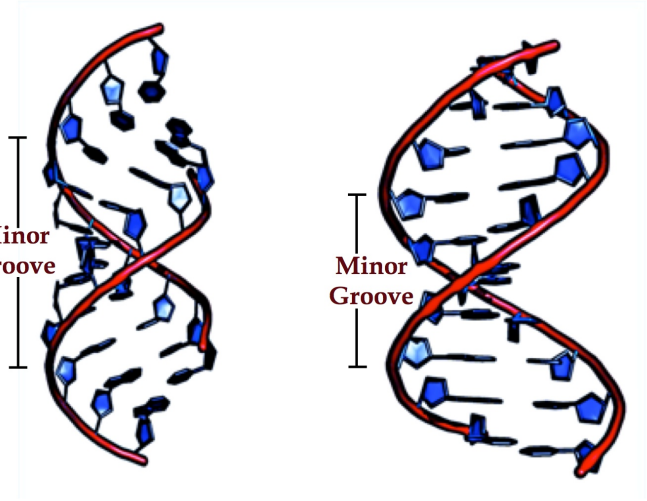


Figure 1.6 Atomistic representation of (left) A-DNA and (right) B-DNA structures generated by an AMBER module NAB (23). Shown in dark blue inside the double helices are the nucleoside parts, in which the non-parallel and parallel base pair orientations of A- and B-DNA are affected by C3′-endo and C2′-endo, respectively. The negatively charged helical backbones (red curves) have formed the anisotropic double helices. Minor grooves are shown in the left-hand side of each structure. The larger minor groove was observed in A-DNA (left panel).

1.3.4) A-DNA and B-DNA

Figure 1.6 shows the two main types of right-handed double helices. B-DNA is most commonly found in aqueous solution, and is characterised by the almost parallel base stacking and small minor groove (approximately zero 'roll' and 'slide'). The A-DNA conformation is observed in the more dehydrated conditions, in which the reduced screening effect of water and positively charged counterions causing more direct electrostatic repulsion between the phosphate groups and the minor groove opening (positive 'roll').

The variety of sequence dependent base pair step conformations mentioned in Section 1.3.3 gives rise to differences in the structure and flexibility of double helices constructed from different primary sequences. AA/TT steps are hardly ever found in A-form, as they maintain zero 'slide' and 'roll'. However, the pyrimidine-purine steps (YR: TA, CA/TG and CG) in Figure 1.5b(top) shows a configuration in which the steric clash between two purines promotes negative 'slide' and positive 'roll', and which therefore has a tendency to develop into A-form. CG/CG are the highest

possible base pair steps to develop an A-form due to its non-zero 'slide' bimodality.

These subtle sequence dependent effects on the structure and flexibility of DNA are important in enabling external biomolecules, such as proteins, to discriminate between the various DNA sequences and interact with specific DNA sites (24, 25). Sequence dependence effect that comes from different unique base pair steps and their response to bending and superhelical stress will be further discussed in Chapter 3 and Chapter 5.

1.4) DNA Topology

The higher order of DNA structural determination involves its global shape and topology, in which a closed loop DNA in Figure 1.7a can be represented by two catenated space curves (see Figure 1.7b). The so called 'tertiary structure' of DNA can be characterised by the invariant linking number (*Lk*), defined by how many times the two strands wrap around each other under topological constraints of forming a closed circular structure, or when the ends of the DNA are restrained by binding to a surface (e.g. through attachment to a protein, which may in turn be embedded in the nuclear membrane in eukaryotes). *Lk* is unaffected by the local contributions from changes in secondary structure. However, the internal changes might affect global twisting (*Tw*) and writhing (*Wr*). *Tw* is defined by the total rotation of a DNA strand about its central axis and *Wr* describes the folding of DNA about itself. The constant linking number can be decomposed into the interchangeable global twist and writhe, as $Lk = Tw + Wr$. Experimental and theoretical studies have been carried out to assess the partitioning between global twist and writhe (26–29), and the response of DNA to supercoiling, quantified by difference in linking number (ΔLk).

1.4.1) Linking Number, Twist, and Writhe

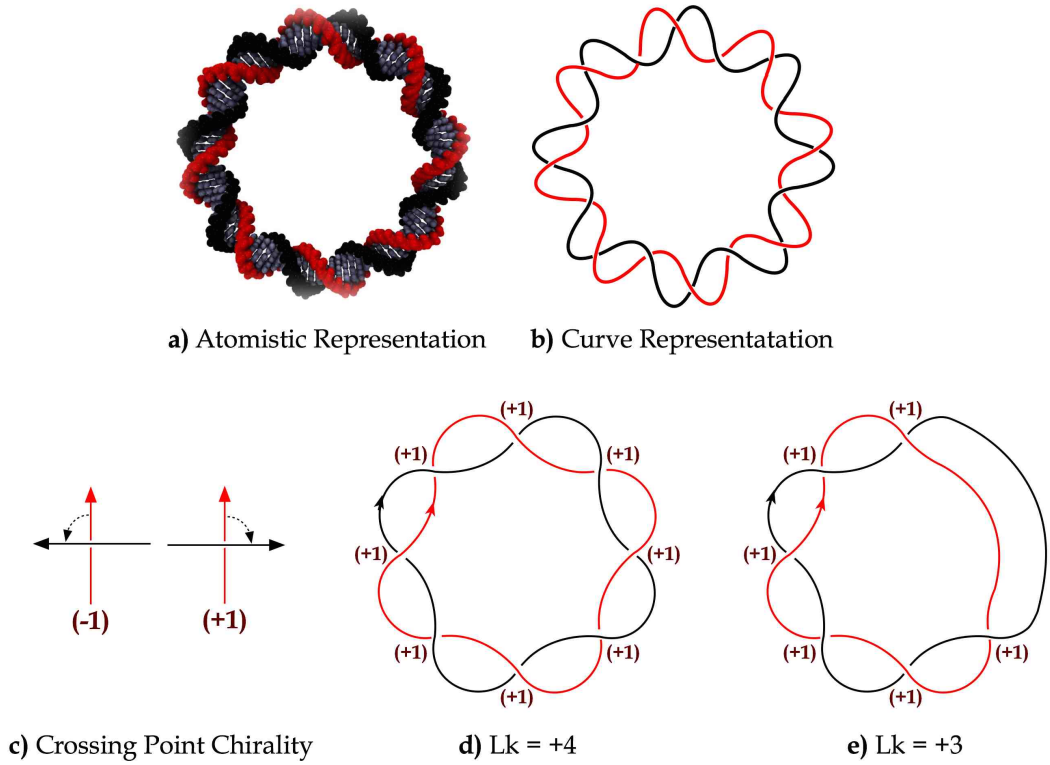


Figure 1.7 a-b) Atomistic and backbone curve representations of a minicircle DNA molecule projected on a 2D plane, **c)** Crossing point chirality: for any two crossing segments viewed from a 2D plane with their directions given, the crossing point is negative/positive when the direction of the top segment (black) points anti-clockwise/clockwise relative to the bottom segment (red), **d)** Two catenated loops with 8 positive crossing points representing *Lk* = +4, **e)** A strand passage has been introduced into the *Lk* = +4 DNA. Two positive crossing points disappear and *Lk* becomes +3 (adapted from (6)).

Linking number can be found by counting the crossing points of the two projected catenated curves shown in Figure 1.7b. Sign conventions of the crossing point chirality are given in Figure 1.7c, in which the two crossing curve segments are given their directions. The point has a positive (+1) crossing when a segment points in the clockwise direction relative to the segment underneath it and a negative crossing (-1) when the top segment is rotated anti-clockwise. Figure 1.7d shows a circular loop of a right-handed double-helix consisting a series of eight positive crossings, which gives *Lk* = +4. This right-handed double-helix can represent a B-DNA loop, although the helical bending is unrealistically sharp. In Figure 1.5e, suppose that the minicircle DNA in Figure 1.7d has engaged in

topoisomerase activity, which breaks a DNA strand and allows strand passage. One positive turn (two positive crossing points) has been removed from the minicircle and *Lk* becomes +3.

The difference between the linking number (*Lk*) of the topoisomer and the pseudo-linking-number of the unconstrained linear DNA (Lk_0) is used to quantify DNA supercoiling. The linking difference $\Delta Lk = Lk - Lk_0$ can be decomposed into $\Delta Lk = \Delta Tw + \Delta Wr$ as the superhelical stress can either cause additional helix twisting or helical axis folding. Figure 1.8 elucidates the interchangeability of twist and writhe using the ribbon representation. In Figure 1.8a, a linear ribbon with no twist and $Lk_0 = 0$ is given. The ribbon is then twisted anti-clockwise from the top so that global twist becomes +1. The supercoiled loop is then made from the twisted ribbon by joining its ends. Two edges of the looped ribbon are the closed catenated loop with two positive crossings, therefore the linking number $Lk = +1$ and linking difference $\Delta Lk = +1$ are fixed. However, the expected values for ΔTw and ΔWr are between 0 and +1 and are not necessarily integral. Figure 1.8a shows two extreme cases: 1) when twist makes the entire contribution to ΔLk, $\Delta Tw = +1$ and $\Delta Wr = 0$, and 2) when twist creates no contribution and the ribbon is positively writhed, $\Delta Tw = 0$ and $\Delta Wr = +1$.

Figure 1.8b shows an example of the supercoiled ribbon that looks closer to DNA. Number of helical turns for the relaxed unconstrained ribbon is $Lk_0 = +4$. The ribbon is overtwisted by one turn and created a minicircle of with $Lk = +5$. Thus, the minicircle is positively supercoiled with $\Delta Lk = +1$. In this example, the ribbon can either remain planar or buckle into the positively writhed form.

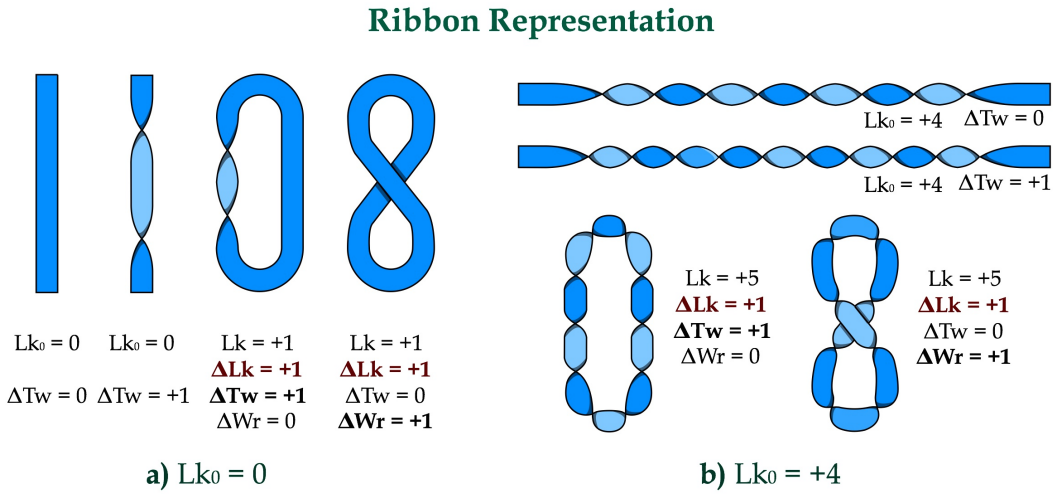
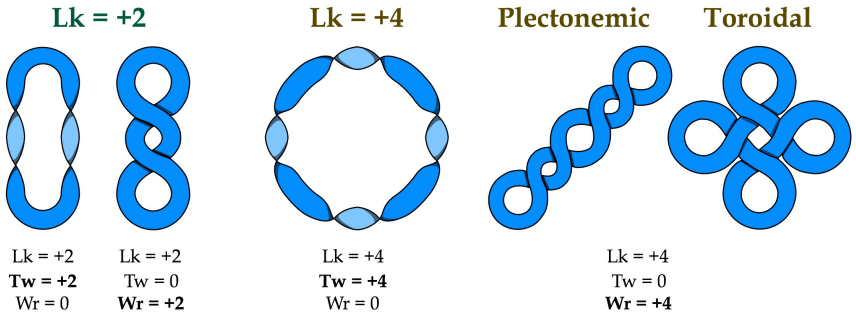
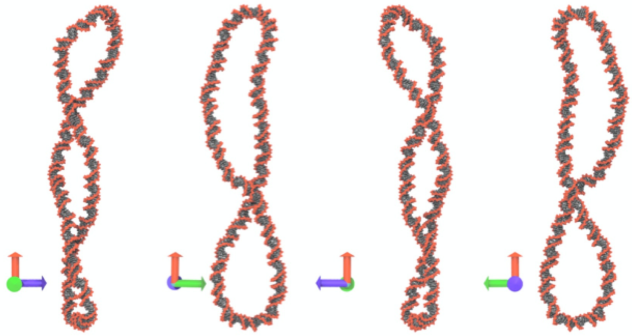


Figure 1.8 Ribbon representation of DNA minicircles: **a)** A relaxed ribbon with Lk_0 = 0 and ΔTw = 0 is twisted for one turn, so that ΔTw = +1. Then, a supercoiled Lk = +1 loop is made from the ΔTw = +1 ribbon, which allows any amount of helical writhing between ΔWr = +0 (where ΔTw = +1) and ΔWr = +1 (where ΔTw = 0) and **b)** A ribbon with relaxed helical turn of Lk_0 = +4 and ΔTw = 0 is over-twisted, so that ΔTw = +1. Then, a supercoiled Lk = +5 (ΔLk = +1) loop is made from the ΔTw = +1 ribbon, which allows any amount of helical writhing between ΔWr = +0 (where ΔTw = +1) and ΔWr = +1 (where ΔTw = 0).

Writhed loops at ΔLk = +2 and ΔLk = +4 are shown in Figure 1.9a. At ΔLk = +4, writhing can occur either in plectoneme (e.g. bacterial DNA) or toroidal (e.g. nucleosomes) forms. Writhe can be evaluated by the total number of crossing points using the chirality defined in Figure 1.5a. However, as we consider a writhed DNA minicircle in 3D, the numbers of crossing points from different viewing orientations can be unequal as shown in Figure 1.9b. As a solution to this problem, the writhe can be found by averaging the numbers of crossing points obtained from every possible viewing angle. This writhe calculation is known as the 'directional writhe' method. The minicircle shown in Figure 1.9b has either -1 or -2 crossing points depending on the viewing angle, which suggests that the actual writhe value is between -1 and -2.



a) Writhing at Lk = +2 and Lk = +4



b) Supercoiled DNA from different angles

Figure 1.9 a) Writhing of Lk_0 = 0 ribbons at the higher orders: ΔLk = +2 and +4 supercoiled ribbons showing the extreme cases where *Tw* or *Wr* entirely contribute to the *Lk*. For $\boldsymbol{\Delta Lk = +2}$, supercoiling of the looped ribbon allows the *Wr* values between 0 and +2 (where two positive crossing points can be seen). For $\boldsymbol{\Delta Lk = +4}$, supercoiling of the looped ribbon allows the *Wr* values between 0 and +4 and writhing can be in either plectonemic or toroidal forms. **b)** A 3D coordinate snapshot captured from a ΔLk = -2 supercoiled DNA minicircle. Different numbers of helical crossings (-1 or -2) have been rendered from four different angles, which implies the writhe value of this supercoiled DNA structure to be -2 < *Wr* < -1.

1.4.2) Calculation of Linking number, Twist, and Writhe

A physical system can be used as a topological analogy for the linking number of two catenated space curves (Figure 1.10a). An electric current I in a solenoid circuit represented by a curve S_2 generates a magnetic field in a vacuum. According to the Biot-Savart law, the magnetic field at any point on a closed path S_1 can be expressed by

$$\vec{\boldsymbol{B}}(s_1) = \frac{\mu_0 I}{4\pi} \oint_{S_2} \frac{(\vec{\boldsymbol{r}}_2(s_2) - \vec{\boldsymbol{r}}_1(s_1))}{|\vec{\boldsymbol{r}}_2(s_2) - \vec{\boldsymbol{r}}_2(s_1)|^3} \times \hat{\boldsymbol{t}}_2(s_2)\, ds_2 \,. \qquad 1.1)$$

With the circulation of the magnetic field described by Ampere's law:

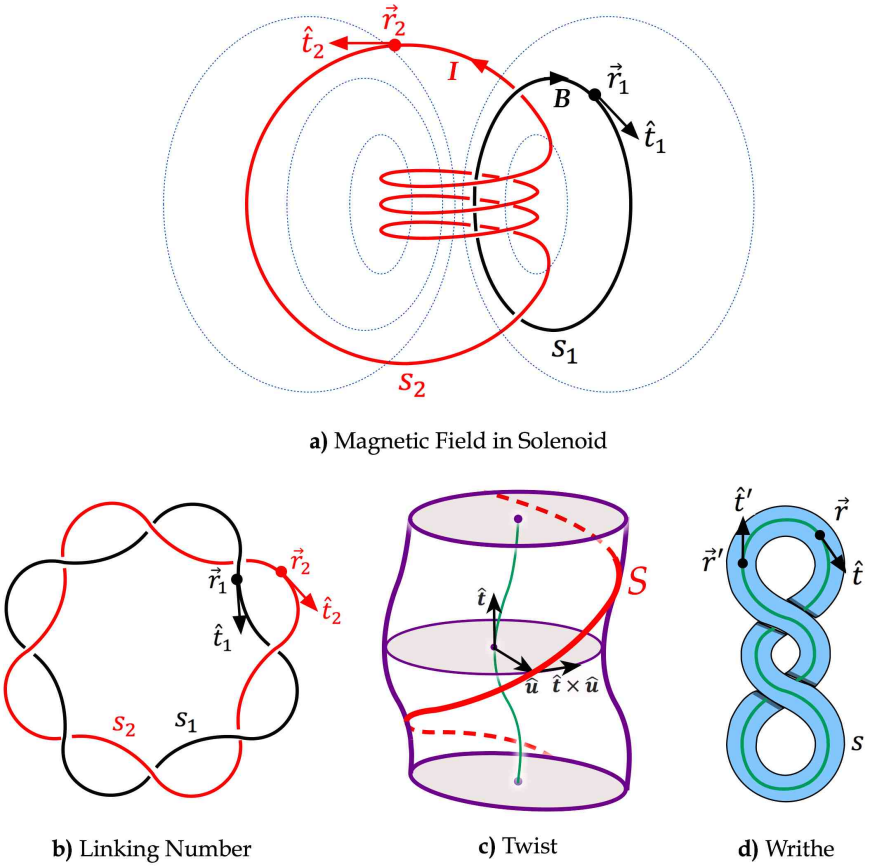
$$\mu_0 nI = \oint_{S_1} \vec{\boldsymbol{B}}(s_1) \cdot \hat{\boldsymbol{t}}_1(s_1)\, ds_1 \,, \qquad 1.2)$$

substituting the magnetic field from Biot-Savart's law into Ampere's law gives an expression for the linking number of the curves S_1 and S_2, which is equal to the number of solenoid turns, n:

$$Lk = \frac{1}{4\pi} \oint_{S_2} \oint_{S_1} \hat{\boldsymbol{t}}_2(s_2) \times \hat{\boldsymbol{t}}_1(s_1) \cdot \frac{(\vec{\boldsymbol{r}}_2(s_2) - \vec{\boldsymbol{r}}_1(s_1))}{|\vec{\boldsymbol{r}}_2(s_2) - \vec{\boldsymbol{r}}_2(s_1)|^3} \, ds_1 ds_2 \,. \qquad 1.3)$$

Global twist is defined by the total rotation of a helix curve or a ribbon edge about the central axis:

$$Tw = \oint_S (\hat{\boldsymbol{t}} \times \hat{\boldsymbol{u}}) \cdot \frac{d\hat{\boldsymbol{u}}}{ds} ds \qquad 1.4)$$



a) Magnetic Field in Solenoid

b) Linking Number **c)** Twist **d)** Writhe

Figure 1.10 a-b) Mathematical descriptions for a topological analogy made between the Linking number of **a)** A magnetic field loop S_1 (black) generated by an N = 4 solenoid current loop S_2 (red) and **b)** A 4-turn DNA minicircle backbones represented by curves S_1 and S_2. The vectors $\vec{\boldsymbol{r}}_i(s_i)$ and $\hat{\boldsymbol{t}}_i(s_i)$ represent the position and tangent direction of any point s_i on the curve i. **c)** Describe the global twist calculation in the Equation 1.4, where a loop integral is performed over the infinitestimal rotation $(\hat{\boldsymbol{t}} \times \hat{\boldsymbol{u}}) \cdot d\hat{\boldsymbol{u}}/ds$ of strand S about the central helical axis (green) **d)** A ribbon representation describing the writhe calculation in the Equation 1.5 over the central helical axis (green), where the similar double integration as in the linking number calculation is performed on one same central axis curve (green). (adapted from (6))

For each point along the helical curve S, $\hat{\boldsymbol{t}}$ describes the tangential direction of the curve S, $\hat{\boldsymbol{u}}$ describes the direction from the central axis to the helix S (see Figure 1.10c), and $d\hat{\boldsymbol{u}}/ds$ describes rotation of the helix about the central axis. For a linear DNA molecule, $d\hat{\boldsymbol{u}}/ds$ is parallel to the cross product $\hat{\boldsymbol{t}} \times \hat{\boldsymbol{u}}$, and integration gives $Tw = \theta/2\pi$, where θ is the total rotation angle about the central axis swept by the vector $\hat{\boldsymbol{u}}$.

From the definition of writhe as the 'self-coiling' of the central axis, a mathematical analogy can be made between writhe and linking number so that the double integration in Equation 1.3 can be performed along the same central axis path (see Figure 1.10d):

$$Wr = \frac{1}{4\pi} \oint_S \oint_{S'} \hat{\boldsymbol{t}}(s) \times \hat{\boldsymbol{t}}(s') \cdot \frac{(\vec{\boldsymbol{r}}(s) - \vec{\boldsymbol{r}}(s'))}{|\vec{\boldsymbol{r}}(s) - \vec{\boldsymbol{r}}(s')|^3} \, ds'ds \,, \tag{1.5}$$

Definition of the central axis is crucial for obtaining the exact writhe values, as will be discussed in the Chapter 4.

1.4.3) Superhelical Density

Superhelical density is defined by the normalisation of the linking difference $\Delta Lk = Lk - Lk_0$ by the total number of helical turns of the torsionally relaxed DNA (Lk_0) (6):

$$\sigma = \frac{(Lk - Lk_0)}{Lk_0} \,. \tag{1.6}$$

Consider a torsionally relaxed minicircle of N_0 base pairs and linking number Lk, the number of base pair per helical turn for torsionally relaxed DNA is $b_0 = N_0/Lk$. Therefore, the total number of helical turns for torsionally relaxed DNA of length N base pairs ($Lk_{0,N}$) can be found by:

$$Lk_{0,N} = \frac{N}{b_0} \,, \tag{1.7}$$

Hence, at the same linking number (Lk_0), the superhelical density of an N base pair minicircle can be found by

$$\sigma = \frac{Lk - Lk_{0,N}}{Lk_{0,N}} \tag{1.8}$$

$$\sigma = \frac{N_0 - N}{N} \tag{1.9}$$

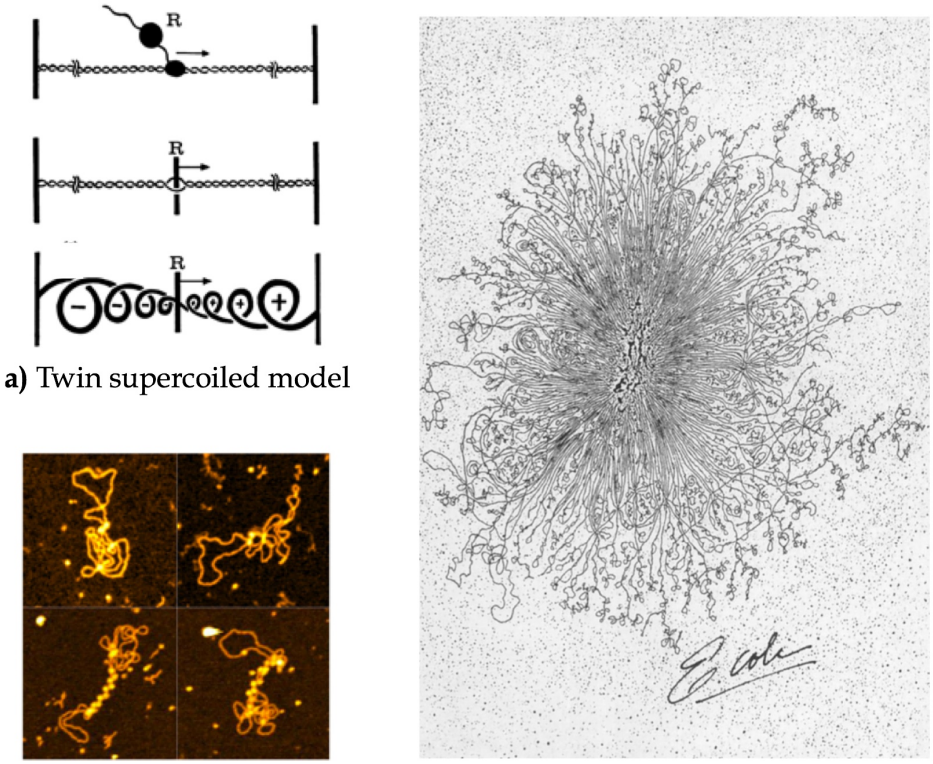
which implies that adding more base pairs into a minicircle at fixed linking number reduces its superhelical density.

1.5) DNA supercoiling from genomic to atomistic scales

DNA is a dynamic molecule *in vivo* and is rarely in its relaxed state. Alteration of the helical twisting as a result of torsional stress generated by genetic processes such as transcription and protein binding (e.g to histones) forces the DNA into a supercoiled state. As discussed in section 1.4.1, not only the global twist (Tw) changes, but also the DNA may adopt a writhed conformation. This property is fundamental to DNA and chromatin (30) and plays a key role in chromosome organisation. This section discusses the studies that have been carried out to observe DNA supercoiling and the effects of superhelical stress propagation, ranging from genomic scale, where supercoiling was measured along the topological domains and supercoiling domains of chromatin, down to the scale where the structure of a nucleosome core particle unit can be viewed at atomistic resolution.

1.5.1) Supercoiling Domains and Topological Domains

In active cells, genetic information stored within the base sequence is expressed through the transcription process. An RNA polymerase unwinds and progresses through the opened DNA strands. As a consequence, positive (overwinding) superhelical stress is generated ahead of the RNA polymerase and negative (underwinding) superhelical stress is behind it, as shown in the 'twin supercoiled domains' model proposed by Liu and Wang (31, 32) (see Figure 1.11a). Torque generated by the ongoing transcription machinery is transmitted along the molecule. Torsional stress would dissipate if the DNA strand was not fixed to boundary and was free to rotate, hence no effect would be transmitted to the DNA elements located far from the transcription site. In the presence of topological constraints, however, superhelical stress can be trapped in the topological domains created by architectural boundary proteins. Figure 1.11b shows the AFM images of looped topological domains created by the *E.coli lac* repressor (LacI) proteins bound to plasmids containing tandem repeats of protein-binding sites (33). LacI proteins at the boundaries are able to trap superhelical stress generated by transcription. At the genomic scale, a bacterial nucleoid is partitioned into topological domains as shown in the TEM image in Figure 1.11c.



a) Twin supercoiled model

b) Topological domains

c) *E.coli* genome

Figure 1.11 Genomic compartmentalisation and DNA supercoiling **a)** The twin supercoiled model (reproduced from (31)) showing an elongation stage of a transcription as an RNA polymerase (R) progressing along a DNA domain with two ends tethered, generating positive and negative supercoiling, **b)** AFM images of topological domain loops created by binding of *E.coli lac* repressor (LacI) proteins (bright dots) on small DNA plasmids embedded with *lac* operators (reproduced from (34)), and **c)** A TEM image of an isolated *E.coli* nucleoid bound by cytochrome C and stained with uranyl acetate showing topological domains branching from the core created by DNA binding to nucleoid associated proteins (TEM image by Ruth Kavenoff).

1.5.2) Genomic Scale: Hi-C and FISH Experiments

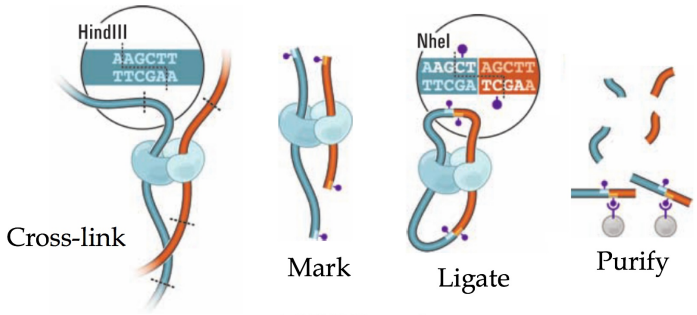
The genomic organisation of prokaryotic and eukaryotic cells has been explored through Hi-C deep sequencing experiments. As shown in Figure 1.12a, pairs of DNA segments in close proximity are cross-linked and digested by restriction enzyme. The digested ends of the cross-linked segments are marked by biotin and re-ligated. Then, the cross-linked, biotin-marked fragments are purified and collected together to create a Hi-C

library. Analysing this library using massively parallel sequencing provides the loci for interacting chromatin regions (35).

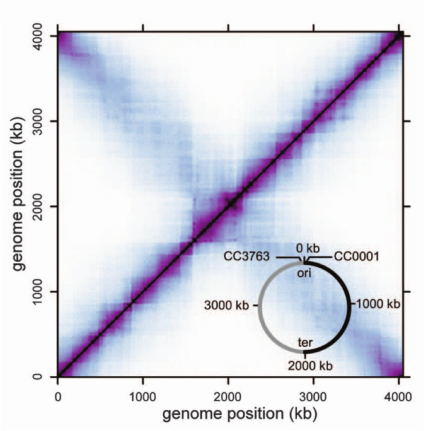
The Hi-C experiments by Lieberman-Aiden *et al*. were carried out on human chromosomes at 1 Mb resolution. The results were consistent with a knot-free 'fractal globule' polymer, which is densely packed but possesses the ability to freely fold and unfold (36). This was followed by Hi-C studies with higher resolution (bin sizes < 100 kbase) by Dixon *et al*., which demonstrated the presence of topological associated domains (TADs) within the self-interacting regions of a mammalian chromosome (37).

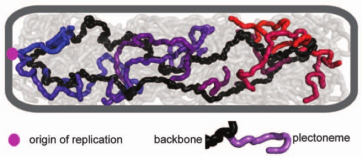
An additional Hi-C experiment was carried out on a 4 Mb bacterial genome at 10 kb resolution. In the contact map shown in Figure 1.12b (top), smaller contacts detected at the off-diagonal elements are the contacts between right and left chromosomal arms about the replication origin. The self-interacting regions at the main diagonal elements are the chromosomal interaction domains (CIDs), the prokaryote counterpart of TADs. Highly expressed genes were found closed to the CID boundaries, which are the plectoneme free regions (PFRs) (see Figure 1.12b (bottom)).

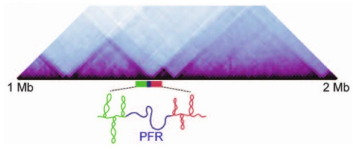
An alternative approach to examine chromosome organisation is the Fluorescent *in situ* Hybridisation technique. DNA probes are introduced into the denatured chromosomal DNA and the physical location of the DNA sequences complementary to the DNA probes can be identified *in situ* (38, 39). Pools of bacterial artificial chromosome (BAC, ~1 Mb size) were used as probes to locate and measure the three dimensional distances between genes (see Figure 1.12c). The results showed the correlation between the counts of chromatin contacts and the three dimensional distance as the chromatin regions were more co-localised within the same topological domain (green) than in the different domains (red) (40). Together with the HiC data, FISH experiment showed that the genomic organisation is not random, but is partitioned into the succession of TADs.



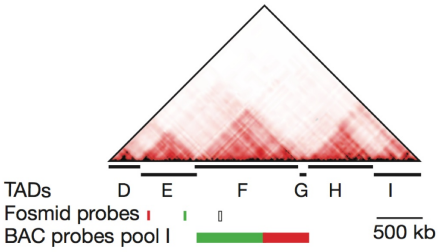
a) HiC Experiment

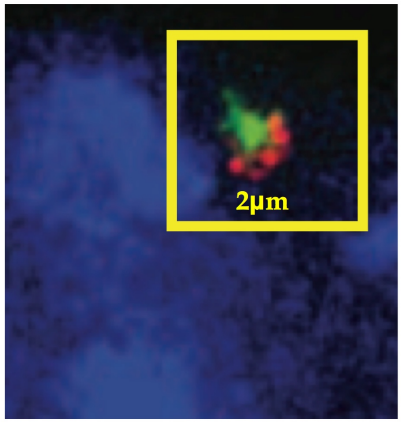






b) Bacterial Genome



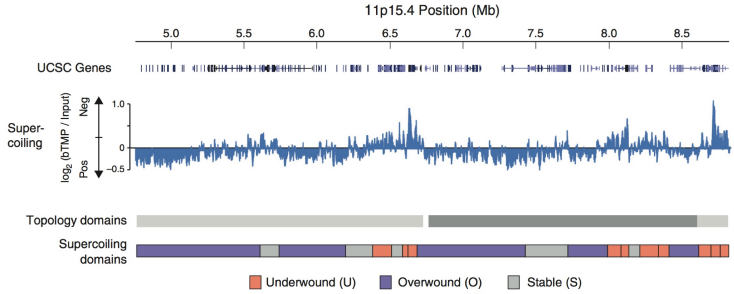


c) HiC - FISH

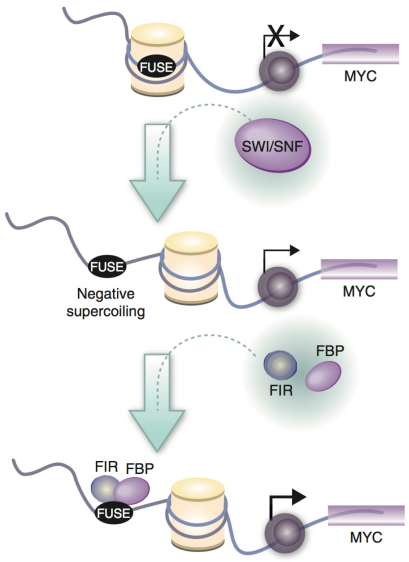
Figure 1.12 a) Schematic diagrams showing the procedure of a Hi-C experiment (reproduced from (35)), **b)** Correlation map showing the amount of chromatin contact within a 4 Mb *caulobacter* genome (middle) polymer model of the same chromosome structure (bottom) relatively small amount of contacts found at a plectoneme free region (PFR) at a chromosomal interacting domain (CID) boundary (reproduced from (41)), **c)** (top) A correlation map showing chromatin contacts within a gene containing six topological associated domains (D-I) and (bottom) The spatial position of the biggest topological domain (F), consisting of a large cluster (green) and a number of smaller clusters (red) corresponding to the contact map above, determined by a FISH experiment (reproduced from (40)).

1.5.3) Superhelical Stress Propagation and the Supercoiling Responsive FUSE Element

A striking demonstration of the importance of DNA supercoiling in gene regulation is provided by recent experiments on 'transcriptional bursting' in bacteria. The occurrence of repetitive stochastic transcription events were monitored in real time through the fluorescent emission of SYTO, an RNA selecting dye. The results clearly showed that positive supercoiling generated ahead of the transcription complexes slows down the transcriptional elongation stage. However, the transcription was resumed after DNA gyrases were introduced (42).



a) Supercoiling distribution over supercoiling and topological domains



b) Consequences of transcription driven negative supercoiling on the FUSE element

Figure 1.13 a) DNA supercoiling distribution measured from the binding of biotinylated-psoralen (bTMP) on a segment of human chromosome 11, showing the supercoiling domains within the two topological domains (reproduced from (43)), **b)** Transcription of c-myc gene facilitated by the nucleosome remodeling complex (SWI/SNF) and its effect on the melting of far-upstream sequence element (FUSE), which can be detected by FBP and FIR proteins (reproduced from (30)).

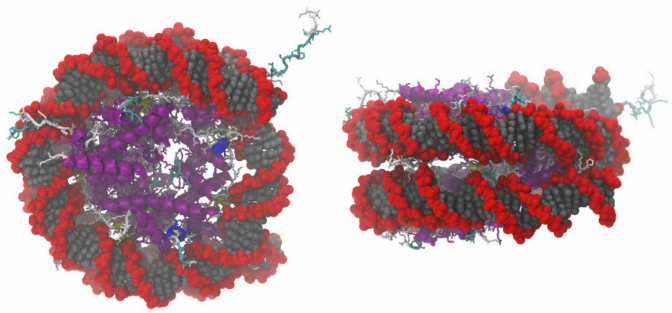
As demonstrated by the twin supercoiled domains model, torsional stress driven by transcription is able propagate through DNA topological domains within chromosomes. This can be measured by psoralen binding experiments (7, 44, 45). Psoralen is a compound that preferentially binds to negatively supercoiled DNA, so the amount of psoralen binding can estimate DNA supercoiling at the genomic scale, including positive and negative supercoiling domains generated by a transcription machinery. An example of the DNA supercoiling distribution mapped by psoralen binding is shown in Figure 1.13a, which shows that there can be a number of positive and negative supercoiling domains lying within a single topological domain (43).

At the kilobase scale, the transfer of torque generated by a transcription complex of c-myc facilitated by the SWI/SNF nucleosome remodelling proteins (see Figure 1.13b), a regulator gene coding a transcriptional factor, can be sensed by the FUSE or 'Far Upstream Sequence Element' located at 1.4 kilobases upstream from the transcription site (45, 46). A FUSE element contains a very high content of A-T base pairs making it sensitive to torsional stress, and is prone to forming a single-stranded region. The FUSE element can be used as a biological 'detector' for supercoiling as single-stranded regions can be detected by the FUSE Binding Protein (FBP), which can further induce the strand opening by increasing the c-myc transcription rate, and FBP-Interacting Repressor (FIR), which represses the transcription mediated by FBP (47). Formation of the ssDNA regions in FUSE elements has shown that the response of DNA to superhelical stress is sequence dependent. In Chapter 3, we examine the mechanical properties of the FUSE element by embedding the sequence into supercoiled DNA minicircles.

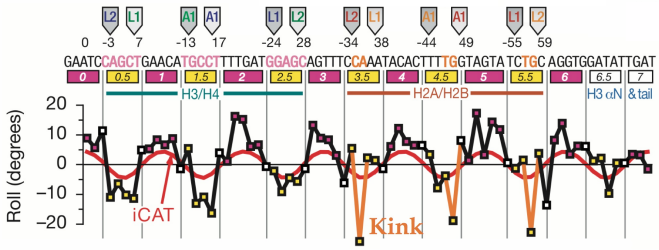
1.5.4) DNA Bending at Sub-Persistence-Length Scale: NCPs and CAP-DNA Complexes

Tight DNA bending can be seen in nucleosome core particles (NCPs). Shown in an x-ray crystal structure in Figure 1.14a, a nucleosome consists of 147 bp DNA which makes ~1.84 toroidal turns around a histone octamer (48). The tight loop of ~80 base pairs per turn is shorter than the DNA persistence length (~150 bp). Bending an object at sub-persistence-length scale requires a large amount energy penalty, and sufficient bending stress

can induce kinked DNA structures. Kinking observed in the x-ray crystal structure in Figure 1.14a can be quantified by 'roll'. In Figure 1.14b, the highly negative roll values (orange) at CA/TG steps compared to the ideal roll values coherent with register angle phase (iCAT: red) signified the dramatic minor groove closure at these steps. Anisotropic bending of nucleosomal sequences into the major groove and the minor groove is termed as 'major groove blocks' and 'minor groove blocks' denoted by pink and yellow blocks in Figure 1.14b. The terms were mentioned in later studies on nucleosomal DNA structures (49–54).



a) Nucleosome Structure



b) Roll parameter values

Figure 1.14 a) X-ray crystallographic structure of a nucleosome core particle [PDB number: 1AOI] showing 147 bp DNA in black (bases) and red (backbone). Alpha helices of the histone core are shown in purple and a histone tail can be seen protruding to the right hand side of the nucleosome, b) Measurement of roll angles (black lines) between each of the neighbouring base pairs along the nucleosomal DNA. The red sinusoidal curve represents the ideal value of roll corresponding to the bending directions: towards major groove (major groove blocks; pink) or towards minor groove (minor groove blocks; yellow). Highly negative roll at the kinked sites are shown in orange (reproduced from (48)).

Moreover, the anisotropic kinking was also apparent in an interaction between DNA and catabolite activator protein (CAP), a dimeric transcriptional activator, of which monomers bind to DNA major groove. Kinking of a CA step located between two CAP binding sites in the absence

of protein contact increases CAP-DNA binding affinity. This recognition process in known as 'indirect readout' (55). DNA bending anisotropy can also be seen in the DNA sequence with an intrinsic bending property, such as A-tract (long runs of AA/TT base pair steps) (56, 57). When phased A-tract sequences alternate with other random sequences, this creates intrinsic curvature within the DNA segment without the need for an external bending force (58), which for example can affect the stability of DNA looped by LacI protein (59).

1.6) Biophysical studies on DNA mechanics

'Passive' experiments on DNA mechanics monitor structure and dynamics without exerting external forces, while the 'active' methods specifically measure how a single DNA molecule responds to an applied force (60). Bending and torsional rigidity can also be assessed by cyclisation experiments and any structural disruptions due to the imposed stress within the tight DNA loops can be detected by endonuclease enzymes. Finally, the formation of highly stressed DNA circles have potential biomedical applications, as supercoiled DNA minicircles have been shown to have potential as novel gene therapy vectors.

1.6.1) Passive methods: SAXS, FRET, AFM, Cryo-EM, and TPM

SAXS and FRET: tracking end-to-end distance through end labelling

Figure 1.15a shows a schematic diagram of optical spectroscopy methods that assess DNA dynamics through labelling the ends. Small-angle X-ray scattering experiment (SAXS) measured the end-to-end length between the gold nano-particles site specifically attached to the 3'-thiol modified deoxycytidine residues. Variance of the distance measured from SAXS implied that stretching (61) and bending (62) flexibility agreed with the worm-like chain model for these linear DNA fragments containing between 40 and 100 base pairs.

Fluorescence resonance energy transfer (FRET) experiments monitor distances between pairs of donor and acceptor fluorophores. The resonance energy emitted from an excited fluorophore pair is proportional to $1/r^6$, where r is the distance between the fluorophore pairs. Fields *et al*. tethered the donor and acceptor fluorescence dyes onto a 'molecular vise' created by a hairpin loop DNA, which bends the targeting short double-stranded DNA. The results showed that free energy required to kink the double helix is

~21k_BT (63). The FRET technique was also used in the single molecule cyclisation studies (discussed in Section 1.6.3).

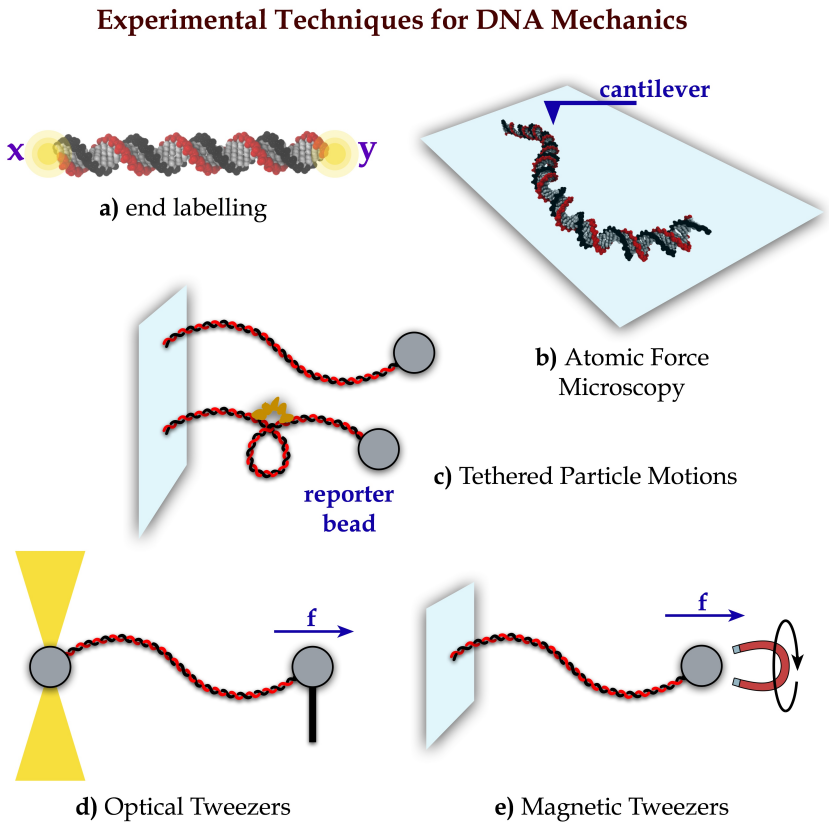


Figure 1.15 (adapted from (60)) Schematic diagrams of *in vitro* techniques to investigate DNA mechanics: a) The methods assessing DNA dynamics by end-labelling. **x** and **y** represent the interacting tethered probes: donor and acceptor fluorophores in FRET experiments or gold nanoparticle scattering agents in SAXS experiments, b) Atomic force microscopy (AFM): A cantilever scans across a 2D surface and a DNA fragment, c) Tethered Particle Motion (TPM): Analysing the Brownian motion of the reporter beads tethered to a free linear DNA or a protein-mediated looped duplex, d) Optical Tweezers: a bead is trapped by counter-propagating laser, while another is pulled by an AFM cantilever, and e) Magnetic Tweezers: DNA is tethered to a surface and is stretched and twisted by a magnetic bead.

AFM: 2D imaging of DNA on a surface

Information about the shape and flexibility of DNA on a 2D surface can be obtained through Atomic Force Microscopy (AFM). As shown in Figure 1.15b, a cantilever tip scans across the surface and produces feedback signals as it is displaced when the molecule is detected. Statistics obtained from the ensemble of AFM measurements can provide information on sequence-dependent DNA curvature and flexibility. The enhanced DNA flexibility was found in A-T rich sequences (64, 65), including the TATA-box

sequence located within promoter regions, forming ssDNA during transcription initiation (66). The AFM study of Wiggins *et al.* demonstrated the distinct mechanical properties of DNA at shorter length scales (~5 nm), in which DNA was bent sharper than would be expected for a perfectly harmonic model (67). AFM imaging has been used extensively to investigate protein-DNA interactions, particularly the mechanism of DNA transcription, which has been the subject of a recent review (68).

Cryo-EM: 3D imaging through the reconstruction of electron density profiles

Three-dimensional structural information on DNA can be obtained from the reconstruction of cryo-electron microscopic (cryo-EM) images. DNA sample is immobilised in liquid ethane within a micron size microscopic grid by rapid freezing into the amorphous ice state. A two-dimensional image of the molecule is provided by the contrast of electron density at the sample compared to surrounding area. Then, multiple images from different incident angles can be reconstructed into a three-dimensional shape. Bednar *et al*. observed the plectonemic writhed conformation of the 178 bp minicircles obtained from the lox-cre recombination on a larger plasmid (26). A subsequent cryo-EM study by Demurtas *et al*. reported that a smaller DNA minicircle (94 bp) did not contain any sharp kinks, unless the DNA was nicked (69), in disagreement with controversial DNA cyclisation experiments (70). A subsequent cryo-EM study of the unwound ~100 bp minicircles showed, however, that in the presence of negative supercoiling stress, cooperative kinking can occur in which two defects form at opposite positions within the minicircle to minimise the bending energy of the looped DNA (71).

Recently, cryo-electron tomography (cryo-ET) has been applied to DNA minicircles, as described in more detail in Chapter 5. Cryo-ET provides images of the sample by sectioning, which provides improved 3D information. The cryo-ET experiments in Chapter 5 have provided 3D tomographic images of 336 bp supercoiled DNA minicircles at different linking numbers (16).

TPM: statistical analysis on the motion of a single molecule

Another technique that combines the study of single molecules with statistical analysis is tethered particle motion (TPM). A single DNA molecule has one of its terminals tethered onto a microscope slide and another

terminal tethered onto a bead, allowing Brownian motion to be tracked. sequence, bracketed with lac operators and looped by a lac repressor protein (see Figure 1.15c). The results showed that the poly-TA sequence has the highest propensity for looping (72).

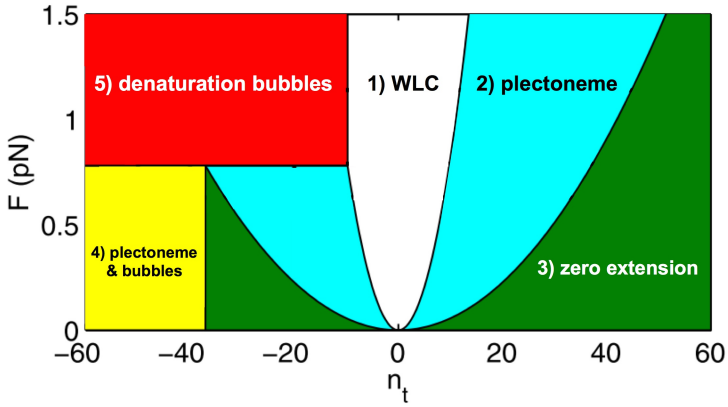


Figure 1.16 Phase diagram of the DNA molecules being stretched and twisted by magnetic tweezers, showing five regions in the force-torsion space representing different phase behaviours: 1) DNA behaves as a semi-elastic polymer in accordance with the WLC model, 2) DNA writhes into a plectonemic form, 3) DNA writhes further until it can no longer be extended, 4) The co-existence of plectonemes and bubbles was observed, and 5) Denaturation bubbles were observed while the plectoneme loops were relaxed.

1.6.2) Active methods: optical and magnetic tweezers

Single DNA molecules can be trapped, tethered and manipulated (stretched and/or twisted) by optical and magnetic tweezers, which provides direct observations of the response to mechanical stress. In an optical tweezers experiment (see Figure 1.15d), DNA is tethered between two polystyrene beads. While one bead at a DNA terminus is trapped within counter-propagating laser beams, another bead is connected to the applied force.

As shown in Figure 1.15e, in a magnetic tweezers experiment, a DNA molecule is tethered on a glass surface and manipulated by a super-paramagnetic bead. The bead can either be stretched or twisted by the magnet, and the applied force can be analysed from the Brownian motion of the bead. A magnetic tweezer experiment by Salerno *et al*. (73) revealed the anisotropic and multiphase behaviour of supercoiled DNA in force-torsion (F, n) space, where F is the applied stretching force and n is the number of over-/under-twisting turns (see Figure 1.16), as has been previously reported by atomistic MD (74), and subsequently using coarse-grained DNA models (75). In spite of the fact that these experiments are on linear DNA,

the anisotropic behaviour observed is generally similar to that reported for atomistic simulations (28) and gel electrophoresis experiments (16) on DNA minicircles with different sizes and superhelical densities.

1.6.3) Experimental Studies on DNA Cyclisation and Structural Disruptions

DNA cyclisation is another technique which examines the bending and torsional rigidity of DNA. The goal of DNA cyclisation experiments is to determine the ring-closure probability of DNA fragments, or the *j*-factor, which is related to the bending and torsional persistence lengths (76). The *j*-factor can be evaluated from the ratio between the amount of circular and linear fragments at a very early stage ($\lim t \to 0$) (77). Combination of single molecule cyclisation and FRET techniques was able to track the real time population of looped DNA (78, 79). The *j*-factor calculated for the cyclisation of 67-106 base pairs DNA were shown to differ from the values predicted by an elastic model (76). However, *j*-factor alone cannot provide the evidence for whether the enhanced loop-closing probability were from the formation of kinks or from the bending rigidity at sub-elastic level (78).

To observe structural disruptions within small DNA loops, enzymatic experiments were carried out with DNA of sizes 63-205 bp. BAL-31 and S1 endonuclease enzymes, capable of digesting ssDNA, were used to detect any kinks or denaturation bubbles that occurred during cyclisation. Torsionally relaxed 65 bp minicircles were digested, suggesting that bending stress alone is sufficient to induce kink formation at ~3.5 nm radius of curvature, while no digestion by endonucleases was observed for the larger 84 bp and 106 bp torsionally relaxed minicircles. However, as the 100 bp minicircles were negatively supercoiled, they were also digested by the enzymes, which suggested that torsional stress promoted structural disruptions (11). The results were consistent with the cryo-EM observations (71). Subsequently, MD simulations were performed to obtain atomistic structural information for the disrupted DNA minicircles (10, 80). As shown in Chapter 3, we have recently performed additional simulations of these minicircles, and have compared them with minicircles of the same size but different DNA sequences to assess the sequence dependence of kink and bubble formation.

1.6.4) Application: supercoiled DNA as minivectors for gene therapy

One of the examples of exploiting the mechanical properties of supercoiled DNA is found in the development of cancer gene therapy. A recent study of supercoiled DNA minicircles (9) has proposed a new method to transfect the siRNA (small interference RNA) into a lymphoma cancer cell using supercoiled DNA minicircles. siRNA is a short RNA targeting the complementary mRNA and forms a double-stranded RNA region, which interrupts the translation, silences the targeted gene and later causes cell death. Previously, viral vectors were used as carriers of siRNA. However, the use of viral vector has been abandoned because of their potential danger (81). Alternatively, using DNA plasmids to deliver the siRNA into cancer cells, has the advantage that they are highly biostable (82), able to reproduce many siRNAs in the form of shRNA (short hairpin RNA) by transcription and relatively safe compared to the viral vector. The only limitation of the plasmid vectors are their size, as relatively large plasmid molecules (>3000 base pairs) possess low transfection efficiency. Therefore, there have been attempts to reduce the size of DNA vectors into minicircles, while still containing essential sequences for replication and vector propagation. Small supercoiled minicircles were shown to be more resistant to shear stress arising from hydrodynamic forces (8) and led to enhanced transfection efficiency *in vitro* (9). The physical properties of one of the minicircles from these studies are described by the MD simulations in Chapter 5, and the 3D structure of the simulated circles is compared to the global structural data from cryo-ET images.

Chapter 2
Introduction to MD Simulation of Duplex DNA

2.1) MD Simulations

Atomistic structural data of DNA can be obtained either from X-ray crystallography or NMR. Both methods, however, have their own limitations. X-ray crystallography is able to provide atomistic structures of linear DNA fragments and protein-DNA complexes, including nucleosome core particles (24, 48). However, systematic error may be introduced by the experimental conditions in the crystallisation process. Moreover, crystal structures lack dynamic information, which can be obtained by nuclear magnetic resonance (NMR) spectroscopy. Distances between pairs of atoms evaluated from the intensity of the NMR resonance peaks then undergo a distance geometry calculation that can create an ensemble of the possible molecular configurations (83, 84). However, the limitation of NMR techniques is when the DNA chains get longer, thermal fluctuation increases and thus softens the resonance peaks.

MD simulations have shown to be a useful tool as they can provide the dynamic properties from the static crystallographic data for timescales up to ~40 μs for a dodecamer (85). Moreover, MD is capable of providing the structure and dynamics of supercoiled DNA of the size up to 336 bp at a ~10ns timescale (see Chapter 4 and 5). In an ideal world where the 'forcefield' parameterisation of all the interatomic forces in the system is perfect, molecular dynamics (MD) simulations perform the role of a microscope with an atomistic resolution (~1 Å), capable of tracking down the motion of all atoms in the systems (86). Later in this chapter, the molecular mechanics forcefield for biomolecules, along with the energy minimisation and molecular dynamics simulation algorithm will be briefly described.

2.1.1) Biomolecular Mechanics Forcefields

Since the DNA molecules considered in this thesis contain more than 100 base pairs (200 nucleotide residues, or about 6000 atoms), quantum mechanical calculations are far too computationally expensive to be feasible for the study DNA dynamics. Thus, electron motions are ignored and only

the nuclear position is considered for each atom. A molecular mechanics forcefield is a set of parameters for all the covalent and non-covalent interactions for all atoms in the system. The molecular mechanics energy of the two sample molecules in Figure 2.1, calculated by a simple forcefield, can be expressed in a functional form as (87):

$$
\begin{aligned}
E(\vec{\boldsymbol{r}}^{\boldsymbol{N}}) = & \sum_i \frac{h_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_i \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 \\
& + \sum_i \sum_{p=1}^{3} \frac{V_i}{2}\left(1 + \cos\left(p\omega - \gamma\right)\right) \\
& + \sum_{m=1}^{N} \sum_{n=m+1}^{N} 4\epsilon_{mn}\left[\left(\frac{\alpha_{mn}}{r_{mn}}\right)^{12} - \left(\frac{\beta_{mn}}{r_{mn}}\right)^{6}\right] \\
& + \sum_{m=1}^{N} \sum_{n=m+1}^{N} \frac{q_m q_n}{4\pi\epsilon_0 r_{mn}}
\end{aligned}
\qquad (2.1)
$$

where h_i ,k_i and V_i represent force constants for stretching $\left(l_i - l_{i,0}\right)$, bending $\left(\theta_i - \theta_{i,0}\right)$ and dihedral torsion $\sum_{p=1}^{3}(1 + \cos\left(p\omega - \gamma\right))$ for each covalent bond i, while ϵ_{mn}, α_{mn}, β_{mn}, q_m and q_n are the parameters describing Van der Waals and electrostatic interactions between atoms m and n separated by the distance r_{mn}.

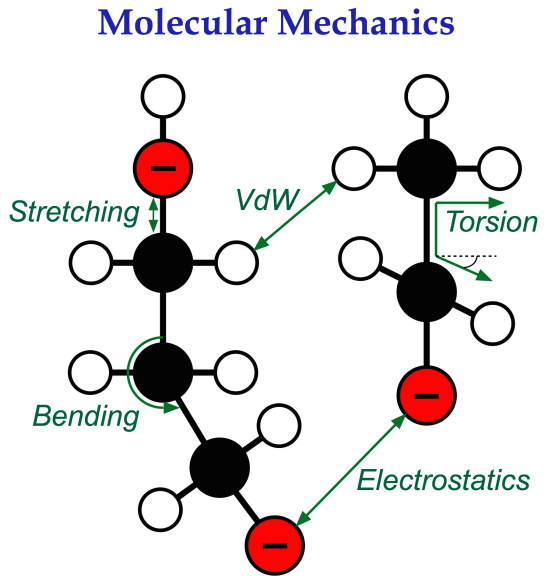
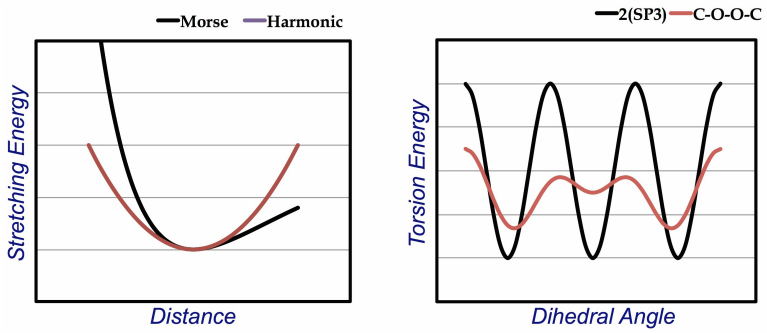


Figure 2.1 Schematic representation of a polymer molecule with covalent (*stretching*, *bending*, and *torsion*) and non-covalent (*VdW* and *Electrostatics*) molecular mechanics interactions as described in Equation 2.1 (adapted from (87)).

Bonded Forcefield Terms

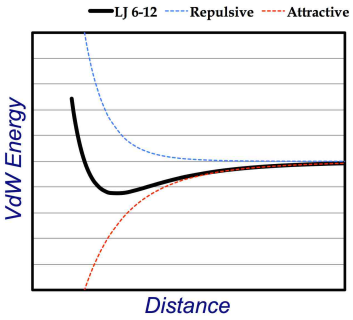
The first three terms are the covalent interactions, representing the stretching, bending and torsion harmonic energy of covalent bonds respectively. The other two terms are the non-covalent interactions: the 12-6 Lennard-Jones potential for Van der Waals interactions and the Coulombic electrostatic interactions. The parameter terms must all be specified by referring to experimental data or quantum mechanical calculations (88).



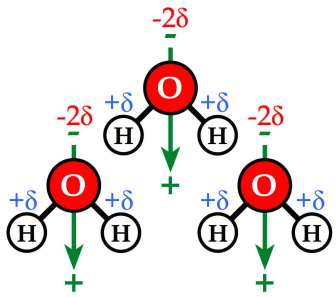
a) Morse and Harmonic Stretching **b)** Different Dihedral Profiles

Figure 2.2 a) Quadratic potential energy profile of the stretching of a covalent bond (red) is compared to the more realistic Morse potential (black). The two functions are consistent for a small stretching about the local energy minimum, **b)** Comparison between two different torsional potentials: one with a single sinusoidal term (black) describing the torsional potential energy of a covalent bond associated with two SP3 hybrid orbitals (e.g. ethane); another (red) with an additional term to describe the electrostatic environment of a C-**O-O**-C dihedral (adapted from (87)).

As electrons are treated implicitly, no covalent bonds can be broken and no additional bonds can be formed during the simulation. The information on the covalent bonding networks must therefore be provided before running a simulation. The topology file of a DNA system must contain all the force constants for the covalent bonds within each nucleotide unit. Energy penalty of non-equilibrium bond length stretching and angle bending can be approximated by Morse potentials, but quadratic potentials are more frequently used for their simplicity in biomolecular forcefields (see Figure 2.2a). Dihedral rotation is relatively soft compared to stretching and bending, therefore DNA can develop a wide variety of conformations due to changes in dihedral angle. The functional form of the dihedral angle for a covalent bond can be expressed by the linear combination of sinusoidal functions, depending on the type of that covalent bond (single, double or triple), constituent atoms and electrostatic environments, as shown in the examples in Figure 2.2b.



a) 6-12 Lennard-Jones Potential



b) Partial Charges of Water Molecules

Figure 2.3 a) A 12-6 Lennard-Jones potential function (black) describing the Van der Waals interaction as the sum of an attractive potential (red) and a weaker repulsive potential (blue), **b)** Schematic diagram showing the alignment of three water molecules on a 2D plane, their dipole moments (green), negative partial charges of oxygen atoms due to their relatively high electronegativity and positive partial charges of oxygen atoms (adapted from (87)).

Non-covalent Forcefield Terms

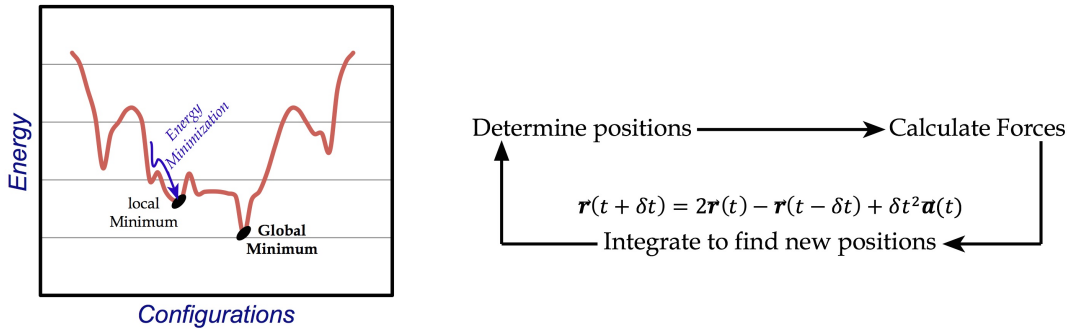
The basic non-covalent interaction terms include the Van der Waals interactions and the Coulomb electrostatic interactions. Van der Waals interactions consist of the attractive forces due to the instantaneously induced dipoles and the repulsive forces due to the Pauli's exclusion principle. Combination of attractive and repulsive Van der Waals forces can be approximated by the 12-6 Lennard-Jones function (Figure 2.3a), providing strong repulsive force at a very small distance, very weak attractive force at a long distance and a local energy minimum for each interacting pair of atoms. Lennard-Jones constants were firstly parameterised to reproduce the motion of classical fluid (89). Further parameterisations were then made towards the thermodynamic parameters of many different biological systems (88). The electrostatic interaction energy is calculated by Coulomb's Law. The approximated partial charge of each covalently bonded atom is determined by its electronegativity or its propensity to attract electrons. Figure 2.3b shows a system of dipolar water molecules. For each molecule, the oxygen atom possesses higher electronegativity than the two hydrogen atoms. Thus, oxygen atoms contain negative partial charges and hydrogen atoms contain positive partial charges. Numerical values of partial charges can be obtained from quantum mechanical calculations (88).

2.1.2) Energy Minimisation and Molecular Dynamics of Circular DNA

The detailed simulation protocols for MD are provided in the Appendix B. To generate a closed circular atomistic DNA starting structure, base pair coordinates are obtained from the standard structural database and reference frames (90, 91). Coordinate transformations produce the uniformly twisted and bent structure. However, DNA structures built by this approach are hypothetical. The accumulation of local strain within the non-equilibrium covalent bond lengths and bending angles brings about structural instability. Therefore, prior to an MD simulation, the hypothetical starting structure needs to be energetically optimised to prevent numerical instability in the simulation.

Biomolecules, including DNA, contain a large number of degrees of freedom, resulting in a very rough energy landscape. Thus, a system progressing through an energy landscape in Figure 2.4a is more likely to be trapped in a local minimum valley and cannot reach the global minimum state. However, for the DNA systems within this thesis, a local energy minimum state in general represents a DNA conformation whose bond lengths and bending angles are adequately corrected that they provide a good starting point for an MD simulation, which will then remain numerically stable over sufficient timescales.

Minimisation methods used in this thesis include the derivative-based algorithms, implemented in both the GROMACS and AMBER packages (92, 93). Derivative-based algorithms are capable of solving a system of linear equations, applicable for energy minimisation problems where the iterative line searching is performed for all the unconstrained atoms. For each iterative step in the steepest descent algorithm, a line searching is defined by the direction of net force (negative energy gradient) acting on each atom, providing a simple conformational search but with some errors. A line searching in the conjugate gradient algorithm also takes into account the information from the direction of the previous move, enabling it to provide exact solutions for any quadratic equations (94). The minimisation protocol generally used in typical atomistic MD simulations begins with the more robust and computationally cheap steepest descent method. Then, the conjugate gradient method is performed to approach the most exact solutions with greater accuracy.



a) Energy Minimisation

b) Molecular Dynamics

Figure 2.4 a) An example of 1D energy landscape along a reaction (configurations) coordinate, showing local and global minima. An energy minimisation (purple) is performed, so that the system configuration changes towards a local minimum. **b)** A basic MD simulation algorithm cycle. Forces are calculated from all the given atomic positions. Numerical integrations are then performed in order to determine the updated atomic positions for the new cycle. (adapted from (87))

After the mechanical energy of the molecule has been minimised, the structure is ready for a molecular dynamics run. The basic algorithm for molecular dynamics is described in Figure 2.4b. Each atom in the starting structure is assigned an initial random velocity at the first timestep. Then, all bonded and non-bonded potential energy and forces acting to each atom are calculated from the molecular mechanics function shown in Equation 2.1 and its gradient at each degree of freedom. Following that, the velocities and positions of all atoms are updated by performing numerical integrals on forces (89, 95), and a similar process is continued as a loop. Additional conditions can be implemented into the calculations to increase their effficiency and to stabilise the system. For example, at the start of each simulation, solute atoms are equilibrated by restraining their starting structure to prevent the impact caused by initial random velocities driving the system into an unstable region of conformational space. The position restraints are set to gradually decrease as the simulation progresses. Moreover, the cut-off distance can be set so that the force calculations can only be performed between each atom and its close neighbours to reduce computational costs without creating significant errors. Temperature and pressure can also be regulated to create a controlled environment, mimicking *in vivo* or *in vitro* situations. The simplest way to keep a constant temperature is to rescale all the atom velocities for their total kinetic energy

to match the temperature, and the volume of the whole system can also be rescaled to match a constant pressure (96).

2.1.3) Implicit Solvent Simulations

In an explicitly solvated simulation, the system is surrounded by water molecules and counter-ions. Since this method represents the solvent environment in full atomistic detail, it provides the highest possible accuracy for the structural representation at the classical level. A disadvantage of this method, however, is its computational expense as forces associated with all atoms need to be calculated. Moreover, friction imposed by the solvent also slows down the molecular motion and exploration of conformational space. Thus, for a simulation that requires only global parameter data, such as writhing and radius of gyration (see Chapter 4), an approximate method using a continuum representation for the solvent electrostatics can be used to increase the computational efficiency.

This continuum approach considers two parts of solvent free energy: non-electrostatic and electrostatic:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel} \qquad 2.2)$$

The non-electrostatic term describes the negative (favourable) energy penalty from Van der Waals interactions between the molecule surface and the solvent, and the positive (unfavourable) energy penalty caused by the cavity within the solvent which the solute has come to replace. The AMBER program estimates the non-electric term to be proportional to the surface area of solute molecule accessible by the solvent. Surface area is calculated by the linear combination of pairwise orbital (LCPO) algorithm (97).

For the electrostatic term, the Generalized-Born (GB) approximation is used as an alternative calculation method instead of the Poisson-Boltzmann equation due to the computational efficiency of the GB method. The GB electrostatic free energy between a pair of solute atoms of partial charges q_i and q_j within the continuum solvent environment can be expressed analytically by:

$$\Delta G_{el} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}\left(1 - \frac{\exp\left(-\kappa f_{GB}\right)}{\epsilon}\right) \qquad 2.3)$$

The exponential term incorporates the electrostatic screening effects, as κ represents the Debye-Huckel screening parameter to account for the

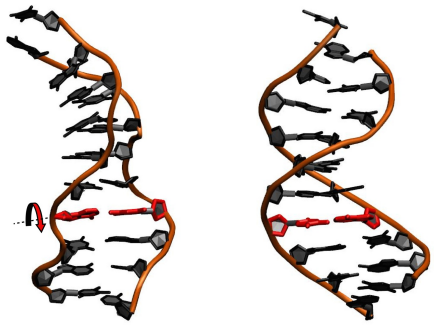
presence of counterions. The function f_{GB} is chosen to represent the distance between two point charges within the implicit solvent approximation:

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_i R_j\right)\right]^{1/2} \quad 2.4)$$

where r_{ij} is the real distance between atoms i and j. R_i and R_j are the 'effective Born radii' describing the depth that an atom is buried under the molecular surface. R_i is approximately equal to the Van der Waals radius when the atom is very close to the molecule surface, and is approximately equal to the distance from the atom to the closest surface area accessible by the solvent when the atom is buried deep into the solute molecule (98). At a very long distance $r_{ij} >> R_i$ and R_j, the second term in 2.4) vanishes and electrostatic interaction can be treated by the simple Coulombic force as f_{GB} becomes r_{ij}. At the limit as r_{ij} approaches zero, atoms i and j become sufficiently close so that $f_{GB} \approx R_i \approx R_j$.

2.2) Previous MD simulation studies of DNA

2.2.1) Forcefield Reparameterisation



a) Wrinkled DNA **b)** Canonical DNA

Figure 2.5 a) A wrinkled structure occurs as a consequence of an artefact due to the flipping of χ angle occurring within an atomistic MD simulation of a negatively supercoiled DNA using ff99BSC0 forcefield. The base pair labelled in red and the red arrow demonstrate the dramatic flipping of the furanose sugar associated with χ. **b)** The same negatively supercoiled DNA simulated using the modified forcefield parameters, showing the canonical right-handed helix conformation.

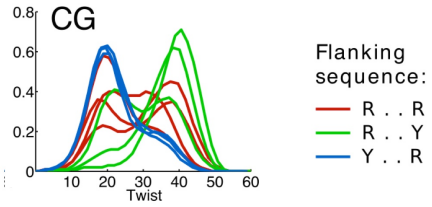
Molecular mechanics forcefield parameters are continually being developed to improve the representation of nucleic acids (99). Parameterisations have been done for α/γ angles to fix the structural artefacts due to backbone flipping (100), which introduced the ff99BSC0

forcefield as a modification from the original AMBER ff99 forcefield. Subsequently, modifications for the χ angle (12, 101) have been implemented to prevent artefacts from the ladder-like structures found in the simulations of folded RNA, which is known as the ff99BSC0$\chi_{OL3/4}$ forcefield (where version 3 is for RNA and version 4 for DNA). We have found that this correction reduced the occurrence of wrinkled structures (see Figure 2.5a) found in the simulations of torsionally stressed DNA (10), as the wrinkles became reversible or became the transient states prior to bubble (small single-stranded regions) formation. Another modification towards the improvement of ε and ζ angles (99) was still under development as the simulations in this thesis were performed. Thus, the ff99BSC0χ_{OL4} were used, providing the stable and artefact-free structure for over 100 ns of MD simulations.

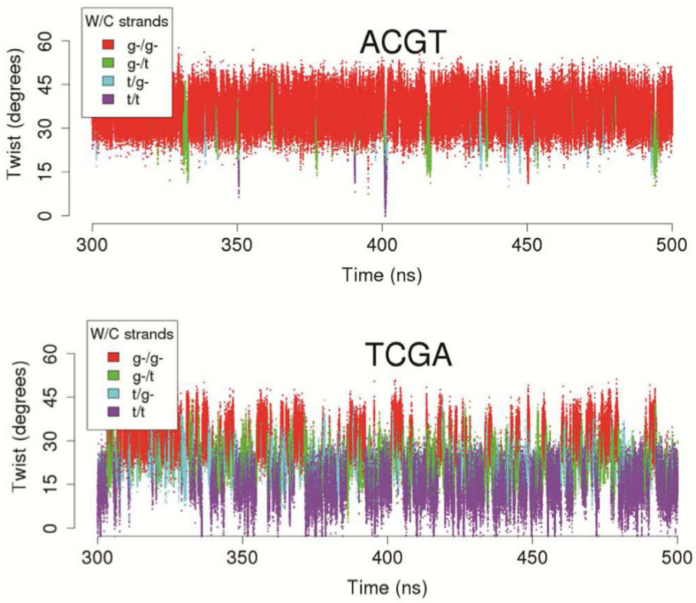
2.2.2) Simulation Studies of Sequence Dependence Flexibility

The ability of the ff99BSC0 molecular mechanics forcefield to maintain the canonical structure of DNA allows simulators to create a large ensemble of structural data for the study of DNA flexibility in aqueous solutions. Collaborative systematic simulations of linear DNA sequences containing each of the 136 unique tetra-nucleotides were performed by the ABC consortium using ff99BSC0 forcefield (100). A series of simulations were firstly run for 50-100 ns (102), and were recently expanded to 1 μs (103).

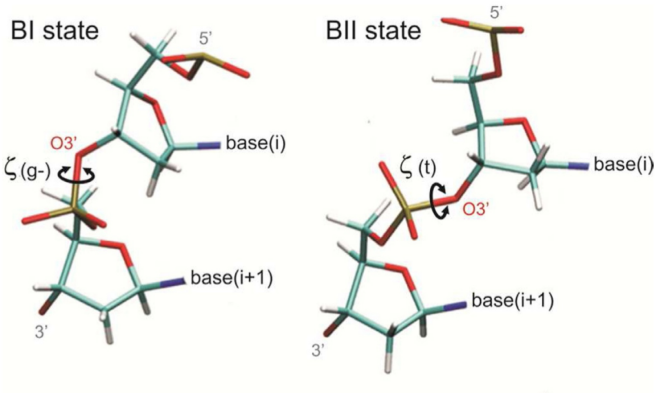
Analysis of the base pair step parameters showed the distinct mechanical properties of the purine-purine (RR), purine-pyrimidine (RY) and pyrimidine-purine (YR) dinucleotide families (102, 103), and enabled the quantitative characterisation of the harmonic spring constants associated with deformations away from the canonical B-helix ground state. However, there are limitations in this approach. For example, bimodality (and therefore anharmonicity) of the base pair step parameters, in particular CG steps, has been observed (104, 105). An example is shown in Figure 2.6. In Figure 2.6a, the probability distribution of local twist at CG dinucleotide steps is given for each unique flanking sequences, showing the nearest neighbour effect on twist bimodality. YCGR tetranucleotides correspond to low-twist CG steps (~20°), while RCGY tetranucleotides correspond to high-twist CG steps (~40°).



a) Twist Bimodality



b) Backbone Analysis



c) BI/BII Substate

Figure 2.6 a) Probability distribution of twist parameter observed in CG steps with all the ten combinations of flanking base pairs (purine-purine: red, purine-pyrimidine: green and pyrimidine-purine: blue), showing the twist bimodality at ~20° and ~40° (reproduced from (103)). **b)** Correlation between twist and g-(gauche-)/t(trans) modes of ζ dihedral angle observed in ACGT and TCGA tetramers. The CG step of ACGT favours the highly twisted *gauche-* mode for ζ dihedrals of both strands (g-/g-), while the CG step of TCGA favours the more untwisted *trans* mode (t/t). **c)** Backbone conformations of BI and BII substates, corresponding to the ζ(g-) and ζ(t) dihedral modes, respectively (reproduced from (105)).

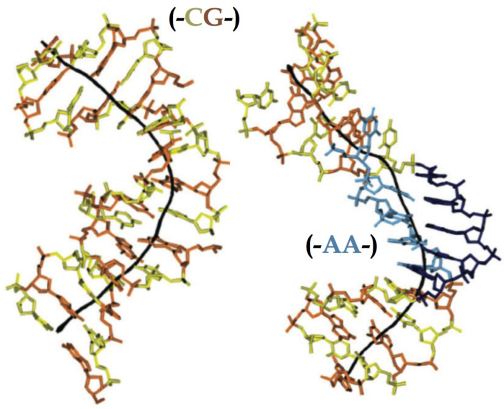
Further analysis of the backbone parameters unravelled the mechanism underlying CG twist bimodality (105). Figure 2.6b shows the correlation between twist bimodality and ζ backbone dihedral substates observed from ACGT (RCGY) and TCGA (YCGR) tetramers. The middle CG step within the ACGT tetranucleotide was mostly in the high-twist state corresponded to gauche- conformation and BI state (see Figure 2.6c (left)), while it was more probable to observe the low-twist state of the CG step within TCGA which corresponded to trans conformation and BII state (see Figure 2.6c (right)). This unique polymorphism of CG steps might also be important to some specific DNA functions, such as the switchable nucleosome arrangement (106).

2.2.3) Simulations of Bent and Twisted DNA

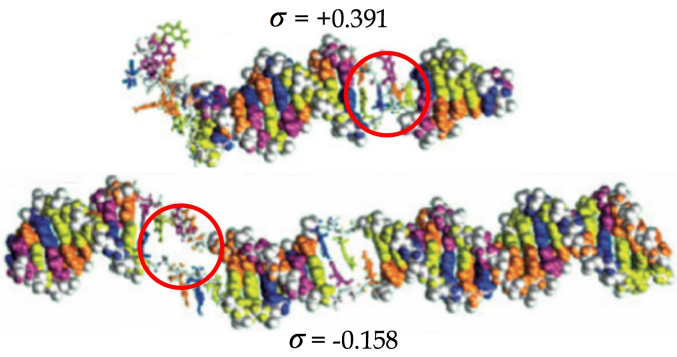
To mimic single molecule manipulation experiments, or to reduce the computational expense of simulating closed circular DNA sequences, simulators have applied artificial constraints to DNA molecules *in silico* to assess DNA deformability in response to bending and torsional stress. The effect of bending stress on DNA structure at the local (e.g. individual base step) and global (e.g. over 15 base pairs) levels was examined by umbrella sampling. In these simulations, the DNA was bent in a controlled manner by using external restraints applied during the calculations (107). The DNA was observed to escape from the quadratic regime when the global bending of an oligomer segment exceeded 50°. Local DNA kinking was observed at strong global bending (>100°). The simulation also showed that the poly-GC sequence preferred bending towards its major groove compared to its minor groove. The straightening effect of embedding a stiff A-tract sequence into the oligomer was also investigated, as shown in Figure 2.7a.

A series of MD simulations depicting the effects due to torsional stress has also been performed. In this study, the umbrella sampling technique was employed to control the DNA twist, showing that DNA deformation due to the applied torsional stress took the biomolecule out of the quadratic regime at smaller forces for negative compared to positive supercoiling (108). As DNA was overtwisted, canonical B-DNA underwent a phase transition to the P-form, in which the base spreading outwards and the phosphate moving to the middle (109). Most recently, MD simulations were performed in which the supercoiling in a linear molecule was

controlled by imposing periodic boundary conditions of a fixed length (74) so that DNA was not affected by bending and writhing. The degree of DNA twisting was varied by altering the number of base pairs. It was possible to observe local base pair flipping when DNA was untwisted, as well as the local transition from B-DNA to P-DNA when over-twisted (see Figure 2.7b). On mechanical failure of the DNA, the stress was observed to be non-uniformly distributed along the double helix, which formed locally denaturated regions allowing the remaining intact double helical regions to relax back to their canonical B-form.

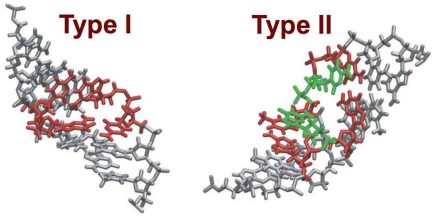


a) Bent DNA

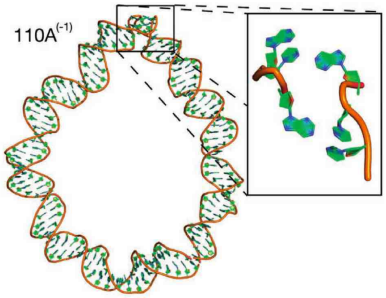


b) Under-/over- twisted DNA

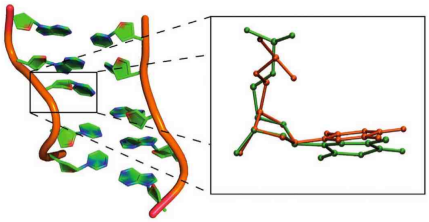
Figure 2.7 a) Conformations of a controlled-bent poly(CG) oligomer and a poly(CG) oligomer embedded with an A-tract (blue). Helical axis (black) demonstrates the straightness of the A-tract part and sharp bending at the A-tract/poly(CG) junctions (reproduced from (107)), **b)** The local transitions (in circles) of a highly overtwisted B-DNA structure at σ = +0.391 to a P-DNA structure and a highly undertwisted canonical DNA structure at σ = -0.158 into ssDNA bubbles (reproduced from (74)).



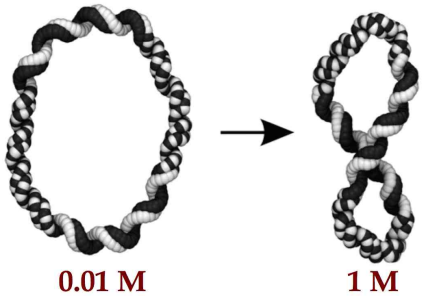
a) Type I and TypeII kinks



b) A Denaturation bubble



c) A Wrinkled Duplex



d) Salt dependent buckling

Figure 2.8 a) Type I and type II kinks observed in an atomistic MD simulation of a 94 bp DNA minicircle (reproduced from (80)). The intact base pair steps with broken base stacking are shown in red and the opened base pair at a type II kink is shown in green. **b)** A denaturation bubble observed in an atomistic MD simulation of a 110 bp DNA minicircle at ΔLk = -1, showing a ssDNA region with three opened base pairs (reproduced from (10)), **c)** A wrinkled structure found in an atomistic MD simulation of a 104 bp DNA minicircle at ΔLk = -0.5, showing an extreme helical unwinding and backbone distortion (reproduced from (10)). d) Buckling transition of a planar circular 178 bp DNA minicircle into a writhed minicircle observed when increasing salt concentration from 0.01 M to 1 M in a GB/SA implicit solvent (reproduced from (29)).

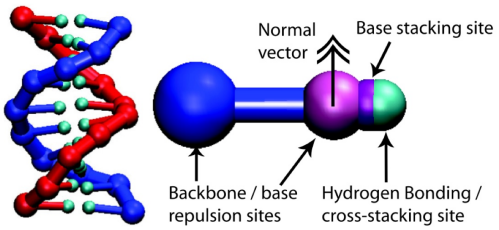
2.2.4) Previous Simulations of DNA Minicircles

In order to understand the mechanical properties of tightly coiled or looped DNA structures such as nucleosomes or lac repressor protein-DNA complexes, small DNA minicircles of a size 65 bp – 110 bp (shorter than the DNA persistence length: ~150 bp) have been used as experimental test systems for the study of extreme bending and torsional stress within DNA

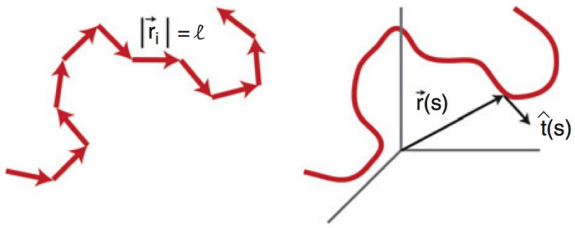
(11, 70, 110), as described in Chapter 1, section 1.6.3. Subsequently, these systems have been studied using atomistic MD simulations. The first simulations of DNA minicircles were performed for 94 base pair circles and were simulated using the AMBER parm94 forcefield (88). The simulations showed for the first time that bending stress within the minicircles induced kink formation. Two types of kink were identified (see Figure 2.8a). The type I kink involves sharp bending of two consecutive base pairs (labelled in red) into the minor groove with the broken stacking interaction with all intact hydrogen bonds. Type II kink involves three consecutive base pairs, in which the middle base pair (labelled in green) is broken.

With the subsequent modified forcefield ff99BSC0, negatively supercoiled 104 and 110 bp minicircles were also simulated and locally disrupted structures, such as kinks, denaturation bubbles and wrinkles were observed as a result of the twisting and bending stress within the closed DNA loops (10, 80) (see Figures 2.8b and 2.8c). Formation of type II kinks and denaturation bubbles involving broken base pairs was consistent with the experimental results, where single stranded DNA minicircles were digested by endonuclease enzymes (11). However, the formation of wrinkles (see Figure 2.8c) contradicted to the ssDNA digestions observed in the experiment suggested the occurrence of base pair disruptions. It was then hypothesised that the wrinkle formation was the artefact due to the unusual flip of the χ dihedrals. In Chapter 3, using the modified ff99BSC0χ_{OL4} forcefield on the same molecule, we showed that wrinkle was the transient state prior to the denaturation bubble formation.

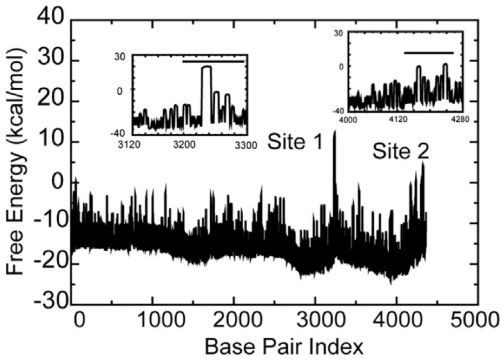
As for the larger DNA minicircles, those longer than the persistence length, DNA writhing could be observed. The 178 bp negatively supercoiled DNA minicircles were simulated in implicit solvent, in which the monovalent salt concentration could be adjusted as wished. Turning on the high salt switch from 0.01 M to 1 M NaCl induced the writhing of a planar circular DNA minicircle, as shown in Figure 2.8b (29).



a) oxDNA model



b) Worm-like chain model



c) Free energy from a thermodynamic model

Figure 2.9 a) Rigid base representation of the coarse-grained oxDNA model showing a point representing base and another point representing backbone for each residue (reproduced from (111)), **b)** Schematic diagram of the worm-like chain model: (left) flexible joints between discrete elements of the length *l* and (right) at continuum limit, a semi-elastic chain becomes an elastic rod (adapted from (18)), **c)** Free energy profile of an untwisted pBR332 DNA of 4 Mb calculated by a thermodynamic model from (112). Local maxima of the free energy profile shown at 'Site 1' and 'Site 2' suggested the locations in the genome, which are prone to denaturation under negative torsional stress. The prediction was consistent with Kowalski *et al* (113).

2.2.5) Coarse-Grained and Statistical Mechanical Models

Due to the computational expense of atomistic MD simulations of larger (>20 base pairs) DNA structures, these calculations are unable to capture the breaking and reforming disrupted base pairs in DNA minicircles (as will be described in Chapter 3). To obtain structural and dynamical data at longer timescales, DNA models can be constructed using a higher degree of approximation. Figure 2.9a shows the coarse-grained representation of the

'oxDNA' model, which considers each nucleotide as a rigid object consisting of 2 points representing both the base and backbone (instead of the 30-33 points representing all the atoms in a nucleotide residue) (114). Hydrogen bonding and base stacking interactions are parameterised from thermal melting data (115) so that the model can reproduce helical pitch, persistence length, torsional stiffness, and structural transitions observed from the single molecule experiments. Moreover, the model can provide the dynamic structural representations for DNA self-assembly (116), cruciforms (14), kissing loops (117), and plectoneme tips (75). In addition to the oxDNA model, which we used to compare with the atomistic MD in the Chapter 3, there are other coarse-grained models of DNA that employ complementary coarse-graining and parameterisation strategies (118–123).

At even coarser levels of representation, the behaviour of DNA under mechanical stress has been described using the worm-like chain (WLC). This model treats the DNA as a semi-flexible polymer, which is flexible only at the joint between the discrete segments (see Figure 2.9b). Analytical application of this model provides good agreement with the data from force-extension experiments (124, 125) and enables the shape of closed circular DNA loops to be deduced by numerical minimisation techniques (126). While most WLC models assume that the duplex structure remains intact, analytical solutions of a modified elastic rod model with the addition of an energetic term for structural disruptions has been used to depict the phase behaviour of DNA minicircles at different sizes and superhelical densities (127), and which provided results in good agreement with a series of MD simulations, which provided the atomistic details (28). More sophisticated numerical calculations using Monte Carlo simulations in which the WLC model is applied has also been used to evaluate the free energy of the kink formation in DNA minicircles (128) and has provided the free energy associated with the formation of DNA knots (129).

At the genomic scale, the stress induced duplex destabilisation (SIDD) statistical mechanical analysis is capable of predicting the location where the DNA double helix becomes destabilised under superhelical stress, which has been hypothesised to correspond to potential genomic regulatory regions (130). The SIDD computational method used the thermodynamic database from Santa Lucia *et al*. (115) to calculate the free energy and probability profile of duplex destabilisation. Another analytical thermodynamic model to predict the local DNA defects takes into account the two body interactions between the neighbouring base pairs and the long range negative correlation

due to topological constraints (112). The model predicted the locations of DNA melting in an untwisted pBR332 DNA (see Figure 2.9c), which was in a general agreement with a biochemical study by Kowalski *et al* (113).

One of the missing theoretical aspect for molecular biology is the ability to describe the molecular functions under multiple environments and lengthscales. The multi-scale modelling allows simplification of the structural representation, while still preserve the fine details of the model at the highest resolution. In Chapter 3, computational models at three different lengthscales (atomistic, coarse-grained and genomic) have been used to represent the small DNA minicircles, which brings about the ideas of reconciling the three models to improve the efficiency and the accuracy in solving the larger systems of DNA under superhelical stress.

Chapter 3
Mechanical Response of Highly Stressed DNA

3.1) Introduction

In order to obtain the structural information for disrupted DNA minicircles under extreme torsional and bending stress due to the cyclisation of short DNA fragments, atomistic MD simulations were performed by Lankas *et al.* They observed two types of kinked DNA structure within the 94 bp minicircles (80), thereby providing insight into the experimental results of Cloutier and Widom (70). The 'type I' kink describes a local disruption, where stacking interaction between two neighbouring base pairs was missing, and the disrupted site became a flexible hinge while the hydrogen bonds were still intact. The other kind of kink, named 'type II' consists of one base pair broken flanked by two highly bent base pair steps (further discussed in section 3.3.1 and figure 3.8). However, errors due to backbone α and γ dihedrals parameterization subsequently became apparent for the AMBER ff99 forcefield used in this study. This structural artefact enhanced the instability of DNA double helix and overestimated the probability of kink formation. Following the introduction of the modified AMBER ff99 forcefield parameters BSC0 (100), the α/γ artefacts were corrected, and these kinks were not seen in subsequent simulation studies of short DNA fragments at the same degree of bending (107).

Subsequently, Mitchell *et al.* (10) performed another series of MD simulations of 65, 104 and 110 bp minicircles, reproduced from later DNA cyclisation experiments by Du *et al.* (11). Four base pairs were added into the 100 and 106 bp minicircles due to the underestimated helical twist in AMBER ff99BSC0. The simulation results generally agreed with the experiment, as kinks and denaturation bubbles (containing two or more broken base pairs) were observed in sequences susceptible to digestion by endonucleases (10, 11). However, for the 104 bp minicircles at ΔLk = -0.5, a wrinkled structure was observed. This irreversible wrinkled structure possessed extremely untwisted base pairs with hydrogen bonds remaining intact. However, in the experiment, the ΔLk = -0.5 minicircles underwent the digestion by both BAL-31 and S1 endonucleases (11), suggesting the presence of single stranded DNA. Moreover, the overly distorted χ dihedral angles observed in the wrinkled structures were analogous to the common

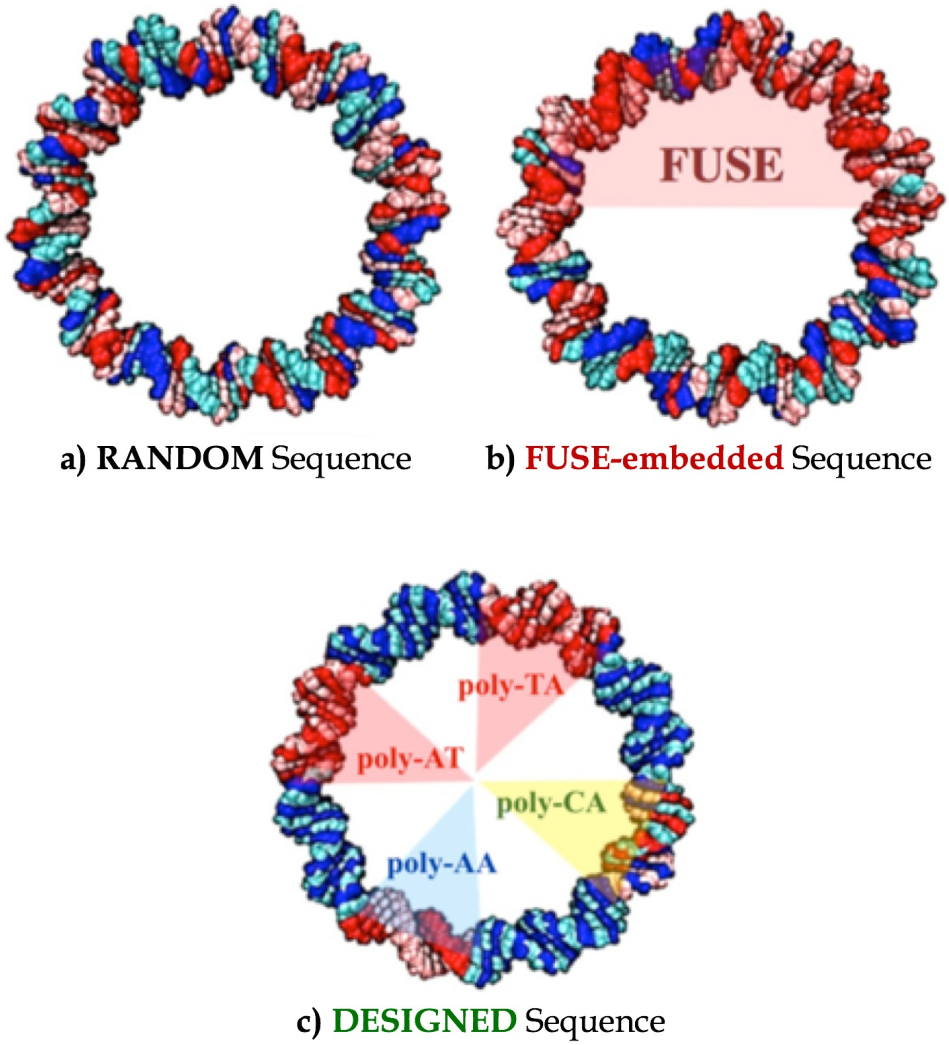
'ladder-like' artefact found in RNA simulations (101). It was later concluded that the irreversible wrinkled structure was an artefact due to the badly parameterized χ dihedral profile, which was later fixed (12), and wrinkles became only a transient state prior to the formation of denaturation bubbles. The ff99BSC0 parameters with χ modification (ff99BSC0$_{\chi OL4}$) has therefore been used in all the MD simulations within this thesis.

The extension of the simulation study by Mitchell *et al.* (10) using the modified forcefield, and including different DNA sequences are shown in this chapter. The results were compared to the other two computational methods, coarse-grained simulations of DNA and a statistical mechanical model of DNA, which demonstrated common features across all three simulation regimes as well as the importance of structural features available only in atomistic MD simulations.

3.2) Methods

3.2.1) DNA Sequences

The Du *et al.* minicircle DNA sequence used in this study consisted of almost no major structural motif, such as a consecutive run of AA base pair steps (A-tract), and was referred to as 'RANDOM'. In order to assess the sequence dependence of the DNA mechanical response, we embedded the A-T rich, stress-sensitive 50 bp 'FUSE' element into the RANDOM sequence. The biological importance of FUSE element was mentioned in Chapter 1, section 1.5.3. Also, we created a rationally 'DESIGNED' sequence, containing repeated sequence motifs to understand further the sequence-dependence properties of DNA. Figure 3.1 shows the starting structures of torsionally relaxed DNA in the three sequences mentioned above, colour coded in accordance with the base sequences. In Figure 3.1b, the A-T rich FUSE element was highlighted. In Figure 3.1c, the 12-bp sequence motif regions were highlighted: poly-TA and poly-AT in red, poly-AA in blue, and poly-CA in yellow. Full details of the DNA sequences investigated in this chapter are provided in Appendix A.



a) RANDOM Sequence **b) FUSE-embedded** Sequence

c) DESIGNED Sequence

Figure 3.1 DNA minicircle sequences mentioned in this chapter. **a)** The RANDOM sequence has equal A-T and G-C content and has no repetitive sequences (e.g. ATATAT…, GCGCGC…, CACACA… or etc.). This sequence is originally obtained from an experimental work of Du *et al.* observing structural disruptions within supercoiled DNA minicircles (11). **b)** FUSE-embedded sequence is a modified RANDOM sequence, in which a 50 bp, A-T rich FUSE element (46) has been embedded. **c)** Artificially DESIGNED sequence containing repetitive elements: poly-TA (red), poly-AT (red), poly-AA (blue), poly-CA (yellow) and poly-CG (unshaded). DNA is colour-coded by base types: red for adenine, pink for thymine, blue for guanine and cyan for cytosine.

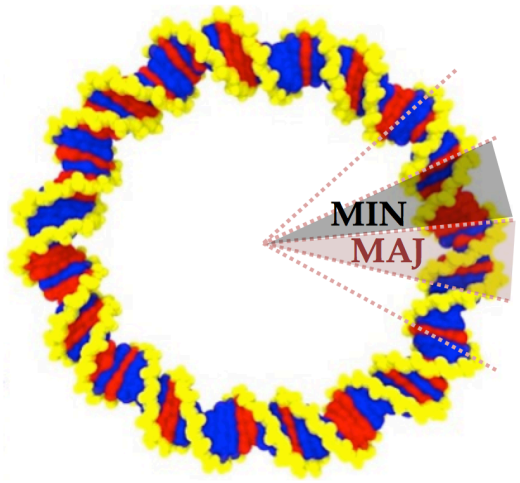


Figure 3.2 Diagram defining a major groove block region (**MAJ**), where the DNA is bent so that the major groove pointing inside the bent DNA curve, and a minor groove block region (**MIN**), where DNA is bent so that the minor groove pointing inside the bent DNA curve.

3.2.2) Bending Register Angles

The DNA helix is anisotropic. Bending DNA towards the major groove and towards the minor groove are not equivalent. This was shown in crystallographic data of nucleosomes, where unique dinucleotide steps favour different bending directions (49). The terms "major groove block" and "minor groove block" have been used to describe the locations where the DNA is bent towards the major groove and the minor groove, respectively (see Figure 3.2). The DESIGNED sequence, which consists of short repetitive sequence motif elements, is more sensitive to bending register than the other two sequences. In order to examine this sequence-dependent bending anisotropy, the simulations of the DESIGNED sequence at $\Delta Lk = -1$ were started at three different bending register angles, as shown in Figure 3.3. In Figure 3.3a, the poly-AA region is positioned in-phase with the major groove block, so this bending register is called the "AA-Major". The other two starting structures, having their helical bending phases 5 bp and 10 bp shifted from the first bending register, were referred in Figures 3.3b and 3.3c as "TA-Major", where poly-TA is located in the major groove block region, and "CA-Minor", where poly-CA is located in the minor groove block region, respectively.

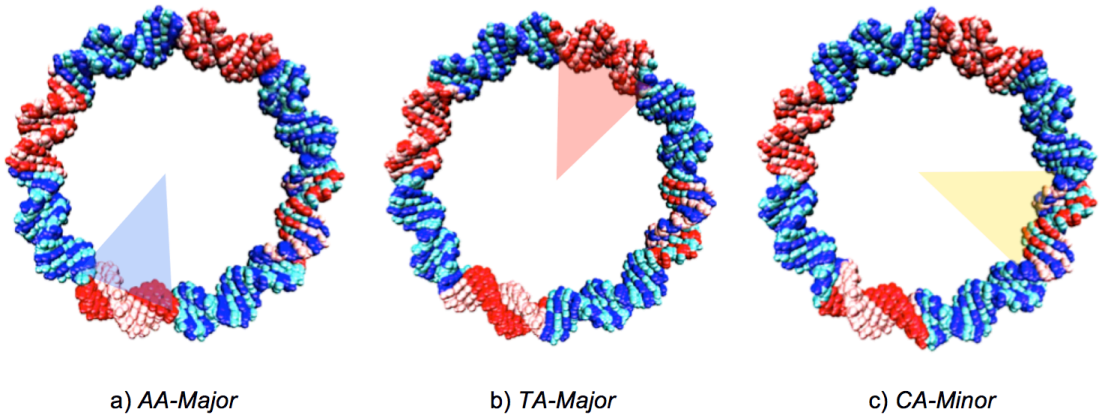


Figure 3.3 Three different bending register angles at the start of ΔLk = -1 MD simulations of the DESIGNED sequence 108 bp minicircles. The bending register angles were named according to the sequence motif regions and their coherence with the major and minor grooves bending phase, **a)** AA-Major: poly-AA in-phase with a major groove block, **b)** TA-Major: poly-TA in-phased with a major groove block, and **c)** CA-Minor: poly-CA in-phased with a minor groove block.

3.2.3) Explicit Solvent Simulations

In order to build a starting structure of circular DNA, an NAB (Nucleic Acid Builder) script was written. For each base pair, the standard coordinates (91) were provided by the AMBERTOOLS package (93). A base pair was firstly added onto the circumference, and was oriented perpendicular to the tangent. The newly added coordinates then underwent a rotation about the vertical axis (twist), followed by a step-wise rotation of the whole system about the circle origin. Repeating the same procedure can create a perfectly planar circular DNA, where the whole double helix is uniformly bent.

Four-stage minimisations, eight-stage equilibrations, and MD production runs were performed in the explicit solvent in the 0.1 M NaCl. The details of the minimisation and equilibration stages are provided in Appendix B. A topology file containing the ff99BSC0$_{\chi OL4}$ forcefield information and a starting coordinate file were generated by the XLEAP module implemented in AMBERTOOLS. The starting structure was solvated in a truncated octahedron TIP3P waterbox with a 30 Å solution buffer and periodic boundary condition (this took approximately 140,000 water molecules). Then, a number of Na+ counterions were added to neutralise the negatively charged DNA backbone (216 Na+ were added for a 108 bp

minicircle) and excess Na+ and Cl- ions were added to create a 0.1 M NaCl salt environment (about 250 Na+ and 250 Cl- were added).

At the first minimisation stage, all-atom position restraints were applied before the restraint weights were decreased in the second and the third minimisation stages. All the atoms were free from restraints in the fourth stage. The minimised structure was then equilibrated. In the first equilibration stage, DNA was frozen and the system temperature was heated up from 100 K to 300 K. Then, in the following stages, the restraint weights on DNA were gradually decreased (see Appendix B). After that, restraint-free productive MD runs were performed under the temperature of 300 K, which was kept constant by the Berendsen weak coupling algorithm (96). The pressure was also kept constant at 1 atm by the NPT ensemble, in which the volume of the system box was rescaled to regulate the pressure. The SHAKE algorithm (95) was implemented to freeze the covalent bonds involving hydrogen atoms, allowing the leapfrog integration (mentioned in Chapter 2, Figure 2.4) to be performed every 2 fs. The trajectory was written every 500 MD steps or 1 ps. Simulations were performed with GROMACS using 32 MPI parallel processors of the University of Leeds Advance Research Computing (ARC1 and ARC2) supercomputer clusters, able to perform 1.25 ns/day.

3.2.4) Series of MD Simulations

For each of the three DNA sequences investigated in this chapter (the RANDOM, FUSE-embedded and DESIGNED sequences), 50 ns atomistic MD simulations of torsionally relaxed minicircles were performed at three different bending register angles. At the $\Delta Lk = -0.5$ negatively supercoiled state, three replicas of 50 ns simulations were performed for each bending register angle in order to improve the sampling. Breathing, or short-lived, reversible base pair disruption occurred within the atomistic MD simulations at $\Delta Lk = -0.5$ will be examined in the Section 3.3.1.

For RANDOM and FUSE-embedded minicircles at $\Delta Lk = -1$, where the irreversible structural defects were observed, the simulations were extended until their defects were established, which in certain cases required trajectories in excess of 50 ns (see Table 3.1). Time and location of defect formation, and the effect due to bending co-operativity will be discussed in the Section 3.3.2, where the statistical mechanical model, SIDD, and the coarse-grained simulation, oxDNA, were performed for comparison.

For the DESIGNED minicircles at ΔLk = -1, in simulations where defect formation was observed after relatively short timescales compared to RANDOM and FUSE-embedded sequences, the simulation lengths were reduced to 40 ns and 15 replicas were performed for each bending register in order to examine the anisotropy effect, as will be discussed in the Section 3.3.3.

In summary, three simulations (containing 150 ns MD trajectory) were performed for torsionally relaxed DNA, while nine simulations (containing 450 ns MD trajectory) were performed for the ΔLk = -0.5 minicircles. The series of RANDOM and FUSE-embedded minicircles simulations consisted of 18 simulations containing at least 900 ns, while 45 simulations of DESIGNED minicircles at ΔLk = -1 contained about 1800 ns. In total, more than 3.3 μs MD trajectories were produced to examine the mechanical response of DNA under high bending and superhelical stress.

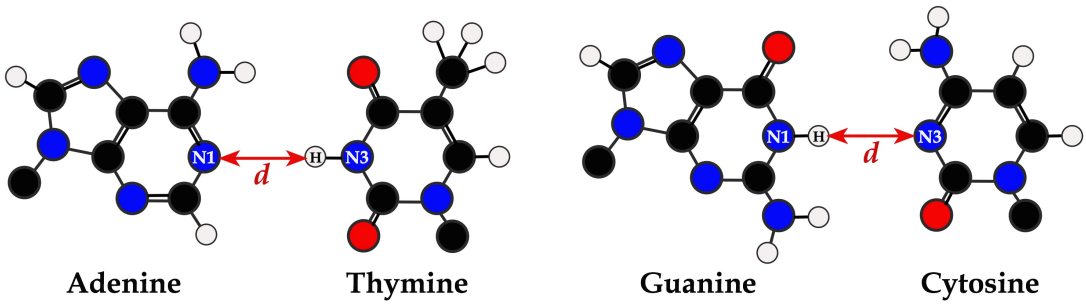


Figure 3.4 Measurement of the hydrogen bond distance (***d***) for an A-T and a G-C base pairs. Atomic nomenclature here is used for inputting the PTRAJ script to analyse the time series of hydrogen bonding distance between the donor hydrogen atom (H3 for A-T and H1 for G-C) and the acceptor nitrogen atom (N1 for A-T and N3 for G-C).

3.2.5) MD Trajectory Analysis

Hydrogen bond distances were measured between the N1 atom of adenine and the H3 atom of thymine for an A-T base pair, and between the N3 atom of cytosine and H1 atom of guanine for a G-C base pair (see Figure 3.4) using the PTRAJ module implemented in AMBERTOOLS (93). A base pair was classified as disrupted when the distance *d* measured exceeded 4 Å. A short (< 1 ns), reversible disruption (see Section 3.3.1) was termed as a 'breathing' event, while a long (> 10 ns), irreversible disruption (see Section 3.3.2 - 3.3.4) was termed as a 'defect'. Types of defects were classified by visual inspection of the trajectories by using the VMD program, in which

conformations of the base pairs with disrupted hydrogen bonds and base pair stacking can be rendered and used as a validation for the PTRAJ h-bond distance analysis.

Base pair step parameters and helical bending data were obtained from the CANAL module implemented in the CURVES+ package (131, 132). All the base pair step parameters were obtained as a time series for the MD trajectories. These were processed along with the sequence information using python scripts employing the NUMPY library, which is suitable for working with the large array output from CANAL. Parameters roll, twist and slide were analysed to address the mechanical response of DNA to superhelical stress (see Section 3.3.1).

The elastic energy stored in the minicircles was calculated using the stiffness matrix method. For each base pair step, stiffness matrix can be obtained by inverting the correlation matrix of the six base pair step parameters (σ):

$$A = k_B T \sigma^{-1} = \begin{bmatrix} a_{twist} & a_{t-r} & a_{t-l} & a_{t-s} & a_{t-s} & a_{t-d} \\ a_{t-r} & a_{roll} & a_{r-l} & a_{r-i} & a_{r-s} & a_{r-d} \\ a_{t-l} & a_{r-l} & a_{tilt} & a_{l-i} & a_{l-s} & a_{l-d} \\ a_{t-i} & a_{r-i} & a_{l-i} & a_{rise} & a_{i-s} & a_{i-d} \\ a_{t-s} & a_{r-s} & a_{l-s} & a_{i-s} & a_{shift} & a_{s-d} \\ a_{t-d} & a_{r-d} & a_{l-d} & a_{i-d} & a_{s-d} & a_{slide} \end{bmatrix} \quad 3.1)$$

The diagonal elements are the force constants of the base pair steps in all six degrees of freedom and the non-diagonal elements describe the coupling between each pair of parameters. The elastic energy of each base pair step can then be evaluated by

$$E = \frac{1}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} a_{ij} \, \Delta p_i \Delta p_j \, , \quad 3.2)$$

when a_{ij} is an element in the stiffness matrix and Δp_i is the deviation of base pair step parameter *i* from the average value.

3.3) Results

3.3.1) Breathing in ΔLk = -0.5 Minicircles

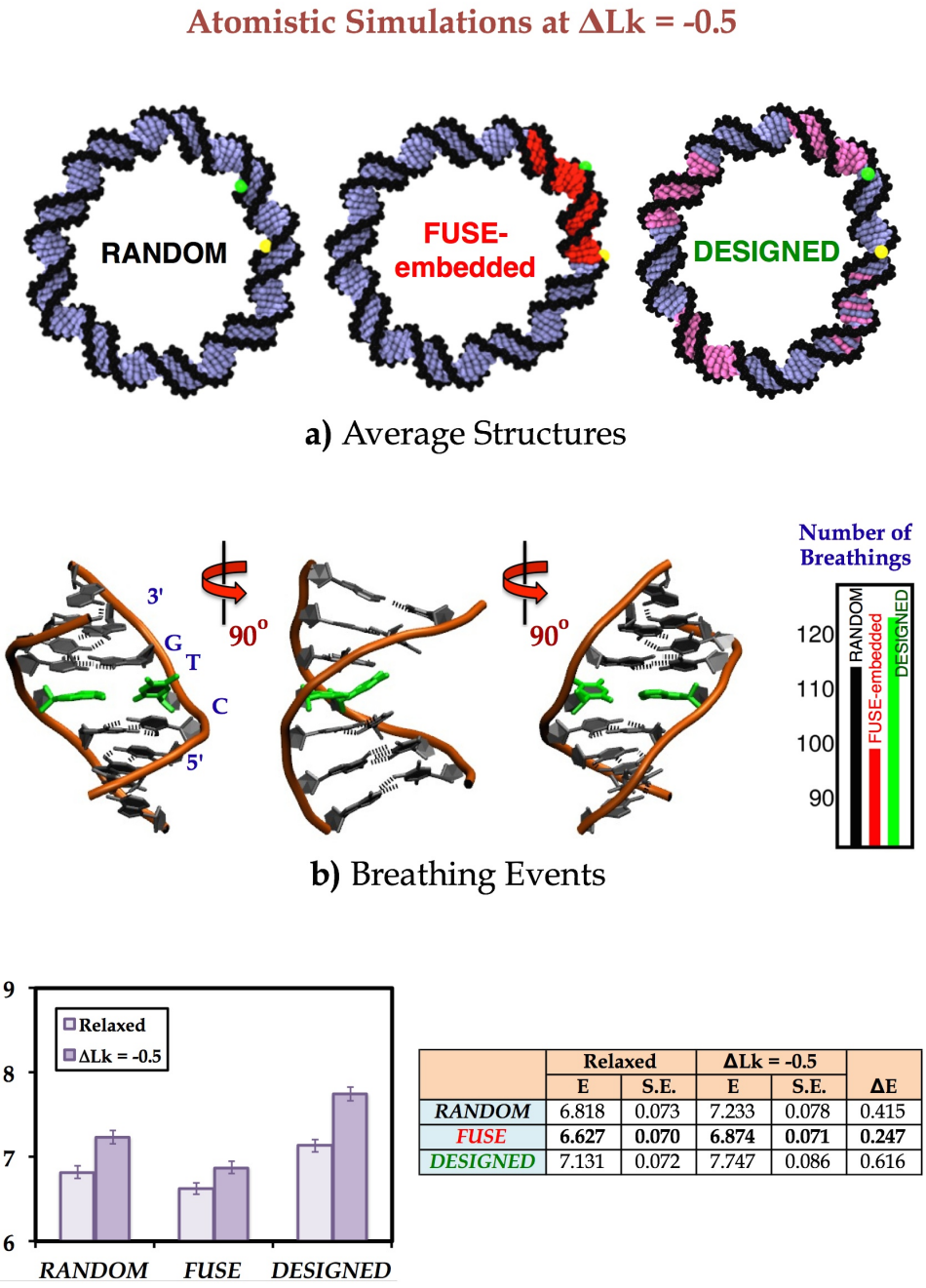
At ΔLk = -0.5, superhelical stress within the 102 bp minicircles was sufficient to induce a number of short-lived (< 1ns), reversible breathing events. For each breathing event, a pyrimidine base (mostly thymine) was flipped out from the DNA strand into the major groove. Hydrogen bonding analysis was performed as described in Section 3.2.5 to count the number of

breathing events from the last 40 ns of simulations. In total, 360 ns trajectories of RANDOM, FUSE-embedded and DESIGNED minicircles at ΔLk = -0.5 were analysed. Then, the time average of elastic energy and basepair step parameters (twist, roll and slide) over the 360 ns trajectory were calculated for each of the three DNA sequences.

Reduced Breathing in A-T rich Minicircles Corroborated Through Elastic Energy Analysis

Figure 3.5a shows the representative structures from the MD simulations of the ΔLk = -0.5 RANDOM, FUSE-embedded and DESIGNED minicircles, in which DNA was untwisted and bent but no permanent structural defects (lasting for longer than 10 ns) were observed. However, the superhelical stress was sufficient to induce transient breathing events. An example of a breathing event at the sequence CAG/CTG is shown in Figure 3.5b (left). A breathing event is counted when the h-bond distance of a base pair exceeds 4 Å and then back to its original state. Figure 3.5b (right) shows that, at ΔLk = -0.5, breathing events occurred most frequently in the DESIGNED minicircles, but surprisingly the simulations of A-T rich FUSE-embedded minicircles contain the least number of breathing events.

This counter-intuitive result corresponded with the analysis of elastic energy stored within the minicircles. Figure 3.5c shows the average values and standard errors of the elastic energy per base pair step calculated by the stiffness matrix method on the torsionally relaxed and ΔLk = -0.5 RANDOM, FUSE-embedded and DESIGNED minicircles. The energy calculation showed that the DESIGNED and FUSE-embedded minicircles had the highest and the lowest elastic energy (E) per base pair, respectively. The elastic energy stored in the FUSE-embedded minicircles was the least affected by negative supercoiling, as the difference in elastic energy (ΔE) between the torsionally relaxed and ΔLk = -0.5 minicircles was the lowest. This is a result of the sequence content of the FUSE element, which contains a large number of flexible pyrimidine-purine (YR) base pair steps. The observation that the FUSE-element is the least energetically stressed of the three sequences explains why the hydrogen bonding analysis identified fewer breathing events for these minicircles than observed for the RANDOM and DESIGNED sequences.



	Relaxed		ΔLk = -0.5		
	E	S.E.	E	S.E.	ΔE
RANDOM	6.818	0.073	7.233	0.078	0.415
FUSE	**6.627**	**0.070**	**6.874**	**0.071**	**0.247**
DESIGNED	7.131	0.072	7.747	0.086	0.616

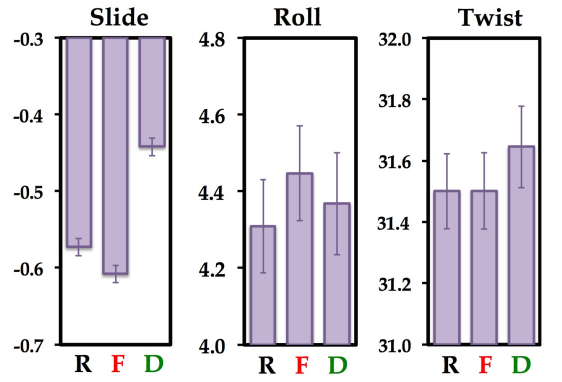
Figure 3.5 a) Representative MD structures of the ΔLk = -0.5 RANDOM, FUSE-embedded and DESIGNED minicircles. Yellow and green dots indicate the residues number 1 and 10. The red region (bp 1-22) within FUSE-embedded minicircle contains high number of the flexible TA steps. The pink regions within the DESIGNED minicircle contain the repeating sequence motifs as described in Figure 3.1. **b)** A sample breathing event at a CAG/CTG tri-nucleotides showing the thymine base flipping into the major groove, and number of breathing events observed within 360 ns MD simulations for each minicircle sequence, **c)** Time averages and standard errors of elastic energy per base pair step calculated from stiffness matrices for RANDOM, FUSE-embedded and DESIGNED minicircles at ΔLk = 0 and -0.5. ΔE values indicate the difference in elastic energy stored within torsionally relaxed and ΔLk = -0.5 minicircles.

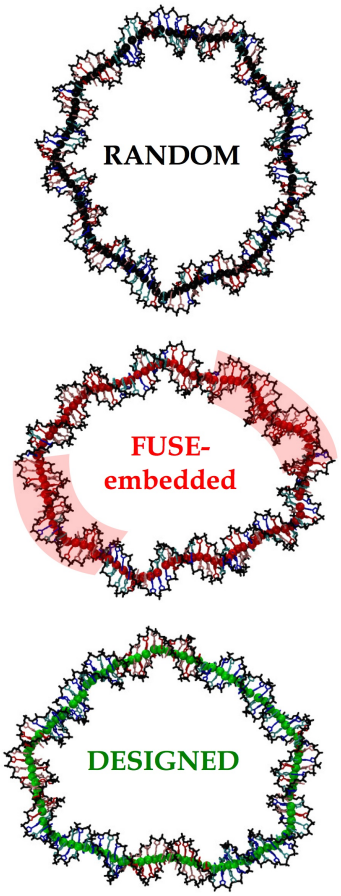
Breathing Suppressed by Minor Groove Opening and Cooperative Bending

A conformational analysis was performed to investigate the ability of ΔLk = -0.5 minicircles to accommodate bending and superhelical stress and to explain the reduced probability of breathing events in the A-T rich FUSE minicircle compared to the RANDOM and DESIGNED sequences. The slide, roll and twist parameters were extracted from the MD trajectories using CURVES+. Figure 3.6a shows the time-averages and standard errors of slide, roll and twist over the last 40 ns of all the MD trajectories of RANDOM, FUSE-embedded and DESIGNED minicircles at ΔLk = -0.5 (average over 360 ns trajectory of each DNA sequence). Base pair steps within RANDOM and FUSE-embedded minicircles were generally more untwisted. Moreover, RANDOM and FUSE-embedded minicircles also had their minor grooves more opened (positive roll) than within the DESIGNED minicircles. Correspondingly, the higher values of negative slide observed in the RANDOM and FUSE-embedded minicircles shifted the base pair origins from the centre and widened the DNA helix (see Figure 3.6b).

The consequence of embedding the FUSE element into the RANDOM minicircles is that the FUSE-embedded minicircles have two opposing flexible regions with a high content of YR steps (highlighted in red), which can bend cooperatively due to the closed-loop topology. DNA segments containing many bi-stable YR steps (19) (see the discussion in Chapter 1 Section 1.3.3) are able to be highly bent into the major groove (positive roll) without completely losing their base stacking interactions. However, when sufficiently large superhelical stress is introduced, these regions containing many flexible YR steps become sensitive and can be disrupted as occurred when the ΔLk was -1, as will be discussed in the next section.

	Slide		Roll		Twist	
	P	S.E.	P	S.E.	P	S.E.
RANDOM	-0.573	0.011	4.309	0.122	31.500	0.089
FUSE	-0.608	0.010	4.448	0.124	31.501	0.084
DESIGNED	-0.442	0.010	4.368	0.133	31.645	0.093





a) b)

Figure 3.6 a) Time-average values and standard errors of slide, roll and twist calculated from the last 40 ns of all the atomistic MD simulations of RANDOM, FUSE-embedded and DESIGNED minicircles at ΔLk = -0.5, **b)** A snapshot picked from the MD trajectory of each sequence with all the local base pair origins defined by X3DNA, showing the toroidal paths for the RANDOM and FUSE-embedded minicircles as a consequence of highly negative slide. Two highly flexible regions of the FUSE-embedded minicircle, containing large numbers of TA and CA steps, are highlighted in pink and located at the bent apices as a consequence of positive roll.

3.3.2) Sequence Dependent Defect Formation of RANDOM and FUSE-embedded Minicircles at ΔLk = -1

When subjected to larger superhelical stress at ΔLk = -1, long-lived denaturations (>10 ns) or defects were observed within the atomistic MD simulations of 108 bp minicircles. Sequence dependence of defect formation of RANDOM and FUSE-embedded sequences was compared to the stress-induced duplex destabilisation model, SIDD, and oxDNA coarse-grained model performed by Christian Matek (111). While the SIDD model and oxDNA coarse-grained simulations could provide the equilibrium

probability of defect formation, visual inspection of atomistic MD simulations suggested that defects were irreversible and non-equilibrium due to limited simulation timescale. The sequence dependence of defect formation in atomistic simulations can only be assessed by repeating the simulation with different initial random velocities and register angles as described in Section 3.2.4. Hydrogen bonding analysis was carried out to specify time and location of defect formation for each of the ΔLk = -1 RANDOM and FUSE-embedded simulations.

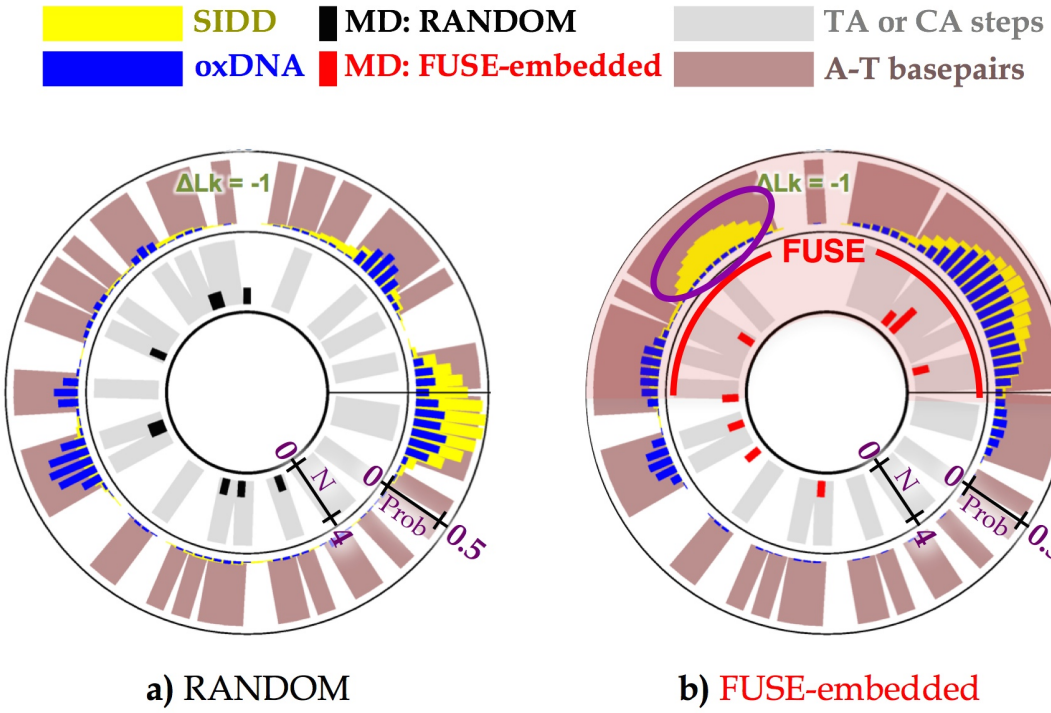


Figure 3.7 (***Outer***) Sequence-dependent denaturation probability (Prob) observed in DNA minicircles at ΔLk = -1 of RANDOM (left) and FUSE-embedded (right) sequences by SIDD model (yellow) and oxDNA simulations (blue). A-T base pairs are shaded in brown. (***Inner***) locations of all defects (N) observed in nine atomistic MD simulations of each minicircle sequence at the flexible TA and CA steps, shaded in grey.

Higher Resolution Models Corroborates the higher Importance of Base Stacking Interactions

Figure 3.7 (outer circles) compares the sequence dependence of base pair opening for the RANDOM and FUSE-embedded minicircles. The opening probability provided by the SIDD and the coarse-grained oxDNA models at ΔLk = -1 are shown in the outer polar plots. Both models suggested that the local maxima of opening probability was found at the centre of A-T rich regions (shaded in brown), due to the lower energy

penalty taken for these base pairs to open compared to GC rich DNA. The difference between these two models arises due to the improved description of the stacking interaction terms within the oxDNA model, which correctly shows a greater probability of the defect formation at the bi-stable YR steps, especially TA and CA (shaded in grey), and the lower probability at RR steps. At the A-T rich regions containing A-tracts (circled in purple), the base pair opening probability predicted by oxDNA simulations was significantly lower that that predicted by SIDD model.

Figure 3.7 (inner circles) shows the positions of the defects that occurred in the MD simulations of RANDOM and FUSE-embedded minicircles at ΔLk = -1. MD simulations predicted that defects usually occurred at YR steps due to their weak base stacking interactions (see Table 3.1). As the timescale of the atomistic MD simulations are too short to provide converged probabilities, the sequence dependence of defect formation can only be described qualitatively from these calculations. However, SIDD and oxDNA do not suffer from this limitation, and it is possible to observe two cooperative denaturation events at the opposing DNA sites at a longer timescale.

Bending Co-operativity Affects the Defect Formation in ΔLk = -1 FUSE-embedded Minicircles

In the atomistic MD simulations at ΔLk = -1, the high level of superhelical stress induced various forms of irreversible base pair disruptions (see Figure 3.8) within ~100 ns. Starting from planar circular conformations, minicircles at ΔLk = -1 developed the different categories of defects due to the bending and torsional stress, as shown in Figure 3.9 in comparison with the global structures obtained from oxDNA simulations. Under-winding due to torsional stress induced the formation of denaturation bubbles (see Figure 3.9a), while extreme bending stress induced base pair step kinking (see Figure 3.9b). Cooperative bending between two opposing regions resulted either in DNA writhing (see Figure 3.9c; not seen in oxDNA) or cooperative kinking (see Figure 3.9d).

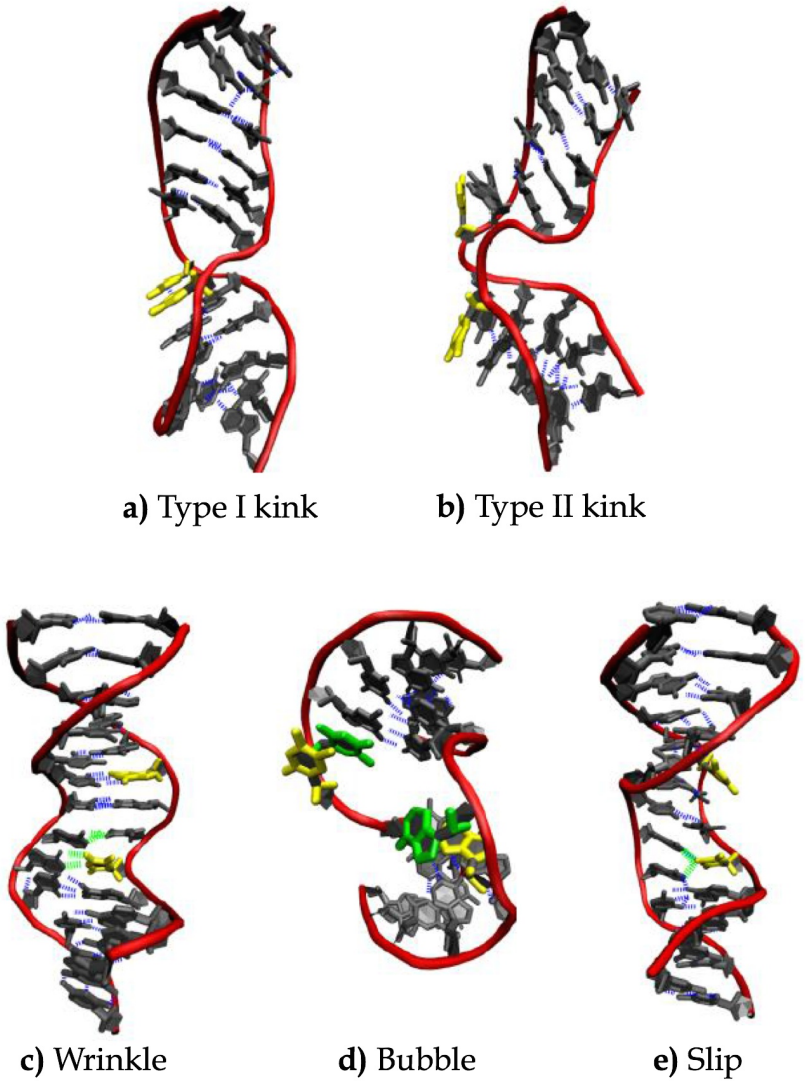


Figure 3.8 Types of the local structural defects observed at ΔLk = -1: **a)** Type I kink – a flexible hinge with broken base stacking and intact base pairs, **b)** Type II kink - a very flexible hinge with broken base stacking and a base pair opening, **c)** Wrinkle – untwisted base pair steps with distorted backbones and intact base pairs, **d)** Bubble – single stranded DNA regions with more than one base pairs opening, and **e)** Slip – base mis-pairing with two unpaired bases.

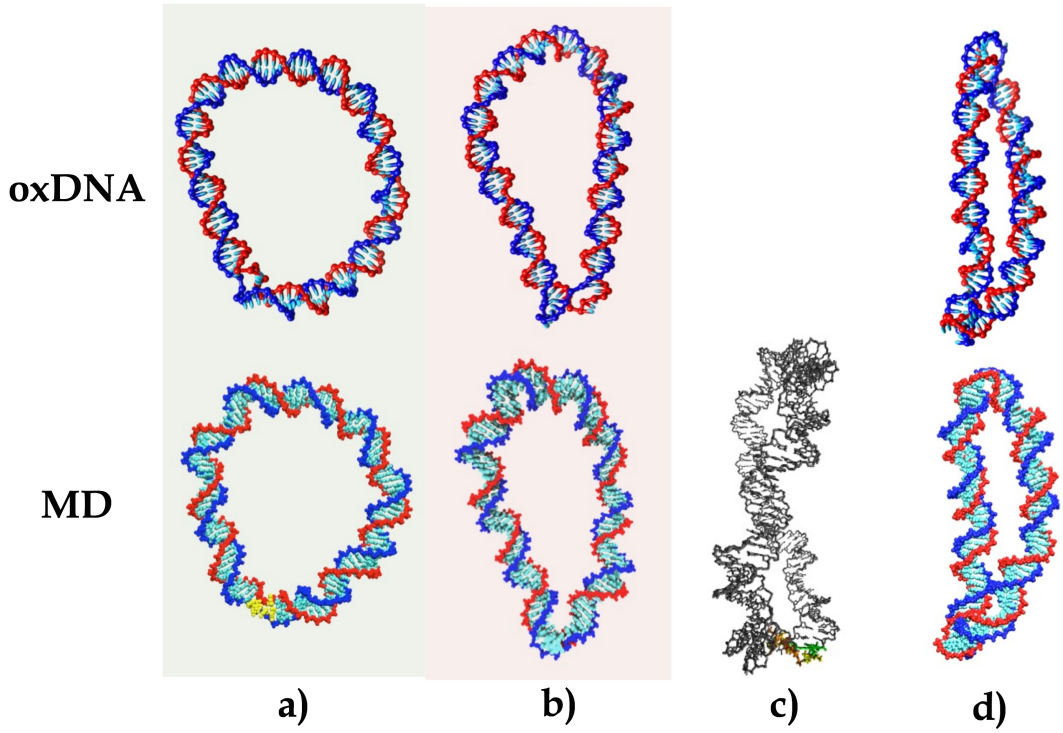


Figure 3.9 OxDNA and MD trajectory snapshots illustrating the possible global shapes of the disrupted minicircles at ΔLk = -1 by **a)** a defect induced by twisting observed in a RANDOM (**VB1** for atomistic MD; yellow) minicircle, **b)** a defect induced by bending observed in a FUSE-embedded (**FB1** for atomistic MD) minicircle, **c)** writhing observed in a FUSE-embedded (**FA1** for atomistic MD) minicircle and **d)** cooperative kinking observed in a RANDOM (**RA0** for atomistic MD) minicircle. oxDNA simulations were performed by Christian Matek.

Name	Denaturations	Time	Sequence	Location
RA1	Untwisting	23	CAAT/ATTG	32-33
RB0	Untwisting	17	CAGG/CCTG	87-88
RB1	Untwisting	10	TTAC/GTAA	61-62
RC1	Untwisting	10	TCAA/TTGA	77-78
RA0*	Bending	28	CCGT/ACGG	27-28
RA2	Bending	52	CAA/TTG	32
RB2	Bending	1	ATG/CAT	80
RC0	Bending	22	TATT/AATA	60-61
RC2	Bending	1	CAAG/CTTG	47-48
FA0	Bending	4	TAA/TTA	15
FA2	Bending	2	TTTG/CAAA	43-44
FB1	Bending	9	ATG/CAT	80
FC1	Bending	9	CGGG/CCCG	65-66
FA1	Writhing	93	TAT/ATA	4
FB0	Writhing	61	TATTA/TAATA	60-62
FB2	Writhing	89	ATAT/ATAT	12-13
FC0	Writhing	47	TATCG/CGATA	54-56
FC2	Writhing	26	TATAA/TTATA	13-15

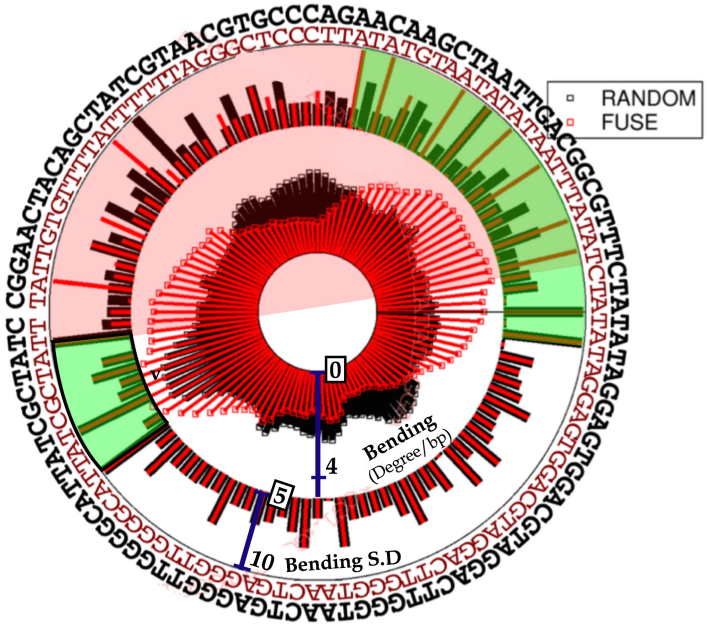
Table 3.1 Details of all the structural defects observed in nine MD simulations from 108 bp ΔLk = -1 RANDOM sequence and FUSE-embedded sequence DNA minicircles. Global shapes of denaturations described in Figure 3.9 (untwisting, bending and writhing) are given, along with the length of time taken for defect formations, and disrupted base sequences and numbers.

* Another defect formed in **RA0** simulation after 44 ns

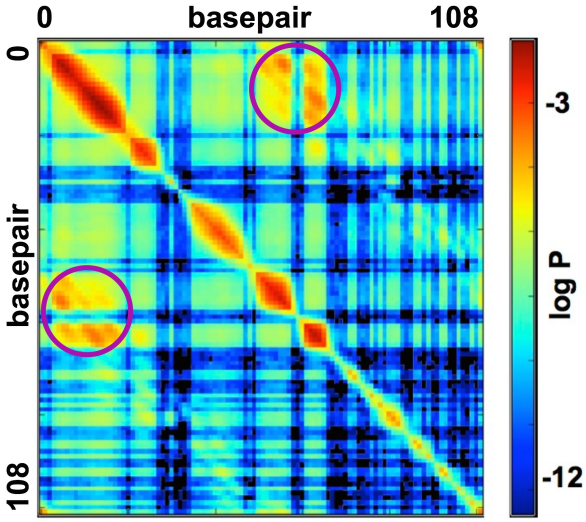
Details of defect formations in RANDOM and FUSE-embedded DNA minicircles at ΔLk = -1 are given in Table 3.1. The simulations were named after their sequences (R for RANDOM and F for FUSE-embedded), starting bending register angles (A, B and C) and replica numbers (0, 1 and 2). The table also indicates that defect formation in RANDOM and FUSE-embedded minicircles at ΔLk = -1 involved only YR steps. Denaturation bubbles (that occurred primarily due to DNA untwisting) were found only in the RANDOM sequence, while kinking defects established after DNA writhing due to cooperative bending were observed only in the FUSE-embedded sequence, as embedding a FUSE element into RANDOM minicircles had placed two flexible regions containing relatively high YR step content opposite to each other.

This was confirmed by the axis-bending analysis by CURVES+ for the non-disrupted MD snapshots (see Figure 3.10a), where the two highly bent regions (highlighted in green) within FUSE-embedded minicircles made the two elliptic/plectonemic apices. In the atomistic MD simulations of the FUSE sequence, larger portion of the superhelical stress is partitioned into DNA writhing compared to the RANDOM sequence. This suppressed defect formation in the FUSE-embedded minicircle MD simulations, which required up to 93 ns to be established, compared to 52 ns for the longest RANDOM minicircle MD simulations (see Table 3.1). This subtle effect cannot be captured either by SIDD or oxDNA because these coarse-grained DNA representations rely on thermodynamic melting studies for their parameterisation, and cannot account for the detailed sequence dependent mechanics of propeller twist illustrated in Figure 1.5, Chapter 1.

While the oxDNA simulations are less detailed than the atomistic models, they do have the advantage that they can explore sufficiently long simulation timescales that the denaturation of the minicircles was reversible, and well-converged denaturation probabilities can be obtained. Figure 3.10b shows the mutual denaturation probability for pairs of base-pairs within a simulation of FUSE embedded 108 bp minicircle at $\Delta Lk = -1$. The diagonal peaks show the probability profile of a single denaturation event, while the cross peaks highlighted by circles indicate cooperative denaturation between two distal regions. These occur at steps with high YR content, and at an equivalent location to the highly bent regions observed in the atomistic MD simulations of the minicircles (see Figure 3.10a).



a) MD: Axis-Bending per base pair



b) oxDNA: Double Defects Probability

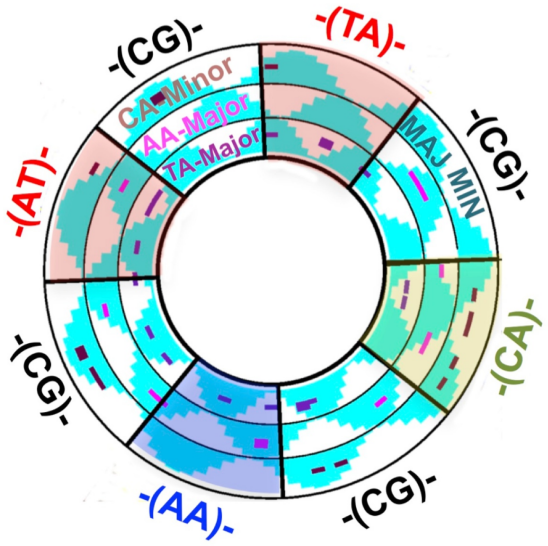
Figure 3.10 a) Polar plots containing the distribution of helical axis bending measured from ΔLk = -1 RANDOM and (red) FUSE-embedded minicircles: (***Outer***) the base sequences, (***Middle***) the relative bending flexibility for each basepair step in both sequences referred from the long timescale simulations performed by the ABC consortium (133), and (***Inner***) time-average bending extracted by CANAL over the non-defective structures from all simulations of ΔLk = -1 DNA of both sequences. Highly flexible regions of FUSE-embedded sequence highlighted in green corresponds to the high bending observed at two opposing regions. **b)** A 2D colour plot of mutual denaturation probability for each two base pairs in FUSE-embedded oxDNA minicircles. (The 2D plot was provided by Christian Matek). An off-diagonal peak (circled) is observed between the same two highly bent regions in atomistic model.

3.3.3) Effect of DNA Anisotropy Observed in Repeated Sequences

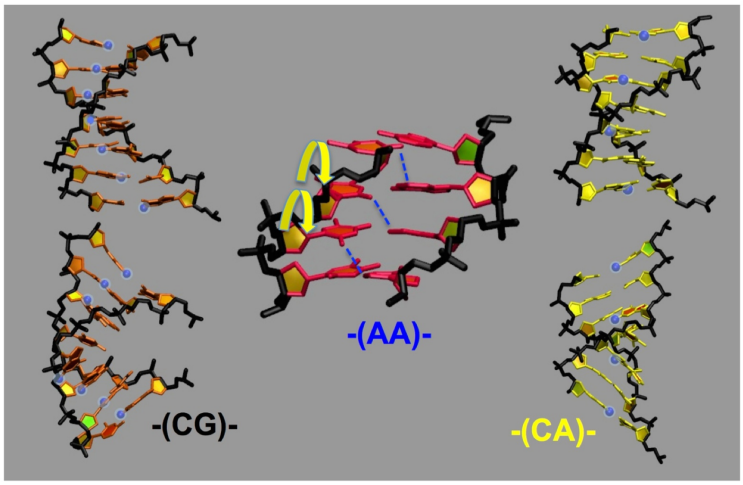
Anisotropy of DNA bending and the consequent structural defects were observed in the simulations of the DESIGNED sequence DNA minicircles at ΔLk = -1. Location of defect formation within 15 simulations for each bending register angle (CA-Minor, AA-Major and TA-Major) is shown in Figure 3.11a, along with the number of defects and the probability of defect formation within the major and minor groove blocks. The detailed description of defects observed within all the 45 atomistic MD simulations of ΔLk = -1 DESIGNED minicircles by the hydrogen bonding analysis are given in Appendix C.

Poly-CG regions preferred defect formations at the major groove block due to the steric clash between the extra amino groups at the minor groove of G-C base pairs (19), which prevented the bending towards the minor groove (see Figure 3.11b (left)) and were in agreement with the statistics gathered from x-ray crystal structures showing the tendency of minor groove opening (positive roll) (24). The poly-AA region also preferred defects that occurred at the major groove block, as the AA steps contained an inter-base pair hydrogen bonding keeps their minor groove narrow so that bending towards minor groove becomes energetically favourable (see Figure 3.11b (left)). This is consistent with a crystal structure consisting an A-tract [PDB number: 1ILC (56)], which is intrinsically bent towards the minor groove. TA and CA steps, however, favour bending towards its major groove, but are also allowed to bend towards the minor grooves as there were no steric clash between amino groups and were more prone to the kink formation (see Figure 3.11b (right)). Kinking found at the minor groove of TA and CA steps were also found in x-ray crystal structures of a nucleosome core particles [PDB number: 1AOI (48) and 3LZ0 (49)].

Regions	N	MAJ	P(MAJ)	MIN	P(MIN)
-(CG)-	60	19	0.007	3	0.0011
-(AA)-	12	4	0.0074	0	0
-(TA)-	24	3	0.0028	7	0.0065
-(CA)-	12	3	0.0056	10	0.0185



a) Anisotropic Kinking



b) Sequence Motifs

Figure 3.11 a) The numbers of defects and the probability of defect formation in major groove (**MAJ**) and minor groove blocks (**MIN**), observed in 45 atomistic MD simulations of ΔLk = -1 minicircles of the DESIGNED sequence at different sequence motif regions (see details in Appendix C). The polar charts show the location of all the defects mentioned in the table. **b)** Sample atomistic structures of three sequence motifs: poly-AA (A-tract) with the extra hydrogen bonds (blue dashed lines) due to the high propeller twists (curved arrows), poly-CG allows most defects to occur in the major groove blocks as the extra amino acid groups (blue dots) at each C-G base pairs prevent the kinking in the minor groove, and poly-CA allows defects to occur at both major and minor groove blocks.

3.4) Conclusion

A multi-scaled, collaborative study using a series of atomistic MD simulations, along with the stress-induced duplex destabilisation (SIDD) model, and a coarse-grained "oxDNA" model, have investigated the sequence dependent denaturation of DNA minicircles. The calculations were performed on the torsionally relaxed and negatively supercoiled (ΔLk = -0.5 and -1) ~100 bp minicircles of three different sequences: the RANDOM sequence was reproduced from previous experimental (11) and simulation studies (134), a 50 bp A-T rich FUSE element was then embedded into half of the RANDOM sequence and the DESIGNED sequence was created from repetitive sequence motifs.

In the atomistic MD simulations with ΔLk = -0.5, DNA base pair breathing was observed. Surprisingly, the smallest number of breathing events was found in the most A-T rich FUSE-embedded sequence. This unexpected result was due to the flexibility of pyrimidine-purine (YR) steps which promoted cooperative bending between two flexible regions with high YR step content, which was also observed for ΔLk = -1. Moreover, the oxDNA simulations observed co-operative denaturation events in the FUSE sequence, but not in the RANDOM sequence. The high bendability of the TA steps within the FUSE region is supported by a tethered particle motion (TPM) experiment (135).

DNA bending anisotropy was demonstrated through 45 atomistic MD simulations of ΔLk = -1 DESIGNED sequence minicircles, started from three different bending register angles. Poly-CG and poly-AA regions showed preferential defect formation at the major groove block, while poly-TA, poly-AT and poly-CA regions formed defects primarily in the minor groove block, which generally agreed with the observations from the X-ray crystal structures of nucleosome cores (48, 49). This DNA major-minor groove anisotropy cannot be seen either with SIDD or oxDNA models, since neither model represents these grooves in adequate detail.

In conclusion, the SIDD model and oxDNA are capable of providing sufficient sampling to quantify the sequence dependent denaturation probability of supercoiled DNA minicircles. Atomistic MD simulations are too computationally expensive to monitor reversible defect formation. However, atomistic MD simulations can provide an atomistically detailed description of the defects that form in mechanically stressed DNA conformations within the minicircles. The multi-scale approach revealed the

importance of base stacking interactions in quantifying the DNA structural disruptions and provided ideas for how the more coarse-grained models could best be modified to provide a more detailed description of the DNA, for example including the major and minor grooves in the coarse-grained oxDNA model.

Chapter 4
A New Definition of the Helical Axis for Quantifying DNA Writhing

4.1) Introduction

To gain insight into how superhelical stress affects local and global DNA structures, we have investigated the partitioning between twist and writhe using atomistic MD simulations. The circles in Chapter 3, which contain only around 100 base pairs, are shorter than the DNA persistence length (~150 bp) and DNA writhing was hardly observed. Instead, this chapter is focused on circular DNA molecules of a length of 258-340 bp. Now the circles are of sufficiently size that DNA writhing can be observed in the simulations, and the question of twist/writhe partitioning can be addressed (136).

Writhe can be calculated by counting the average number of crossing points over all the different viewing angles (137) or by the discretized Gauss' Integral (138), as described in Chapter 1. To calculate the writhe from an atomistic DNA structure, the mathematical definition of the DNA central axis is crucial. In this chapter, we compare three definitions to extract the DNA central axis from MD simulation data; X3DNA, CURVES+ and WrLine. WrLINE is a new technique we have developed for assigning the DNA central axis, which we have optimised to measure the writhe of supercoiled DNA structures obtained by MD. This new algorithm provides a reference point representing each base pair step from the running average over an exactly full helical turn, in which the flanking atoms are weighted in accordance with the increment in total twist from 2π.

We used a series of MD simulations of supercoiled DNA minicircle topoisomers to demonstrate the functionality of WrLINE. The MD simulations were run using the approximate GB/SA implicit solvent model due to its ability to explore conformational space rapidly in the absence of solvent friction. For each topoisomer, time-averaged writhe values were calculated using all three central axis definitions and compared with the radii of gyration of the writhed minicircle DNA.

4.2) Methods

4.2.1) Circle Sizes and Supercoiling Levels

The DNA minicircles chosen for this study correspond to those investigated by electrophoretic mobility experiments (27) and those used to demonstrate the potential of supercoiled DNA minicircles to act as gene therapy agents (9). The two smallest minicircles reported by these studies (260 bp and 336 bp) were selected, as their sizes were small enough for the all-atom calculations, but large enough for DNA writhing to be observed. Topoisomers having a superhelical density ranging from σ = -0.07 to +0.07 were chosen.

To obtain DNA topoisomers with as close to zero superhelical density and therefore zero writhe as possible, we also investigated a series of circle sizes in the vicinity of the most relaxed linking numbers (Lk_m = 24 for 260bp and Lk_m = 30 for 336bp minicircles), as described in Table 4.1. The differences in circle sizes caused slight changes in superhelical density. The criterion for assigning zero superhelical density was then the circle size for which the radius of gyration was found to be a maximum. Then, it was possible to assign the most accurate values of σ to all the other DNA minicircles, using Equations 1.6) and 1.9). Details of all the minicircle DNA simulations performed at different sizes and superhelical densities are given in Table 4.1 along with the writhe calculated from all three methods.

4.2.2) Implicit Solvent MD Simulations of DNA Minicircles

The circular DNA starting structures were built by using the same method described in Section 3.2.3. From this starting structure, a two-stage energy minimisation, and a multi-stage MD equilibration was performed prior to the productive MD runs, as described in detail in Appendix B. MD was performed using the GB/SA implicit solvent model with a screening parameter set to mimic 0.1 M salt concentration (98). For the equilibration and the productive MD runs, the velocities and coordinates of all the atoms were updated every 2 fs time step. The SHAKE algorithm (95) was implemented to freeze the covalent bonds involving hydrogen atoms. The temperature was kept constant using the Berendsen weak coupling scheme (96). NMR distance restraints were applied to all hydrogen bonds between complementary base pairs in the productive MD runs to stabilise the duplex DNA structure during the large conformational changes associated with writhing.

4.2.3) WrLINE Helical Axis

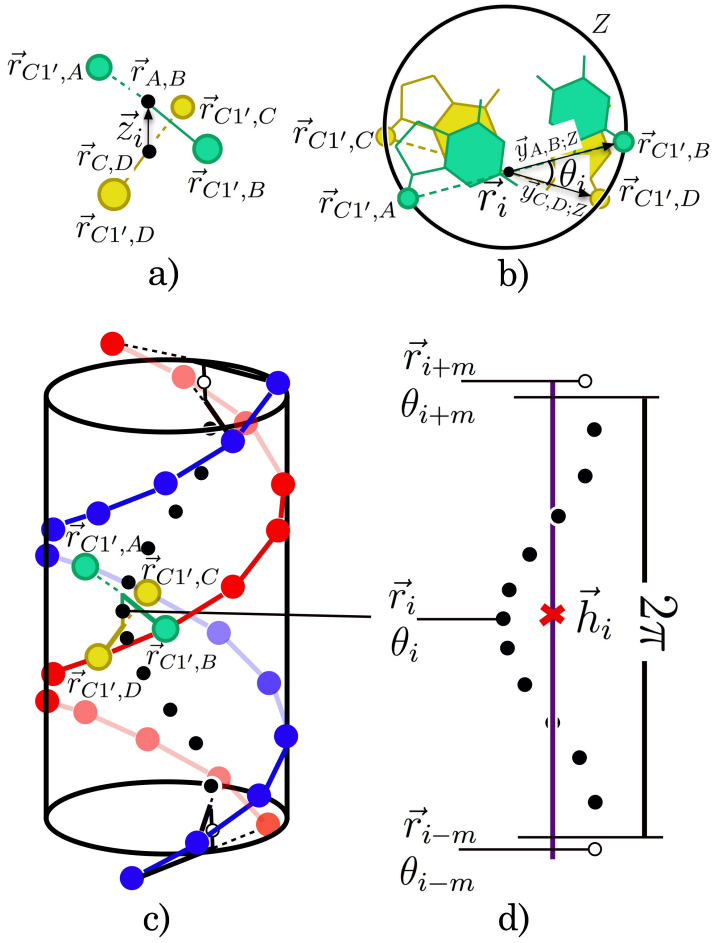


Figure 4.1 Diagrams showing the algorithm of WrLINE helical axis path **a)** The four C1′ atoms of two neighbouring base pairs and the corresponding local helical axis $\vec{z}_i$, **b)** simplified description of twist θ_i is given to the base pair step i, **c)** A full helical turn denoted by a cylinder, and **d)** The helical axis path $\vec{h}_i$ (red cross) obtained from averaging the path $\{\vec{r}_i\}$ over a full helical turn, where the flanking $\vec{r}_{i+m}$ and $\vec{r}_{i-m}$ points are corrected by the weighting term w.

The new helical axis calculation method, WrLINE, was defined based on the running average of the base pair step coordinates over an exact helical turn surrounding it. The running average was able to remove any irregularities and artefacts in the writhe calculation due to local changes and the incompleteness of the helical turn. An in-house Python script was written for the algorithm described in Figure 4.1.

A primary reference frame $\vec{r}_i$ is created for each of the neighbouring base pairs, located at the centre of four C1′ atoms, as shown in Figures 4.1a and 4.1b:

$$\vec{r}_i = \frac{\vec{r}_{C1',A} + \vec{r}_{C1',B} + \vec{r}_{C1',C} + \vec{r}_{C1',D}}{4}. \quad 4.1)$$

$\vec{r}_{A,B}$ and $\vec{r}_{C,D}$ vectors represent the midpoint between two C1′ atoms of each base pair. Then, the local helical axis $\vec{z}_i$ is defined by:

$$\vec{z}_i = \vec{r}_{A,B} - \vec{r}_{C,D}\,. \quad 4.2)$$

The vectors $\vec{y}_{A,B}$ and $\vec{y}_{C,D}$ connecting C1′,A with C1′,B and C1′,C with C1′,D are projected onto the Z plane. Then, the simplified definition of local twist at i^{th} step is:

$$\theta_i = \cos^{-1}(\vec{y}_{A,B;Z} \cdot \vec{y}_{C,D;Z}). \quad 4.3)$$

In a helical turn, the total twist of base pair steps surrounding the central step i with the local twist θ_i is evaluated. Consider adding the twist of $2m$ base pairs onto θ_i so that:

$$\Theta_m = \theta_i + \sum_{k=1}^{m} \theta_{i+k} + \theta_{i-k} \quad 4.4)$$

makes the total twist Θ_m larger than and closest to 2π:

$$\Theta_m > 2\pi > \Theta_{m-1}, \quad 4.5)$$

where

$$\Theta_{m-1} = \theta_i + \sum_{k=1}^{m-1} (\theta_{i+k} + \theta_{i-k})\,, \quad 4.6)$$

the total twist can be rearranged as:

$$\Theta_m = \Theta_{m-1} + (\theta_{i+m} + \theta_{i-m})\,. \quad 4.7)$$

In order to make an exactly full helical turn, replace Θ_m by 2π and introduce a weighting factor w so that:

$$2\pi = \Theta_{m-1} + w(\theta_{i+m} + \theta_{i-m}), \quad 4.8)$$

and

$$w = \frac{2\pi - \Theta_{m-1}}{\theta_{i+m} + \theta_{i-m}} = \frac{2\pi - \Theta_{m-1}}{\Theta_m - \Theta_{m-1}} \quad 4.9)$$

Finally, the reference point $\vec{h}_i$ at i^{th} step associated with the helical axis can be found by averaging the coordinates over the central i^{th} step, the surrounding 2(m-1) steps, and two flanking steps weighted by the factor w:

$$\vec{h}_i = \frac{1}{2(m+w)-1}\left(\vec{r}_i + w(\vec{r}_{i+m} + \vec{r}_{i-m}) + \sum_{k=1}^{m-1}(\vec{r}_{i+k} + \vec{r}_{i-k})\right). \quad 4.10)$$

The set $\{\vec{h}_i\}$ defines a path for the DNA minicircle helical axis for an MD timestep.

4.2.4) Writhe and Radius of Gyration

The global structural properties of supercoiled DNA minicircles in this chapter can be quantified by writhe (Wr) and the radius of gyration (R_g). Writhe directly addresses how the twist and writhe are partitioned in response to linking number changes (ΔLk), while the radius of gyration is related to the global compaction of the supercoiled DNA, as could be quantified by its mobility within gel electrophoresis experiments.

Consider a DNA minicircle helical axis represented by $\vec{h}(s)$ as a closed space curve of the length L. Writhe can be evaluated from Gauss' Integral:

$$Wr = \frac{1}{4\pi}\oint_0^L \oint_0^L \left(\frac{d\vec{h}(s)/ds}{\left|d\vec{h}(s)/ds\right|} \times \frac{d\vec{h}(s')/ds'}{\left|d\vec{h}(s')/ds'\right|}\right) \cdot \frac{\left(\vec{h}(s) - \vec{h}(s')\right)}{\left|\vec{h}(s) - \vec{h}(s')\right|^3} ds'ds. \quad 4.11)$$

Given a discrete set of representative coordinates of an n base pairs minicircle $\{\vec{h}_i\}$, the discretised Gauss' Integral can be written as:

$$Wr \cong \frac{1}{2\pi}\sum_{i=1}^{n}\sum_{j\geq i}^{n}\left(\frac{\vec{h}_{i+1} - \vec{h}_i}{\left|\vec{h}_{i+1} - \vec{h}_i\right|} \times \frac{\vec{h}_{j+1} - \vec{h}_j}{\left|\vec{h}_{j+1} - \vec{h}_j\right|}\right) \cdot \frac{\left(\vec{h}_i - \vec{h}_j\right)}{\left|\vec{h}_i - \vec{h}_j\right|^3}. \quad 4.12)$$

For each of the MD simulations in Table 4.1, time-average and SD of the writhe were calculated from the snapshots over the last 10 ns.

The radius of gyration is defined by the root mean square distance (RMSD) of the elements about the centre of mass $\vec{a}_{CM}$:

$$R_g \cong \sqrt{\frac{1}{M}\sum_{k=1}^{N} m_k(\vec{a}_k - \vec{a}_{CM})^2}. \quad 4.13)$$

The summation was performed over all the N atoms at coordinates $\{\vec{a}_k\}$ with masses $\{m_k\}$ for each timestep, and the time-average and SD of the R_g over the last 10 ns are shown in Table 4.1 for each simulation. DNA with lower R_g is more compact and would possess higher mobility in a gel electrophoresis experiment.

Length	Lk	Rg	σ	Wr(WrLINE)	Wr(CURVES+)	Wr(X3DNA)
336	**28**	132.2 ± 11.7	-0.067	-1.10 ± 0.12	-1.43 ± 0.13	0.12 ± 0.27
336	**29**	144.2 ± 12.9	-0.033	-0.62 ± 0.17	-0.91 ± 0.18	0.43 ± 0.31
340	**30**	153.2 ± 8.8	-0.012	-0.42 ± 0.21	-0.67 ± 0.22	0.66 ± 0.38
338	**30**	161.0 ± 5.8	-0.006	-0.18 ± 0.11	-0.41 ± 0.12	0.97 ± 0.23
336	**30**	170.9 ± 3.2	0.000	0.05 ± 0.07	-0.20 ± 0.08	1.05 ± 0.19
334	**30**	165.9 ± 3.8	0.006	-0.01 ± 0.07	-0.25 ± 0.09	1.02 ± 0.20
332	**30**	162.3 ± 6.7	0.012	0.06 ± 0.09	-0.17 ± 0.10	1.06 ± 0.19
336	**31**	157.2 ± 7.9	0.033	0.50 ± 0.14	0.26 ± 0.14	1.20 ± 0.23
336	**32**	138.7 ± 10.0	0.067	1.11 ± 0.12	0.86 ± 0.12	1.56 ± 0.21
260	**22**	101.1 ± 5.6	-0.069	-1.08 ± 0.13	-1.36 ± 0.14	-0.43 ± 0.24
260	**23**	111.6 ± 6.3	-0.027	-0.31 ± 0.19	-0.54 ± 0.20	0.35 ± 0.33
266	**24**	131.0 ± 4.2	0.007	-0.04 ± 0.11	-0.25 ± 0.11	0.60 ± 0.18
264	**24**	133.5 ± 2.9	0.000	0.03 ± 0.06	-0.19 ± 0.08	0.61 ± 0.14
262	**24**	131.7 ± 2.5	0.007	0.11 ± 0.05	-0.11 ± 0.16	0.62 ± 0.14
260	**24**	130.4 ± 3.0	0.015	0.14 ± 0.07	-0.08 ± 0.25	0.57 ± 0.15
258	**24**	128.3 ± 3.1	0.023	0.16 ± 0.12	-0.02 ± 0.22	0.57 ± 0.18
260	**25**	121.8 ± 7.5	0.058	0.76 ± 0.16	0.56 ± 0.16	1.05 ± 0.23

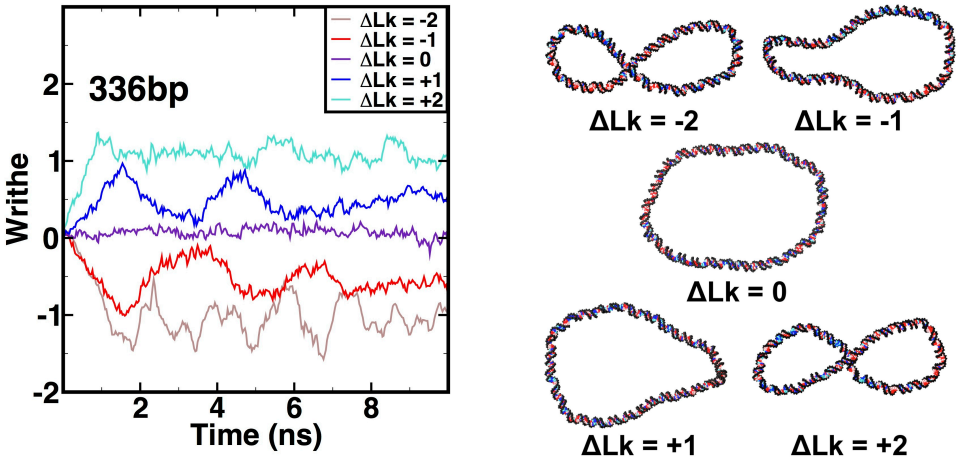
Table 4.1 Average and SD values of writhe and radius of gyration calculated over the last 10 ns of each implicit solvent MD simulation in this chapter. Writhe values were calculated from three definitions of helical axis: WrLINE, CURVES+, and X3DNA.

4.3) Results

4.3.1) Dynamics of DNA Writhing

Figure 4.2a, shows the writhing behaviour of the five topoisomers of 336 base pairs DNA minicircles in GB/SA implicit solvent as the simulations progressed. Starting from the planar circular conformation at $Wr = 0$, torsional stress stored within the over-/under-twisted structures was partly converted into bending stress and minicircles developed their writhed conformations. Time-dependent writhe values were calculated using WrLINE helical axis paths for the first 10 ns. Only small fluctuation was observed in writhing of the relaxed topoisomer (purple line in Figure 4.2a). For the other four supercoiled topoisomers, rapid changes were observed over the first 2 ns, after which the writhe oscillated around the mean values.

The corresponding atomistic structures of the writhed minicircles are shown in Figure 4.2b. The torsionally relaxed (ΔLk = 0) minicircle maintained its open conformation which corresponded to zero writhe. For the highly supercoiled minicircles at ΔLk = ±2, the writhe values were found to fluctuate around +1 for ΔLk = +2 and -1 for ΔLk = -2, which suggested the figure-8 writhed configurations. However, for the slightly supercoiled minicircles at ΔLk = ±1, writhe values fluctuated between 0 and ±1. In these cases, both open and figure-8 configurations were observed.



a) b)

Figure 4.2 a) Writhe calculated using the WrLINE helical axis for the first 5 ns of five 336 bp topoisomers, **b)** Sample snapshots showing the writhed global shape of the five topoisomers.

4.3.2) Writhe Calculations Compared to X3DNA and CURVES+

In order to validate the use of WrLINE helical axis, writhe values calculated using helical axis paths from WrLINE were compared with those from X3DNA and CURVES+. To map writhe values from the three helical axis definitions with the superhelical densities of different minicircle topoisomers, zero superhelical density (σ = 0) needed to be assigned for torsionally relaxed minicircles with the maximum radius of gyration R_g.

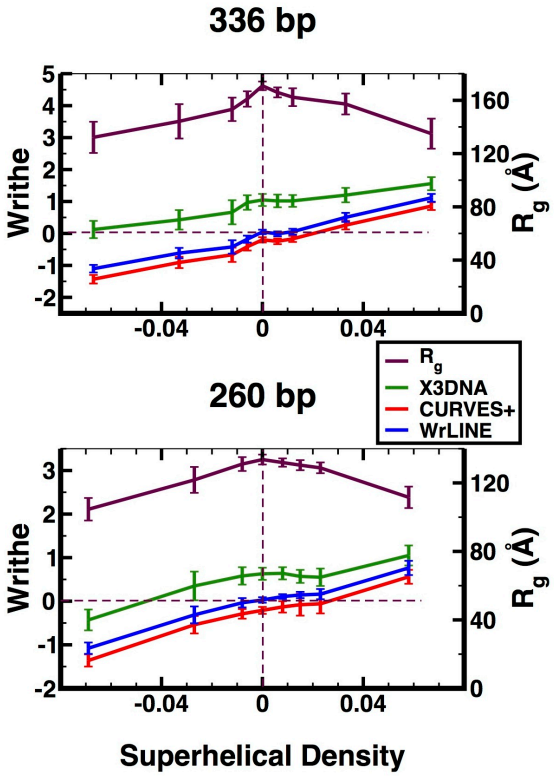


Figure 4.3 Time averages and standard deviations of radius of gyration (maroon) and writhe as functions of superhelical density of the 260 base pairs and 336 base pairs minicircles calculated from the last 10 ns of implicit solvent MD simulations in Table 4.1, using three different definitions for the helical axis paths: X3DNA (green), CURVES+ (red), and WrLINE (blue).

In Figure 4.3, the vertical dashed lines indicate zero superhelical density at the maximum radii of gyration observed for 264 bp at Lk = 24 and for 336 bp at Lk = 30, and the horizontal dashed lines indicate Wr = 0. Then, σ values were assigned by Equation 1.6 for the original simulations in which the linking numbers vary, and by Equation 1.9 for the additional simulations where the DNA lengths are also changed (see Table 4.1).

The helical axis path definition should give zero writhe for zero superhelical density (e.g. the intercept between the vertical and the horizontal dashed lines in Figure 4.3). While the X3DNA helical axis path at maximum radius of gyration greatly overestimated writhe at σ = 0 (+0.61 ± 0.14 for ~260bp and +1.05 ± 0.19 for ~336bp minicircles), writhe evaluated from CURVES+ helical axis paths were slightly underestimated (-0.19 ± 0.08 for ~260bp and -0.20 ± 0.08 for ~336bp minicircles). The WrLINE helical axis paths provided the most accurate writhe values (+0.03 ± 0.06 for ~260bp and +0.05 ± 0.07 for ~336bp minicircles) as the $Wr(\sigma)$ functions of WrLINE were found closest to the $(\sigma = 0, Wr = 0)$ intercepts.

4.3.3) Local Periodicity

In this section, the source of writhe calculation artefacts were investigated. Figure 4.4 shows a sample snapshot taken from ΔLk = -2 (top panels) and the ΔLk = +2 (bottom panels) topoisomers of 336 bp minicircles. X3DNA paths (green) were superimposed onto the WrLINE paths (blue) and made a right-handed toroidal superhelix about the WrLINE paths, resulting in the overestimation of writing. The left-handed superhelix was observed for the CURVES+ paths (red), which signified the underestimation of writhing.

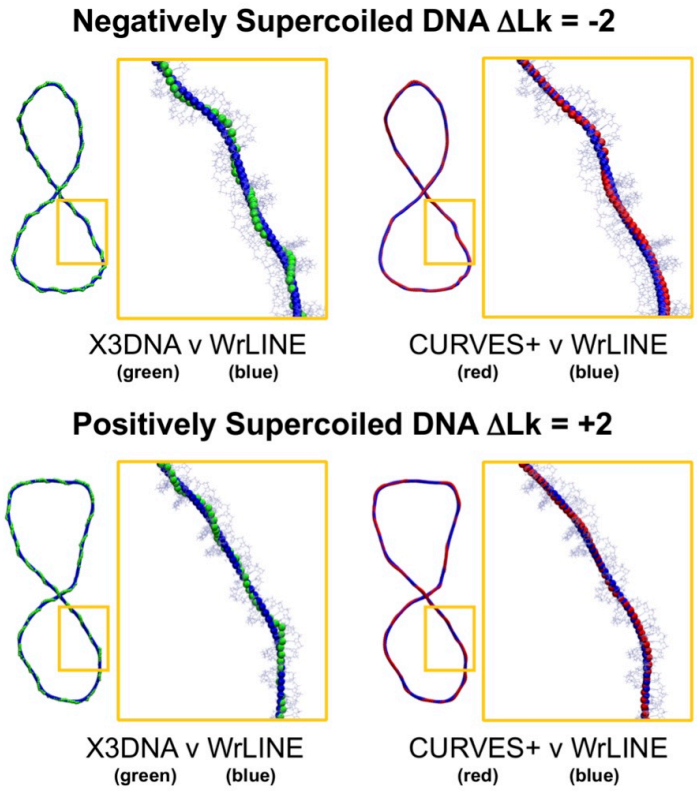


Figure 4.4 Sample conformations of the (***upper***) negatively supercoiled (ΔLk=-2) and (***lower***) positively supercoiled (ΔLk=+2) 336 base pairs minicircles. WrLINE helical axis paths (blue) are compared with those of X3DNA and CURVES+. The latters showing their local periodicity as sources of writhe calculation errors (shown in yellow frames).

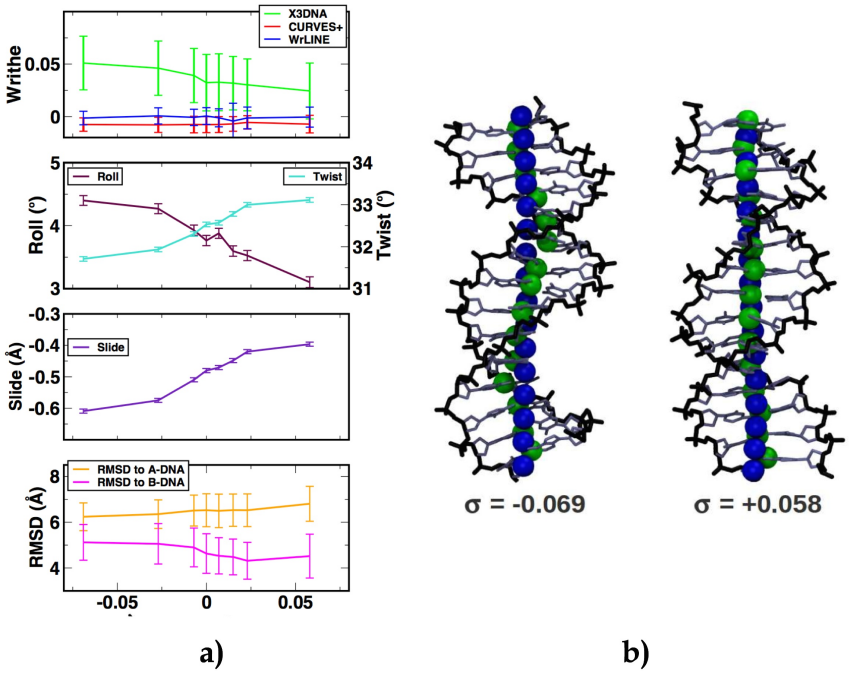


Figure 4.5 a) Writhe, roll, twist and RMSD compared to A-DNA and B-DNA structures calculated from a 20 bp fragment cut from each of the ~260 base pairs minicircle simulations. Time-averaged and standard deviations of these values are calculated from the last 10 ns of the trajectories, and **b)** Molecular structures indicating the WrLINE (in blue) and X3DNA (in green) helical axis paths of the 20 bp DNA fragment at σ = -0.069 and σ = +0.058.

To demonstrate that the local periodicity of the toroid-shaped helical paths of X3DNA and CURVES+ were the source of over-/under-estimation in writhe calculation, we considered a 20 bp DNA segment from each simulation of the ~260 bp minicircle located far from the plectoneme apices and unaffected by bending. Local contribution of the 20 bp segment to DNA writhing was calculated by using Gauss' integral over the helical axis path of that segment (see Figure 4.5a). Local writhe measured from the WrLINE helical axis paths of the 20 bp segment were close to zero, proving the appropriateness of the method for calculating writhe from atomistic MD structure. While CURVES+ slightly underestimated DNA writhing, the values calculated using X3DNA helical path definition were greatly overestimated, in agreement with the apparent local superhelicity shown in Figure 4.5b. However, the largest systematic writhe discrepancy was observed when DNA is negatively supercoiled.

We then extracted the average values of roll, twist, and slide for the 20 bp segment from each simulation of the ~260 bp minicircle. Roll and twist were negatively correlated (see Figure 4.5a), as observed in the x-ray crystal

structures (24). Negative supercoiling directly led to DNA untwisting and minor groove opening (+ve roll). Positive correlation between twist and slide corresponded to further negative slide. This resulted in the deviation of the local base pair reference frame in X3DNA from the centre, as seen in green in Figure 4.5b (left); this generated a toroidal path and was the source of the discrepancy with the axis defined by WrLINE.

These observations on local periodicity provided further insight into how DNA responds to superhelical stress. The ideal B-DNA structure has zero slide and roll, and coincides with the minimum helical diameter, while A-DNA features positive roll, negative slide and wide helical diameter. Figure 4.6a (bottom) shows that the linear subsegment became more closely matched with the A-DNA form and deviated further from the B-DNA form as the DNA was undertwisted.

4.4) Conclusions

Writhe calculation by the discretised Gauss' integral method (137) was performed on the helical axis paths from three different definitions of the helical axis: X3DNA, CURVES+ and WrLINE. A series of implicit solvent MD simulations of ~260bp and ~336bp supercoiled DNA topoisomers were performed in order to provide sample trajectories. In order to rationalise the writhe calculation results, we considered the relaxed DNA, corresponding to the maximum radius of gyration, obtained from fine-tuning the DNA by adding or cutting off a few base pairs to create small changes in superhelical density. The new method for defining DNA helical axis paths from MD trajectory, WrLINE, provided the writhe measurement closest to zero at zero superhelical density, while the other methods X3DNA and CURVES+ (131, 139), introduced artefacts due to their local helix axis periodicity.

In order to examine local periodicity carefully, the local writhe was calculated from the non-apical 20 bp segments. The local writhe of the 20bp X3DNA contour was greatly overestimated, especially when the DNA was negatively supercoiled. However, the errors were larger as DNA was negatively supercoiled. The analysis of base pair parameters and the RMSD calculations comparing the negatively supercoiled DNA structure to A- and B-DNA structures were performed. The results showed that DNA responded to negative superhelical stress by adapting itself towards the A-form, as illustrated by the decrease of RMS deviation from the A-form

structure. The widened DNA diameter shown by the slide in Figure 4.5a resulted in the larger helical periodicity and created artefacts in writhe calculation by X3DNA.

In conclusion, this chapter presented an alternative method of defining the helical axis path, which has been demonstrated to be more compatible with the calculation of writhe and is therefore useful for further studies on DNA topology based on the atomistic DNA structures. Moreover, the structural analysis performed in this chapter provided another insight to the mechanical response of DNA to superhelical stress.

Chapter 5
Atomistic Representations of Supercoiled DNA Minicircles in Explicit Solvent

5.1) Introduction

DNA deformability plays an important role in protein recognition and DNA packing (3). As an example, human topoisomerase 1B preferentially binds to supercoiled DNA and allows the DNA to relax. To study the details of topoisomerase-DNA interactions, the atomistic structure of topoisomerase-DNA complex was determined by x-ray crystallography (140). However, the DNA fragment bound to the topoisomerase was torsionally relaxed and was not subjected to any topological constraint, therefore, the information of supercoiled DNA structure cannot be provided.

To understand the mechanism of how DNA responds to mechanical stress imposed under the different topological states, DNA minicircles have been used as the probes for DNA topology (136). As an example, the writhing and compactness of a 336 bp supercoiled DNA minicircle has been addressed by electrophoresis experiments and compared with AFM images for a spectrum of topoisomers (27). Here, the latest cryo-electron tomography (cryo-ET) experiments, which have provided 3D tomographic images of 336 bp supercoiled DNA minicircles at different linking numbers (16). At the atomistic resolution, however, the conformational diversity of writhed DNA structures, even for a single topoisomer, makes it difficult to obtain the structures from x-ray crystallography. Closed DNA topologies are also too large for standard NMR structure determination experiments. To bridge the gap between the system size and resolution, we have used MD simulations in conjunction with cryo-ET to provide an atomistic representation of supercoiled DNA.

The series of implicitly solvated MD simulations performed in the Chapter 4 demonstrated the writhing behaviour of DNA at different topological states. Moreover, a numerically exact method to address the twist/writhe partitioning problem from the atomistic DNA structures has been developed. However, implicit solvent simulations are unable to provide the most precise atomistic representations of the DNA structures due to the structural artefacts introduced by the continuum solvent approximation.

To remove these artefacts, a hybrid implicit/explicit solvent simulation protocol has been developed for the MD simulations in this chapter. Combining the explicitly solvated MD structures with the cryo-ET results, we present the most exact atomistic representation of supercoiled DNA to date. In related calculations, structural parameters were then analysed and used to explain the mechanism of topoisomerase binding to a supercoiled DNA minicircles.

5.2) Methods

5.2.1) Cryo-ET Images and Representative Traces

Cryo-ET density maps and their computational traces were provided by Rossitza Irobalieva (*personal communication*). 336 bp DNA minicircles were produced by λ-Int recombination technique. Negative and positive supercoiling were introduced by using ethidium bromide and HMfB. Minicircle DNA topoisomers were then separated by gel electrophoresis. After that, sample grids containing 100ng/μl of DNA minicircles in 0.01 M $CaCl_2$ were prepared for the electron cryo-tomography and the -60° to +60° tilt-series of images were obtained through a 200 kV microscope and a 4k×4k CCD camera. From these image series, the reconstructed 3D density maps were normalised and filtered. And then, traces representing the minicircle shapes were optimised by an iteration method performed on a composite polygon of 20 vertices over the density gradient (see (16) for more detailed method description). These computational traces were used to compared with the explicitly solvated atomistic MD simulations in this chapter.

5.2.2) Hybrid Implicit/Explicit Solvent MD Simulation Setup

336 bp minicircles

A combination of implicit and explicit solvent (see Appendix B) MD simulation protocols was developed to obtain the structure of supercoiled DNA in aqueous solution in atomistic detail (see Figure 5.1).

For each of the seven 336 bp topoisomers (Lk = 27, 28, …, 33), two replicas of 20 ns AMBER GB/SA implicit solvent MD simulations were performed with the NMR distance restraints applied to maintain Watson-Crick hydrogen bonding. Each implicit solvent simulation was started from the planar circular configurations. The neglect of solvent friction in the GB/SA calculations enables the DNA to rapidly relax into writhe

configurations, and then the molecule continues to undergo large thermal fluctuations as it explores conformational space. Once the writhed structure obtained from the implicit solvent simulation is solvated in TIP3P water box and 0.1 M NaCl ions, the friction from the water "freezes" the global conformation, and only local structural changes are observed.

To choose the best starting structures for the explicit solvation of each topoisomer, the average linkage algorithm (141) was performed to cluster the supercoiled DNA conformations from the last 15 ns of each implicit solvent MD simulation replica into six clusters. Then, from each of the highly populated clusters, a snapshot with no structural disruption was chosen. After that, it was necessary to perform a 3000-step minimisation in vacuum to reduce the large electrostatic repulsion between phosphate groups that resulted from the minor groove collapsing artefact (further discussion in section 5.3.2), at which point the supercoiled DNA structure was considered ready for the 10 ns explicitly solvated MD run (see Section 3.2.3 for the detailed explicit solvent simulation protocols). The explicitly solvated systems, consist of over 2,000,000 atoms including water and ions, were run by using GROMACS 4.5 on the UK supercomputer nodes, ARCHER, with 256 processors and are able to produce MD trajectory of ~5ns /day.

240 bp minicircle-topoisomerase complex

A similar approach was applied to a negatively supercoiled 240 bp minicircle-topoisomerase complex (17). To supercoil the circle containing the catalytic site, the 22 bp topoisomerase binding DNA sequence was embedded into the supercoiled 240 bp minicircle and a writhed conformation was obtained from an implicit solvent simulation. To build the protein-DNA complex into a minicircle, the x-ray crystallographic structure of the linear 22 bp DNA-topoisomerase complex was obtained from the protein data bank (PDB number 1A36 (142)). The topoisomerase 1B enzyme was then introduced into the embedded catalytic site in the writhed DNA minicircle by rms fitting and docking, the complex was explicitly solvated, and further MD productive runs were performed for 50 ns.

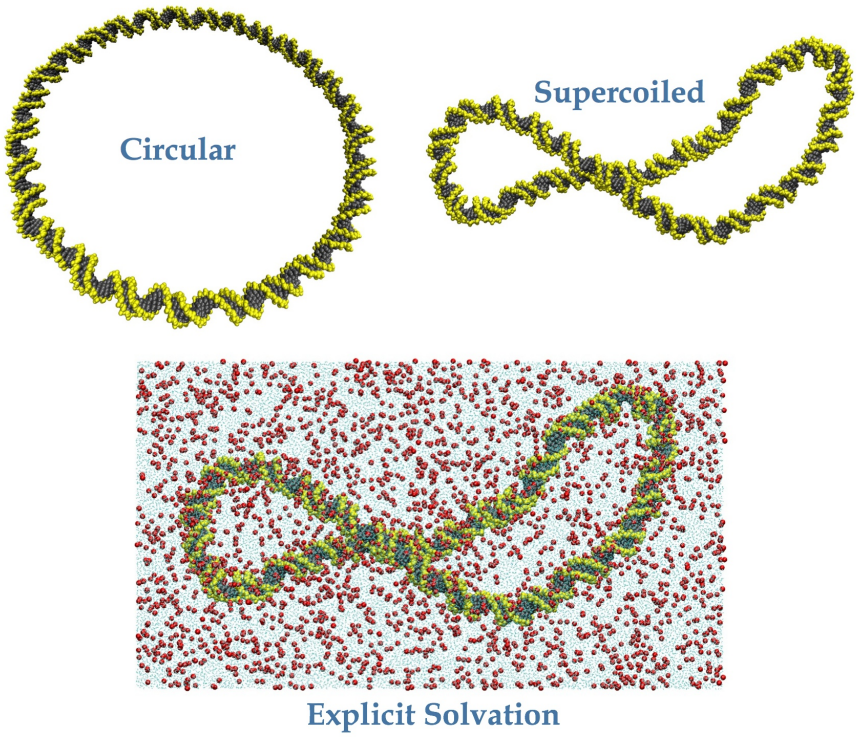


Figure 5.1 Hybrid implicit/explicit solvation: A planar circular 336 bp DNA minicircle at ΔLk = -2 were folded in GB/SA implicit solvent into the supercoiled structure. Then the folded (writhed) structure was solvated in 0.1M NaCl and TIP3P waterbox. Similar approach has been used to solvate the other 336 bp minicircle topoisomers and the 240 bp minicircle.

5.2.3) Analysis

336 bp minicircles

The writhe and the radius of gyration were calculated for each of the 336 bp topoisomers using the methods defined in Chapter 4. Helical and groove parameters were extracted by the program CURVES+. Time average values were then calculated for all the parameters at each base pair.

In order to compare the supercoiled DNA structures from MD simulations to the structures obtained from cryo-ET density maps, computational traces of the cryo-ET structures were given (see Section 5.2.1). A trace consisted of 20 coordinate points representing 336 base pairs. Then, the equivalent computational traces were produced from the corresponding MD trajectory. After that, the RMSD between the cryo-ET trace and the MD trace was calculated for each MD snapshot. The snapshot with minimum RMSD value was then superimposed into the cryo-ET density trace, as shown in Figure 5.2.

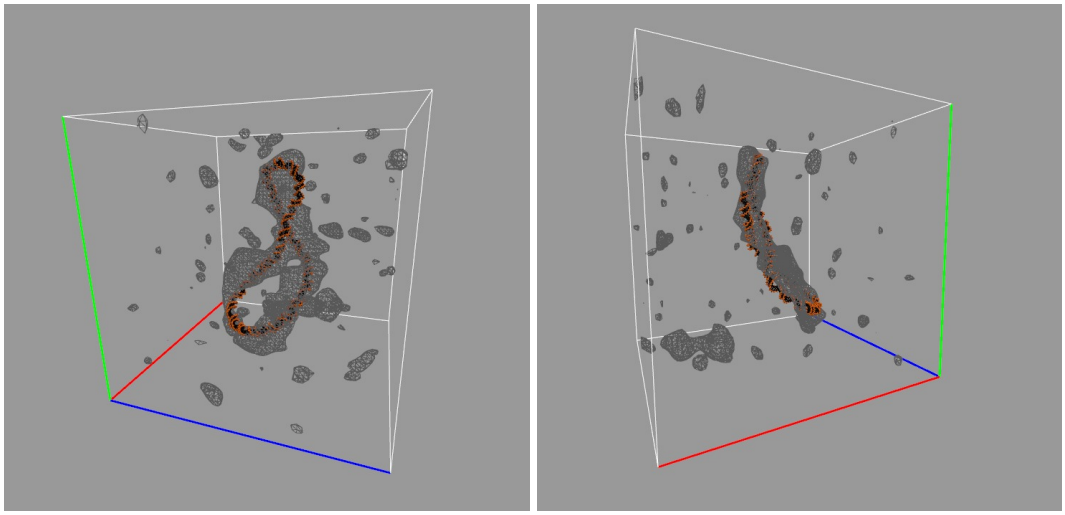


Figure 5.2 An atomistic structure of a ΔLk = -2 DNA minicircle, obtained from an explicity solvated MD simulations.. The structure is superimposed into a cryo-ET density map and visualised by VMD at two different viewing angles.

240 bp minicircle-topoisomerase complex

Protein-DNA interactions were quantified by measuring the occupancy of hydrogen bonds between amino acid and nucleotide residues. The PTRAJ module implemented in the AMBERTOOLS package provided a tool for keeping track of the distances and angles. All possible donor atoms from backbone phosphate ions and all possible acceptor atoms from the positively charged amino acid functional groups were tracked. For each MD snapshot, hydrogen bonding occurred if the interatomic distance between donor and acceptor atoms was smaller than the 3.5 Å cut-off and the angle made by the three atoms (donor--H--acceptop) was smaller than 60° cutoff. The hydrogen bonding occupancy for each donor-acceptor pair was defined by the probability that the hydrogen bonding criterion was met.

5.3) Results

5.3.1) Radius of Gyration Compared to Gel Mobility

A series of implicit solvent simulations of 336 bp DNA minicircle topoisomers was performed. These simulations provided an effective conformational searching tool, and writhed DNA conformations were rapidly explored in the absence of hydrodynamic forces from the collisions with the water molecules. Despite the appearance of localised structural artefacts (discussed later in the Section 5.3.2), supercoiled DNA structures in

the approximated solvent environment were still acceptable for the calculations of global properties, such as writhe and radius of gyration.

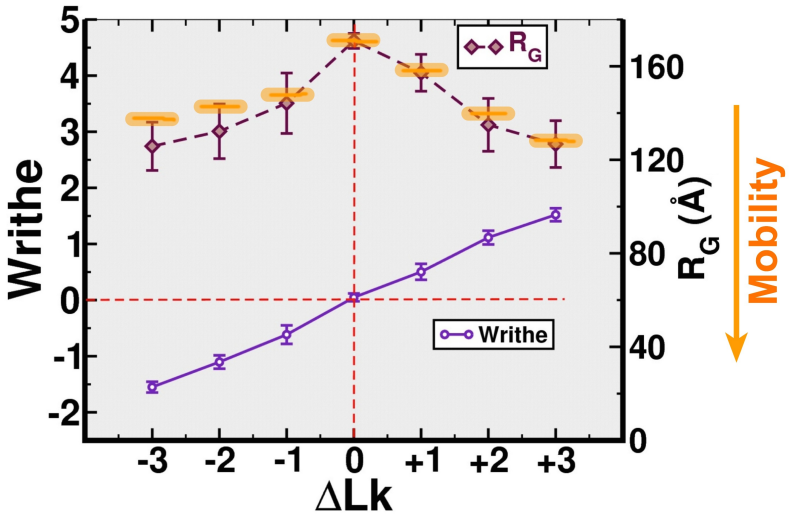


Figure 5.3 Time averages of writhe (purple) and radius of gyration (maroon) over the last 10 ns of the implicitly solvated DNA minicircles at seven different ΔLks. Orange marks indicate the relative mobility (down the vertical axis) of the seven minicircle topoisomers within polyacrylamide gel determined from an electrophoresis experiment on the same DNA minicircle topoisomers (16).

Figure 5.3 shows the values of writhe and radius of gyration averaged over the last 10 ns for each 20 ns simulation of DNA minicircle topoisomer ranging from ΔLk = -3,-2,…,+3. Torsionally relaxed minicircles at ΔLk = 0 had zero writhe and were in the open circular form with the minimum compactness, hence the maximum radius of gyration. The positive and negative supercoiled topoisomers had non-zero writhe and the structures were more compact, hence the reduced radius of gyration.

A comparison was made between the radii of gyration and the mobility (orange bands) of DNA minicircles obtained by electrophoresis experiments (16). Topoisomers with smaller radius of gyration are expected to propagate faster in the gel (down the vertical axis of the plot in Figure 5.3) when accelerated by an electric field. For the positively supercoiled topoisomers, the radius of gyration calculated from the MD trajectories generally agreed with the gel mobility results. However, differences were seen at the ΔLk = -2 and ΔLk = -3 topoisomers with the decreased gel mobility, suggesting that DNA was less compact in the experiments than predicted by MD simulations. This is to be expected, as in the experiments the formation of flexible single stranded DNA regions relieves superhelical

stress within the DNA and enables it to unwrithe, making a less compact structure. In the implicitly solvated calculations, however, denaturation is prohibited by hydrogen bond restraints, so that the superhelical stress purely contributes to twisting and writhing.

5.3.2) Supercoiled DNA in Explicit Solvent Compared to Cryo-ET

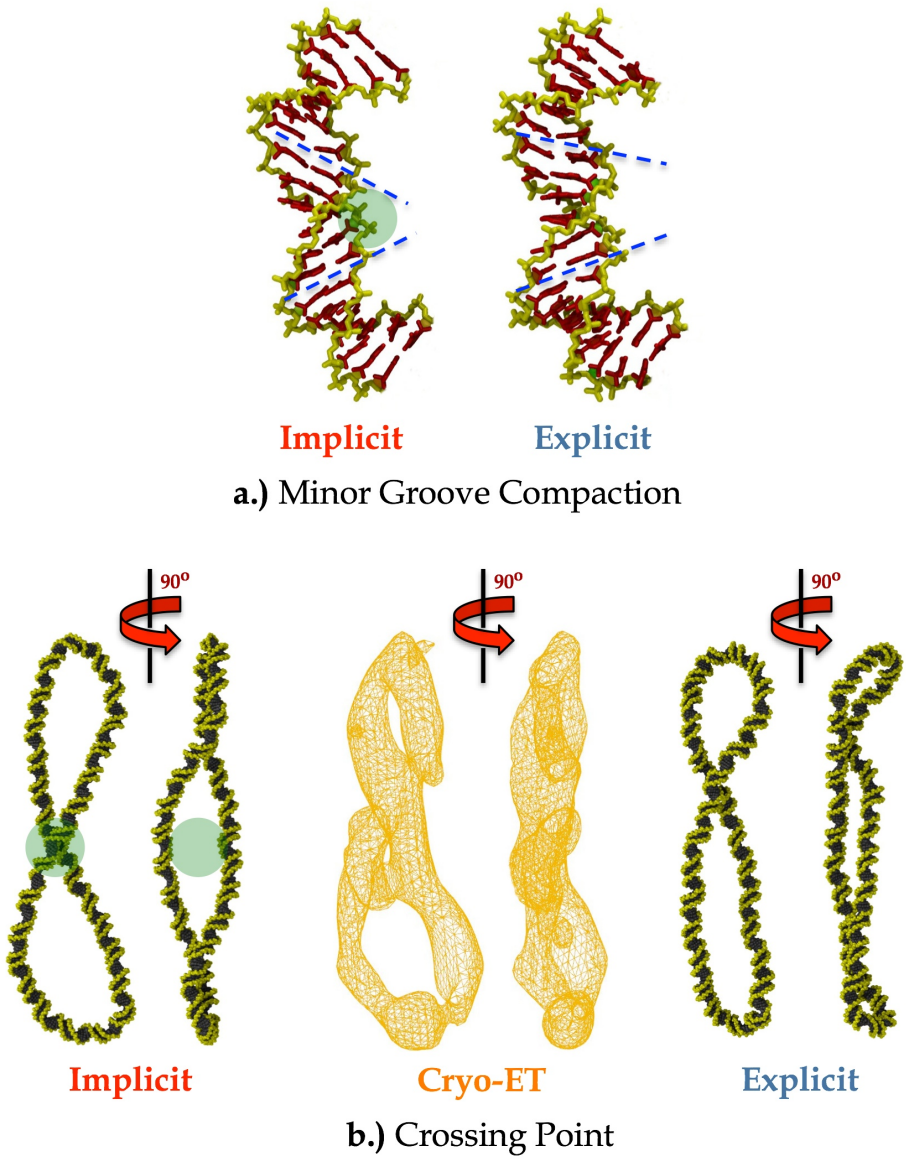


Figure 5.4 Structural artefacts (highlighted in green) found in the implicitly solvated DNA: **a)** Minor groove compaction was found at the bent plectoneme tip in the implicitly solvated DNA but absent in the explicitly solvated DNA, and **b)** Crossing point repulsion resulting in a gap between two juxtaposed segments, which was also absent in structures from cryo-ET and explicitly solvated simulations.

Implicit VS Explicit Solvation

To obtain more accurate structural information for the DNA minicircles, implicitly solvated supercoiled minicircle structures from the most populated conformational clusters were selected for subsequent explicitly solvated MD calculations. Figure 5.4 shows the comparison between DNA structures in implicit and explicit solvents. Minor groove compaction was commonly found at the implicitly solvated DNA plectoneme tips (see Figure 5.4a; highlighted in green). This artefact was not seen in the explicitly solvated DNA as the counterions bound into the minor groove could prevent the collapse. Moreover, the invariance of the salt distribution in the approximate GB/SA implicit solvent calculations limited the charge screening at the plectoneme crossing points. Thus, the crossing DNA segments were highly affected by electrostatic repulsion, resulting in the larger distance between the two segments (see Figure 5.4b (left); highlighted in green). However, in explicit solvent simulations, the positive counterions were free to diffuse into the highly charged region of the crossing points. The enhanced screening effect by the counterions could reduce the electrostatic repulsion and brings the two crossing segments closer together (see Figure 5.4 (right)), and make a better agreement with the cryo-ET, also having two crossing segments in close proximity (see Figure 5.4 (centre)).

Explicitly Solvated MD Structures VS Cryo-ET Density Maps

Explicitly solvated 336 bp minicircle structures from MD were compared to the electron density maps observed from cryo-ET (Rossitza Irobalieva, *personal communications*) by picking the representative molecular traces from both methods with the best RMSD fitting between each other. Figure 5.5 shows the examples of MD structures fitted to computational traces of cryo-ET density maps. The degree of difference in cryo-MD fitting were given by the RMSD values. For each conformation, RMSD values were not larger than 10.3% compared to the end-to-end distance of the minicircles.

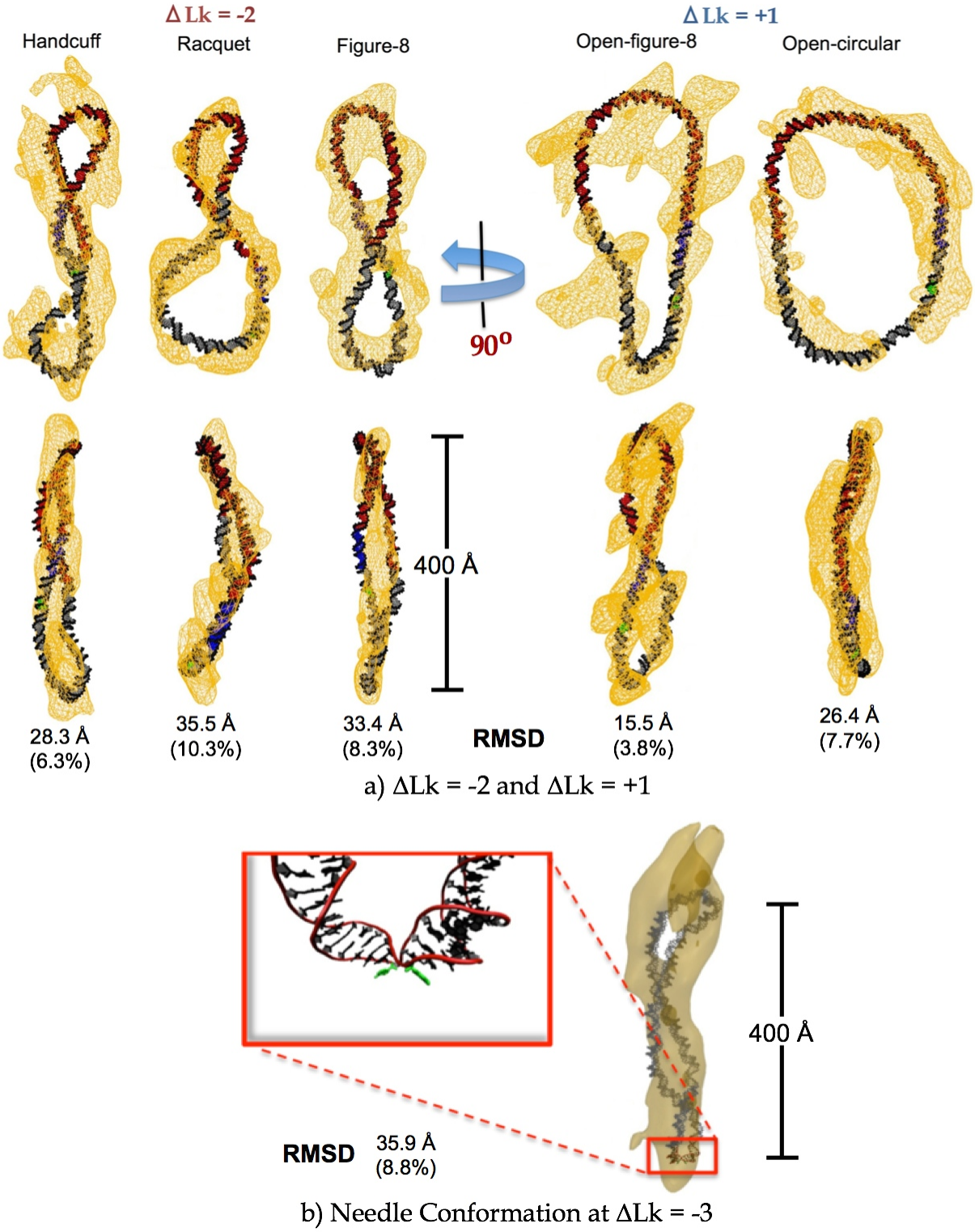


Figure 5.5 a) Different minicircle conformations observed in Cryo-ET and MD simulations at ΔLk = -2 and ΔLk = +1 at two different viewing angles and **b)** A kink observed at a plectoneme tip of the ΔLk = -3 minicircle in the 'Needle' conformation. The fitting RMSD value and the percentage compared to the end-to-end distance of the minicircle were given below each conformation.

The conformations in Figure 5.5a were observed from the ΔLk = -2 and ΔLk = +1 minicircles, in which a topoisomer can develop into more than one conformational type. Presented in order of compactness, with the most compact structure on the left: 'Handcuff' represented the plectonemic structure with a thin interwinding region with two large apical loops, 'Racquet' represented the plectonemic structure with one crossing point and two unequal plectoneme loops, 'Figure-8' represented the plectonemic structure with one crossing point and two equal plectoneme loops, 'Open-figure-8' represented a slightly writhed conformation having either one or no crossing point, and 'Open-Circular' represented the planar structure with almost no writhe.

Another highly compact conformational shape in Figure 5.5b was observed from the ΔLk = -3 minicircle. The 'Needle' conformation consisted of intertwining regions and the sharp bending of a plectoneme tip, which was identified as a kink by our explicitly solvated MD simulation.

5.3.3) Sequence Dependent Effect of DNA Supercoiling

In this section, we examined the sequence dependent effect of positive and negative supercoiling on twist, roll and slide parameters extracted from the MD trajectories of explicitly solvated supercoiled DNA from the Section 5.3.2. To discount the effect of bending at the apices, we picked a 100 bp non-apical segment from ΔLk = -2 and ΔLk = +2 minicircles for analysis. These two segments were equivalent in base sequences (5' – TACCG AGTTC CGACA CTTTC ATTGA GAAAG ATGCC TCAGC TCTGT TACAG GTCAC TAATA CCATC TAAGT AGTTG ATTCA TAGTG ACTGC ATATG TTGTG T – 3') and base positions (101-201).

Local Effects on RY and YR Base Pair Steps

Figure 5.6 shows the average values of twist, roll and slide of the pyrimidine-purine (YR), purine-purine (RR), and purine-pyrimidine (RY) steps calculated from the 100 bp segments of the handcuff-shaped ΔLk = -2 minicircle and the figure-8-shaped ΔLk = +2 minicircle. For both topoisomers, average parameter values generally agreed with the latest ABC simulations of short DNA fragments (103), for example, the distinctively positive roll of the YR steps.

Parameter	Steps	ΔLk = -2		ΔLk = +2		Δ
		Average	S.D.	Average	S.D	
Twist	YR	29.15	6.49	31.19	7.01	-2.04
	RR	32.57	5.82	34.93	5.57	-2.37
	RY	31.17	4.09	32.16	4.17	-0.99
Roll	YR	11.40	6.64	10.21	6.46	1.19
	RR	2.64	5.98	2.39	5.68	0.26
	RY	0.58	5.34	-0.12	5.34	0.70
Slide	YR	-0.41	0.67	-0.05	0.72	-0.36
	RR	-0.73	0.76	-0.34	0.67	-0.40
	RY	-0.86	0.50	-0.69	0.43	-0.17

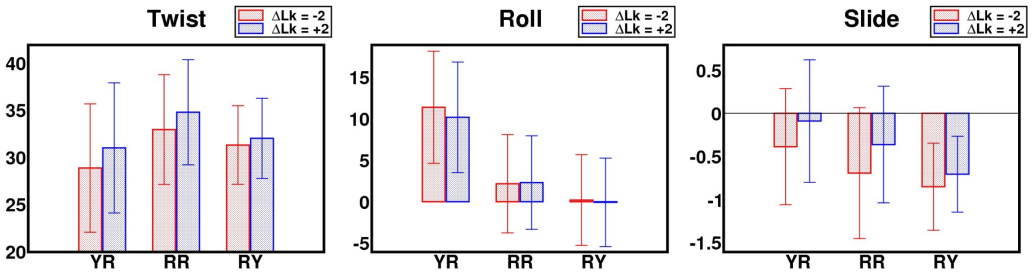


Figure 5.6 Time-average twist, roll and slide over all the pyrimidine-purine (YR), purine-purine (RR) and purine-pyrimidine (RY) steps of a 100 bp non-apical segments from the explicit solvent simulations of the ΔLk = -2 and ΔLk = +2 minicircles shown in Figure 5.7. Difference (Δ) in parameter values between positively and negatively supercoiled DNA at YR, RR and RY steps are indicated.

Differences in twist, roll and slide between ΔLk = -2 and ΔLk = +2 minicircles, shown in Figure 5.6, demonstrated the sequence dependent effect of supercoiling. Lower twist and negative slide values were observed in the negatively supercoiled segment within the ΔLk = -2 minicircle. However, differences in twist and slide between negatively and positively supercoiled segments were smaller in RY steps, as well as the fluctuations, which implied the higher torsional stiffness of the RY steps. The difference in roll was observed only in YR steps where negative supercoiling resulted in the increase of positive roll, which implied that the minor groove is opened at unwound YR steps.

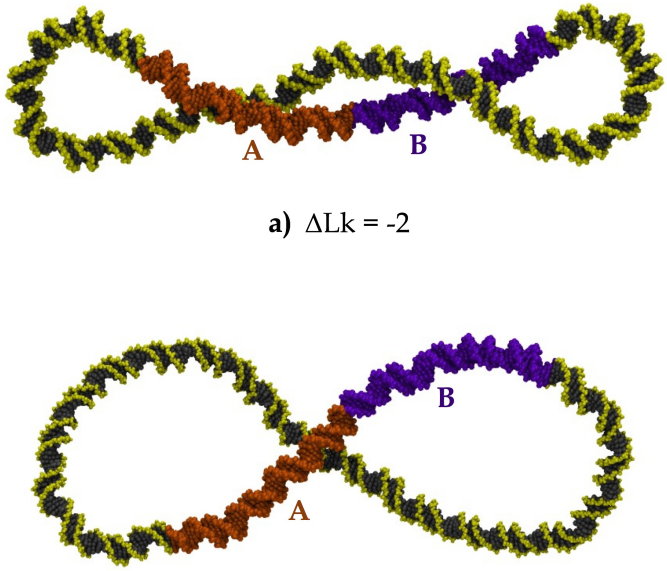


Figure 5.7 MD snapshots from explicit solvent simulations of 336 bp minicircles at a) ΔLk = -2 and b) ΔLk = +2. Segment **A** contains base pair steps 101-150 and segment **B** contains base pair steps 151-200

Collective Effect of YR and RY Base Pair Step Contents

To examine the collective properties of the sequence dependence of the DNA response to superhelical stress, we split the 100 bp non-apical segments into two 50 bp sub-segments A and B (see Figure 5.7), each containing unequal number of YR, RR, and RY base pair steps: sub-segment A (base pair number 101-150) contained 11 YR steps and 10 RY steps and segment B (base pairs 151-200) contained 16 YR steps and 17 RY steps.

As shown in Figure 5.8, time averaged values of the three base pair step parameters were measured for all the base pair steps in each 50 bp sub-segment. In the sub-segment A, we observed highly supercoiling responsive regions (highlighted in green), where the difference in helical parameter values between positively and negatively supercoiled DNA was significant. These regions contained a large number of RR base pair steps but no RY base pair steps. In sub-segment B, which contains a large number of RY steps, the global difference in parameter values (shown in ***Italic***) was smaller than that observed in the sub-segment A, which demonstrated the collective effect of the stiff RY steps.

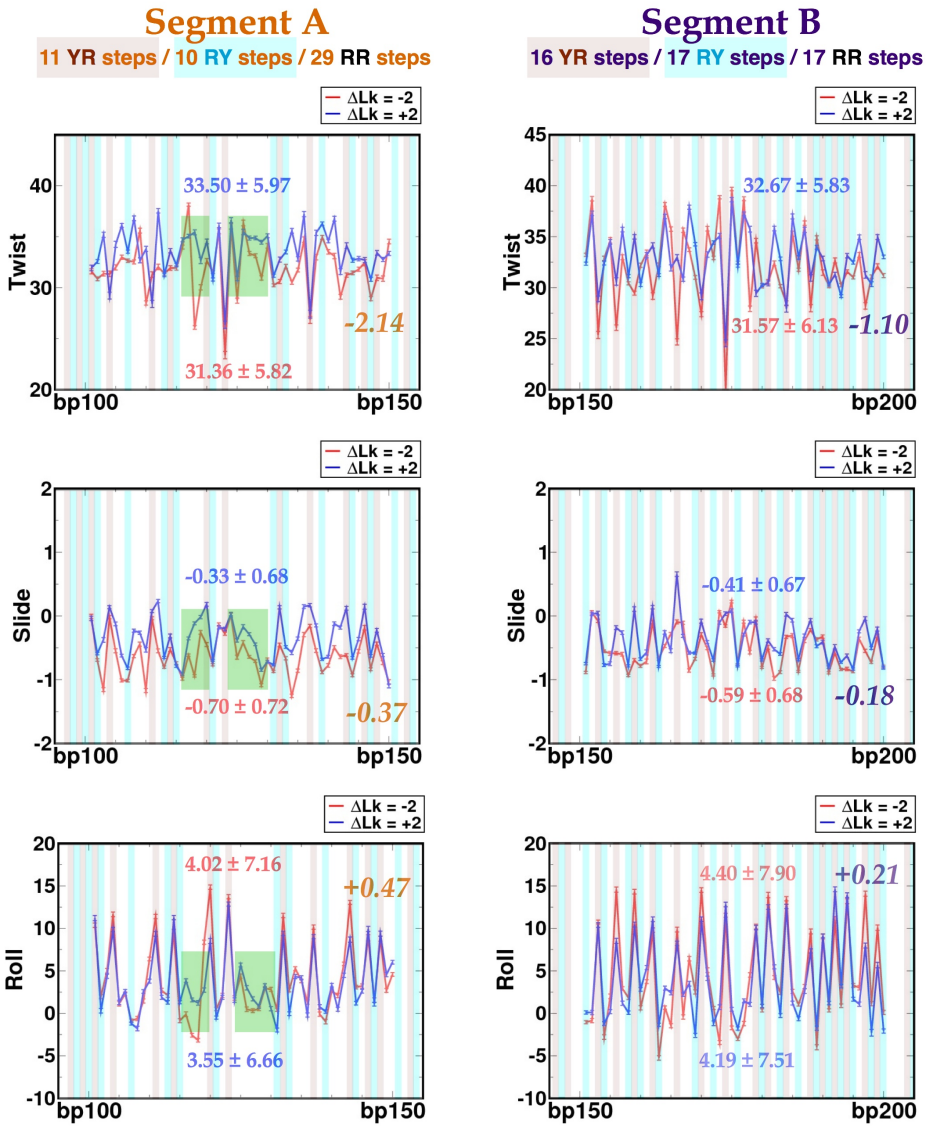


Figure 5.8 Time-averaged twist, roll and slide measured from each base pair step within the segment **A** (left) and segment **B** (right) of a ΔLk = -2 negatively supercoiled minicircle (red) and a ΔLk = +2 positively supercoiled minicircle (blue). The plotted backgrounds were shaded according to base pair step types (grey for pyrimidine-purine, YR ; white for purine-purine, RR ; light-blue for purine pyrimidine, RY). The regions highly responsive to supercoiling, in the absence of the stiff RY steps, are shaded in green. For each parameter measured from segments A and B, averages and standard deviations over all the base pair steps, along with the difference between average parameter values at ΔLk = -2 and ΔLk = +2 were labelled.

5.3.4) DNA-Topoisomerase 1B interaction

Explicitly solvated MD simulations were performed on the human topoisomerase 1B enzyme bound to (i) a 22 bp linear DNA fragment and (ii) a 240 bp negatively supercoiled DNA minicircle. Figure 5.9 shows the

atomistic structure of the negatively supercoiled 240 bp DNA minicircle bound to the subdomain I and the subdomain III of a topoisomerase 1B enzyme after 50 ns of MD simulation. Protein-DNA interactions were quantified by the number of hydrogen bonds between the amino acid residues and the nucleotides. The h-bonds are mostly formed at the negatively charged phosphate groups of the DNA backbones or the ring-atoms with high-electronegativity (e.g. oxygen or nitrogen atoms).

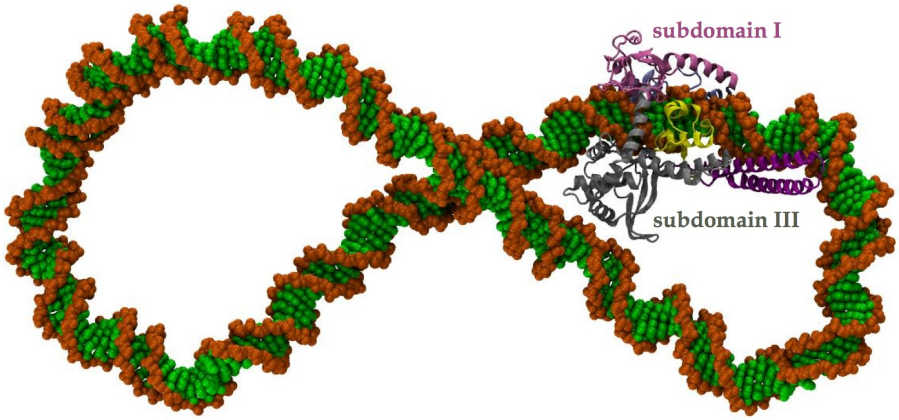


Figure 5.9 An MD snapshot showing a 240 bp DNA minicircles bound to a subdomain I (pink) and a subdomain III (grey) of a topoisomerase 1B. (17)

Figure 5.10 (top) shows the protein-bound catalytic sites of a) linear, torsionally relaxed DNA and b) supercoiled DNA. The supercoiled DNA experienced more binding interactions with the catalytic region, specifically the DNA major groove with the protein subdomain I (shown in grey), and the DNA minor groove with the protein subdomain III (shown in pink). Figure 5.10 (bottom) shows the binding network diagrams of all the hydrogen bonds formed between amino acids and nucleotides with >60% occupancy (see criteria for h-bond in the Section 5.2.3). For a linear DNA fragment with topoisomerase 1B in Figure 5.10a, 20 amino acid residues were found binding to the DNA, which was close to the number observed in the x-ray crystal structure [PDB number: 1A31] (143). In Figure 5.10b, when the same DNA fragment was embedded into a 240 bp negatively supercoiled minicircle, after 50 ns of MD, 14 additional amino acid residues were found to be bound to the DNA, in which 12 of the additional residues were from the subdomain III, mainly interacting with the minor groove.

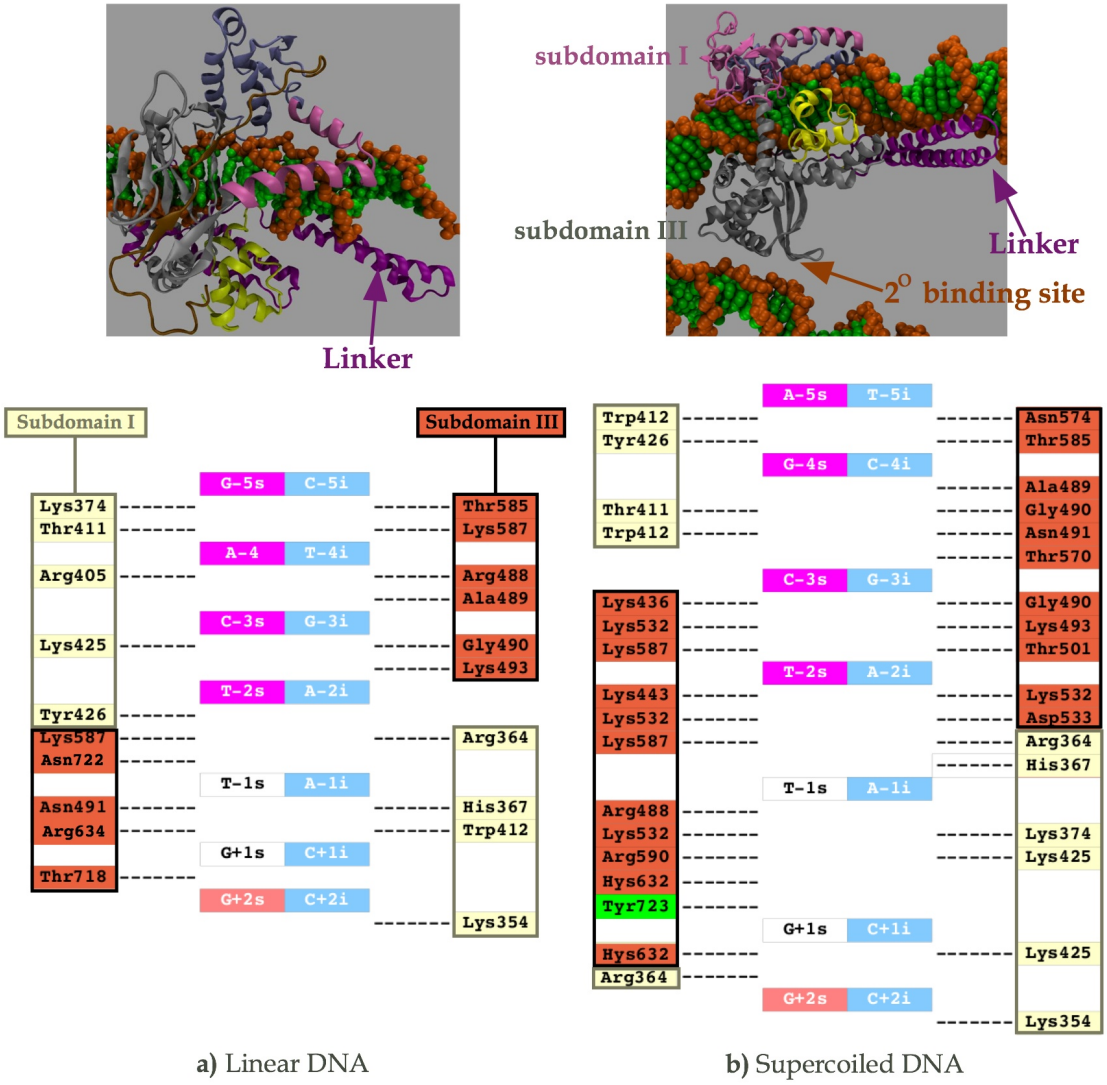


Figure 5.10 (***top***) Snapshots after 50 ns of MD simulations showing the subdomains I and III of an enzyme topoisomerase 1B bound to **a)** 22 bp linear DNA and **b)** the catalytic site, linker domain and a secondary binding site of 240 bp minicircle DNA. (***bottom***) Diagrams showing all the amino acid residues of the subdomain I (yellow) and the subdomain III (red) making hydrogen bonds to phosphate groups between two neighbouring bases or intruding the base step with probability higher than 60% at the catalytic sites of **a)** 22 bp linear DNA (20 h-bonds) and **b)** 240 bp supercoiled minicircle DNA (34 h-bonds). The `Tyr723` residue highlighted in green is the phosphorylated tyrosine binding to the cleaved CA/TG base step consists of `T-1s` and `G+1s` bases.

A possible factor facilitating protein-DNA interactions is the secondary structure of the supercoiled DNA. As described in the earlier section, negative supercoiling opened the minor groove at the YR base pair steps, including the cleaved CA/TG step (T-1s : G+1s) and the regions around it to further interact with the subdomain III. Most strikingly, the DNA curvature induced by DNA supercoiling promoted the additional

binding by the linker domain and the secondary binding at a distal DNA region by subdomain III (indicated by arrows in Figure 5.10).

These findings have demonstrated that DNA-protein interactions are affected by DNA topology, as negative supercoiling induces more amino acid binding at the catalytic and linker domains. Additional protein-DNA interactions at the distal secondary binding site also provides a sample case for the larger interacting supercoiled DNA systems, such as the interactions between the distal pairs of genes, which were observed by the HiC and FISH experiments, as discussed in Chapter 1.

5.4) Conclusions

In this chapter, the secondary and tertiary structures, and the helical parameters of supercoiled DNA have been investigated in atomistic detail. Supercoiled DNA structures were folded in GB/SA implicit solvent in the absence of hydrodynamics (144). Global parameters, writhe and radius of gyration were measured and found to have a general agreement with the results from gel electrophoresis (27). However, simulation artefacts were found in implicitly solvated DNA simulations. Therefore, we chose representative structures from the most populated implicitly solvated conformational clusters (141) and performed explicitly solvated MD in the absence of hydrogen bonding restraints. The trajectory from the explicit MD runs was then compared to the folded supercoiled DNA structures from cryo-ET. MD simulations performed independently from the cryo-ET experiment then present similar types of structures, namely handcuff, racquet, figure-8, open-circle and needle. With the hydrogen bond restraints removed, the DNA was free to respond to the large mechanical stress within $\Delta Lk = -3$ by forming a kink at the tip of a needle-like conformation.

The local base pair step parameters (131) were extracted from both positively and negatively supercoiled DNA in explicit solvent in order to assess the effect of supercoiling on the parameters. The analysis showed that the mechanical response of DNA to superhelical stress was sequence-dependent. Pyrimidine-purine (YR) steps respond to negative supercoiling by increasing the roll values, which signifies minor groove opening, while purine-pyrimidine steps (RY) were stiff and showed less structural deviation in response to supercoiling. These results were used to show how topoisomerase-DNA interactions are affected by DNA supercoiling (17)

(143). In the supercoiled DNA, the cleaved site located at the minor groove of a CA step was opened, facilitating more protein-DNA interactions, and most significantly the more complex topology of the supercoiled DNA provided the opportunity for a secondary DNA binding site for topoisomerase distal to the catalytic region.

In conclusion, by combining atomistic MD simulations with gel electrophoresis and cryo-ET imaging experiments, we have provided the most detailed representation of the atomistic structure and conformational flexibility of 336 bp supercoiled DNA minicircles. Helical parameters extracted from the MD trajectories showed the sequence-dependence of DNA mechanical response to superhelical stress, which provided further understanding of DNA-protein recognition and provide the further insight to the communication between genes located in two or more distal regions within the genome.

Chapter 6
Conclusion and Future Work

6.1) Conclusions

The importance of the primary DNA sequence, which carries the encodes the protein content of the cell, is well known. The role of sequence dependent local secondary to protein/DNA recognition, and therefore gene regulation, is similarly well established. In Chapter 1, the growing evidence that higher order DNA topology is also biologically relevant was presented. However, far less information is available for the sequence dependent structure and dynamics of higher order DNA topologies, because all of the applicable experimental techniques (such as gel electrophoresis, cryo-EM and AFM) do not provide atomic level resolution. This thesis has used atomistic MD simulations, which are discussed in detail in Chapter 2, to study the structure and dynamics of topologically closed DNA minicircles, and has compared the results with lower resolution experimental information.

Tightly looped DNA minicircles of ~100bp < DNA persistence length were investigated in Chapter 3. Minicircles of this small size contained the more subtle mechanical properties beyond that of elastic models. Sequence-dependent flexibility, anisotropy were elucidated by atomistic MD simulations of DNA minicircles. Explicit solvent MD simulations were performed on the ~100bp DNA minicircles of three different sequences. At ΔLk = -0.5, we observed the transient breathing events, each with a pyrimidine base flipped into the major groove. The counter-intuitive results, where the A-T rich FUSE-embedded minicircles were the least probable for breathing, could be explained by the bi-stable YR base pair steps absorbing more superhelical stress than the other steps. Moreover, cooperative bending between two flexible regions also reduce the breathing probability. Compared to the Stress-Induced Duplex Destabilisation (SIDD) model and oxDNA coarse-grained simulations, atomistic MD simulations exhibited some unique characteristics of DNA under high superhelical stress: 1) DNA flexibility is sequence-dependent, so that defects usually occur at the most flexible pyrimidine-purine steps, and 2) DNA anisotropy affects the defect formations. The anisotropic effects became more apparent for the DESIGNED sequence containing the repeated sequence motifs.

In Chapter 4, a new mathematical description for the DNA helical axis was defined in order to obtain writhe values from an atomistic MD trajectory and to address the twist/writhe partitioning problem. Comparative tests were performed with other two helical axis definitions from X3DNA and CURVES+. Local coordinate reference frames from X3DNA overestimated the writhe values as the non-ideal B-DNA structures contain non-zero roll and slide. CURVES+ corrected this error in periodicity by considering the global helical axis. However, writhe calculated from CURVES+ was still slightly underestimated. The new WrLINE method described in Chapter 4 reduced the systematic errors in writhe calculations to a minimum by weight-averaging all the C1′ atoms in a full helical turn. The method correctly produces zero writhe at zero superhelical density.

In Chapter 5, a hybrid implicit/explicit solvent MD simulation protocol was used to observe the writhing of 336 base pair supercoiled DNA minicircles in atomistic detail. Implicit solvent simulations were firstly performed to "fold" the DNA structures rapidly. The writhe and the radius of gyration were calculated and were consistent with the gel electrophoresis results. Selections of the writhed DNA structures were then solvated by explicit water molecules and ions. 3D structures generated by MD were compared with the cryo-ET results for the possible shapes supercoiled DNA could fold into, which provided the most exact atomistic representations of supercoiled DNA accessible by any experimental or computational method to date. Base pair step parameter analysis on the explicitly solvated supercoiled minicircles showed that YR base pair steps responded to negative supercoiling by opening their minor grooves. To study the effect of supercoiling on protein-DNA recognition, simulations of a 240 base pair writhed DNA circle bound to human topoisomerase 1B were also performed. In these simulations, the CA step at the topoisomerase 1B binding site presented an open minor groove, which is likely to promote the phosphorylation reaction to cut and relax the supercoiled DNA.

In conclusion, we have used atomistic simulations to provide information on the atomistic structures of supercoiled DNA minicircles at both the sub-persistence length scale, where extreme helical unwinding and bending induce structural disruptions (Chapter 3), and at the longer length scale where the minicircles can adopt writhed conformations (Chapter 4 and 5). Both sets of calculations were in broad agreement with and complementary to the available data from low resolution experiments. These results improved our understanding of the mechanical response of DNA to

supercoiling, which is ubiquitous in living cells and plays an important role in gene regulation and functional organisation of the genome.

6.2) Future Perspectives

The mechanical properties of supercoiled DNA have been studied by the Computational Biophysics group at the University of Leeds since 2008. In Chapter 3, the work built on Mitchell *et al*. (10), but with an improved version of the forcefield, and two additional DNA sequences were studied. The ongoing forcefield modification of ε and ζ dihedrals is expected to solve the long-standing issue of helical twist error, which will improve the ability of the AMBER forcefield to model closed circular DNA. Therefore, with this modification, it would be desirable to repeat the simulations. However, knowledge from the findings in Chapter 3 will be invaluable for the design of new sequences that test our ability to control the effects of sequence-dependence flexibility and anisotropy when DNA is tightly bent or untwisted. The combination of the use of the newly modified forcefield and the newly designed sequences could make good progress towards a systematic consensus views on the mechanical properties of stressed DNA.

In Chapter 5, due to the dramatically increased supercomputing resources provided by ARCHER, it was possible to explicitly solvate the writhed 336 bp minicircles. However, far more analysis could be done to observe more global properties given increased computational resources in the future. The sequence-dependence of DNA writhing could be analysed, to see which sequences prefer the plectoneme apices or a location at a crossing point with another double-helical segment. The sequence dependence could be quantified by developing a method for calculating the elastic energy from the stiffness matrix, based on the helical parameters calculated from atomistic structures sampled by MD. However, as the tightly bent DNA is proved not to be in elastic regime (107), novel methods will be required to estimate the non-linear free energy associated with defects. The dynamics and distributions of the counterions could also be monitored.

In this thesis, I have focused mainly on negatively supercoiled DNA in great detail, as we were interested in local DNA denaturation into single-stranded regions. It is also intriguing to understand the mechanism of topoisomerase 1B (also able to unwind positively supercoiled DNA) binding to positively supercoiled DNA after simulations and analysis have been performed on a topoisomerase 1B binding to negatively supercoiled DNA. A

high affinity of amino acid binding is expected, in spite of the contraction of the groove sizes due to DNA overwinding.

As the forcefield parameterisation continues to be improved, and with ever growing computational resources, the application of MD simulations to study the mechanical properties of supercoiled DNA promises to be highly fruitful. Using the hybrid implicit/explicit solvent simulation protocols described in Chapter 5, a series of MD simulations of minicircles varying in size and sequence can be performed to further explore the DNA writhing energy landscape and to comprehensively map out the phase diagram of the supercoiling and structural disruptions within minicircles. As longer timescales will be achieved in the future, it might be possible for the supercoiled DNA conformations in explicit solvent MD to explore a sufficiently large region of conformational space that thermodynamic properties can be quantified.

When simulation timescales reach those that are biologically relevant, we will be able to construct model systems for "living DNA", in which we will be able to simulation the process of transcription itself. The simplest case study would be to mimic the twin supercoiling domain model to monitor the transferring of torque generated by the transcription machinery. Additionally, a larger model of supercoiling domains would allow us to examine the time-dependent buckling transition of the DNA, as well as the interactions with proteins, such as topoisomerases, nucleosome core particles (NCPs) and other regulatory proteins. It will be possible to calculate the binding affinity of topoisomerases as a function of supercoiling, and we will be able to visualise the sequence dependent and superhelical density dependent association and dissociation of histones as a function of time at atomistic resolution.

To understand the role of DNA supercoiling in gene regulation and to quantitatively define the implications for biomedical applications, a comprehensive picture of the structure, dynamics and mechanical response of DNA will be required. This thesis starts the detailed study of the mechanical properties of supercoiled DNA by using the MD simulations to unlock the mystery hidden underneath the double-helical DNA, renowned as 'the Molecule of Life'.

Bibliography

1. Travers, A.A., Muskhelishvili, G. and Thompson, J.M.T. (2012) DNA information: from digital code to analogue structure. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, **370**, 2960–2986.

2. Levens, D. and Benham, C.J. (2011) DNA stress and strain, in silico, in vitro and in vivo. *Phys. Biol.*, **8**, 1–9.

3. Fogg, J.M., Randall, G.L., Pettitt, B.M., Sumners, D.W.L., Harris, S. a and Zechiedrich, L. (2012) Bullied no more: when and how DNA shoves proteins around. *Q. Rev. Biophys.*, **45**, 257–299.

4. Koster, D., Crut, A., Shuman, S., Bjornsti, M.-A. and Dekker, N.H. (2010) Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. *Cell*, **142**, 519–530.

5. Watson, J.D. and Crick, F.H. (1953) Molecular Structure of Nucleic Acid. *Nature*, **171**, 737–738.

6. Bates, A.D. and Maxwell, A. (2005) DNA Topology 2nd ed. Oxford University Press.

7. Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M. and Levens, D. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.*, **20**, 396–403.

8. Catanese, D.J., Fogg, J.M., Schrock, D.E., Gilbert, B.E. and Zechiedrich, L. (2012) Supercoiled Minivector DNA resists shear forces associated with gene therapy delivery. *Gene Ther.*, **19**, 94–100.

9. Zhao, N., Fogg, J.M., Zechiedrich, L. and Zu, Y. (2011) Transfection of shRNA-encoding Minivector DNA of a few hundred base pairs to regulate gene expression in lymphoma cells. *Gene Ther.*, **18**, 220–224.

10. Mitchell, J.S., Laughton, C.A. and Harris, S.A. (2011) Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.*, **39**, 3928–3938.

11. Du, Q., Kotlyar, A. and Vologodskii, A. (2008) Kinking the double helix by bending deformation. *Nucleic Acids Res.*, **36**, 1120–1128.

12. Krepl, M., Zgarbová, M., Stadlbauer, P., Otyepka, M., Banáš, P., Koča, J., Cheatham III, T.E., Jurečka, P. and Sponer, J. (2012) Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of

the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput*, **8**, 2506–2520.

13. Fye, R. and Benham, C. (1999) Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA. *Phys. Rev. E*, **59**, 3408–3426.

14. Matek, C., Ouldridge, T.E., Levy, A., Doye, J.P.K. and Louis, A.A. (2012) DNA cruciform arms nucleate through a correlated but asynchronous cooperative mechanism. *J. Phys. Chem. B*, **116**, 11616–11625.

15. Sutthibutpong, T., Harris, S.A. and Noy, A. (2015) Comparison of Molecular Contours for Measuring Writhe in Atomistic Supercoiled DNA. *J. Chem. Theory Comput.*, 10.1021/acs.jctc.5b00035.

16. Irobalieva, R.N., Fogg, J.M., Catanese, D.J., Sutthibutpong, T., Chen, M., Barker, A.K., Ludtke, S.J., Harris, S.A., Schmid, M.F. and Zechiedrich, L. (2015) Structural diversity of supercoiled DNA. *Submitt. to Nat. Commun.*

17. D'Annessa, I., Coletta, A., Sutthibutpong, T., Mitchell, J., Chillemi, G., Harris, S. and Desideri, A. (2014) Simulations of DNA topoisomerase 1B bound to supercoiled DNA reveal changes in the flexibility pattern of the enzyme and a secondary protein-DNA binding site. *Nucleic Acids Res.*, **42**, 9304–9312.

18. Daune, M. (2006) Molecular Biophysics: Structures in motion Oxford University Press, Strasbourg.

19. Calladine, C.R., Drew, H.R., Luisi, B.F. and Travers, A.A. (2001) Understanding DNA: The Molecule & How It Works Third. Elsevier Academic Press, Chennai.

20. Babcock, M.S., Pednault, E.P.D. and Olson, W.K. (1994) Nucleic Acid Structure Analysis: Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable Between Structures. *J. Mol. Biol.*, **237**, 125–156.

21. El Hassan, M.A. and Calladine, C.R. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.*, **251**, 648–664.

22. Mo, Y. (2006) Probing the nature of hydrogen bonds in DNA base pairs. *J. Mol. Model.*, **12**, 665–72.

23. Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., et al. (2010) Amber 11.

24. Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, **95**, 11163–11168.

25. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S. and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.

26. Bednar, J., Furrer, P., Stasiak, A., Dubochet, J., Egelman, E.H. and Bates, A.D. (1994) The Twist, Writhe and Overall Shape of Supercoiled DNA Change During Counterion-induced Transition from a Loosely to a Tightly Interwound Superhelix. *J. Mol. Biol.*, **235**, 825–847.

27. Fogg, J.M., Kolmakova, N., Rees, I., Magonov, S., Hansma, H., Perona, J.J. and Zechiedrich, E.L. (2006) Exploring writhe in supercoiled minicircle DNA. *J. Phys. Condens. Matter*, **18**, S145–S159.

28. Harris, S.A., Laughton, C.A. and Liverpool, T.B. (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res.*, **36**, 21–29.

29. Mitchell, J. and Harris, S. (2013) Thermodynamics of Writhe in DNA Minicircles from Molecular Dynamics Simulations. *Phys. Rev. Lett.*, **110**, 148105.

30. Gilbert, N. and Allan, J. (2014) Supercoiling in DNA and chromatin. *Curr. Opin. Genet. Dev.*, **25**, 15–21.

31. Liu, L.F. and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*, **84**, 7024–7027.

32. Tsao, Y.-P., Wu, H.-Y. and Liu, L.F. (1989) Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell*, **56**, 111–118.

33. Leng, F., Chen, B. and Dunlap, D.D. (2011) Dividing a supercoiled DNA molecule into two independent topological domains. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 19973–19978.

34. Fulcrand, G., Zhi, X. and Leng, F. (2013) Transcription-coupled DNA supercoiling in defined protein systems and in E. coli topA mutant strains. *IUBMB Life*, **65**, 615–622.

35. Van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. a, Dekker, J. and Lander, E.S. (2010) Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.*, **39**, 1–7.

36. Lieberman-Aiden, E., Berkum, N.L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, **326**, 289–294.

37. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.

38. Gall, J.G. and Pardue, M.L. (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. U. S. A.*, **63**, 378–383.

39. Rudkin, G. and Stollar, B.D. (1977) High resolutin detection of DNA-RNA hybrids in situ by indirect immunofluorescent. *Nature*, **265**, 472–473.

40. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.

41. Le, T.B.K., Imakaev, M. V, Mirny, L.A. and Laub, M.T. (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, **342**, 731–734.

42. Chong, S., Chen, C., Ge, H. and Xie, X.S. (2014) Mechanism of Transcriptional Bursting in Bacteria. *Cell*, **158**, 314–326.

43. Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B. and Gilbert, N. (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.*, **20**, 387–395.

44. Sinden, R.R., Carlson, J.O. and Pettijohn, D.E. (1980) Torsional Tension in the DNA Double Helix Measured with Trimethylpsoralen in Living E . coli Cells : Analogous Measurements in Insect and Human Cells. *Cell*, **21**, 773–783.

45. Kouzine, F., Sanford, S., Elisha-Feil, Z. and Levens, D. (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.*, **15**, 146–154.

46. Kouzine, F., Liu, J., Sanford, S., Chung, H.-J. and Levens, D. (2004) The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol.*, **11**, 1092–1100.

47. Cukier, C.D., Hollingworth, D., Martin, S.R., Kelly, G., Díaz-Moreno, I. and Ramos, A. (2010) Molecular basis of FIR-mediated c-myc transcriptional control. *Nat. Struct. Mol. Biol.*, **17**, 1058–1064.

48. Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145–150.

49. Vasudevan, D., Chua, E.Y.D. and Davey, C.A. (2010) Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.*, **403**, 1–10.

50. Wu, B., Mohideen, K., Vasudevan, D. and Davey, C.A. (2010) Structural Insight into the Sequence Dependence of Nucleosome Positioning. *Structure*, **18**, 528–536.

51. Volle, C.B. and Delaney, S. (2012) CAG/CTG Repeats Alter the Affinity for the Histone Core and the Positioning of DNA in the Nucleosome. *Biochemistry*.

52. Volle, C.B. and Delaney, S. (2013) AGG/CCT interruptions affect nucleosome formation and positioning of healthy-length CGG/CCG triplet repeats. *BMC Biochem.*, **14**, 1–12.

53. Wang, D., Ulyanov, N.B. and Zhurkin, V.B. (2010) Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove. *J. Biomol. Struct. Dyn.*, **27**, 843–859.

54. Olson, W.K. and Zhurkin, V.B. (2011) Working the kinks out of nucleosomal DNA. *Curr. Opin. Struct. Biol.*, **21**, 348–357.

55. Napoli, A. a, Lawson, C.L., Ebright, R.H. and Berman, H.M. (2006) Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: recognition of pyrimidine-purine and purine-purine steps. *J. Mol. Biol.*, **357**, 173–183.

56. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. and Shakked, Z. (2001) DNA bending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8490–8495.

57. Geršata, T., Špačková, N., Jurečka, P., Zgarbová, M., Šponer, J. and Lankaš, F. (2014) Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning. *Nucleic Acids Res.*, **42**, 7383–7394.

58. Haran, T.E. and Mohanty, U. (2009) The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.*, **42**, 41–81.

59. Haeusler, A.R., Goodson, K.A., Lillian, T.D., Wang, X., Goyal, S., Perkins, N.C. and Kahn, J.D. (2012) FRET studies of a landscape of Lac repressor-mediated DNA loops. *Nucleic Acids Res.*, **40**, 4432–4445.

60. Peters, J.P. and Maher, L.J. (2010) DNA curvature and flexibility in vitro and in vivo. *Q. Rev. Biophys.*, **43**, 23–63.

61. Mathew-fenn, R.S., Das, R. and Harbury, P.A.B. (2009) Remeasuring the Double Helix. **322**, 446–450.

62. Mastroianni, A.J., Sivak, D., Geissler, P.L. and Alivisatos, P. (2009) Probing the conformational distributions of subpersistence length DNA. *Biophys. J.*, **97**, 1408–1417.

63. Fields, A.P., Meyer, E.A. and Cohen, A.E. (2013) Euler buckling and nonlinear kinking of double-stranded DNA. *Nucleic Acids Res.*, **41**, 9881–9890.

64. Scipioni, A., Zuccheri, G., Anselmi, C., Bergia, A., Santis, P. De, Chimica, D. and Moruzzi, B.G. (2002) Sequence-Dependent DNA Dynamics by Scanning Force Microscopy Time-Resolved Imaging. *J. Chem. Biol.*, **9**, 1315–1321.

65. Scipioni, A., Anselmi, C., Zuccheri, G., Samori, B. and Santis, P. De (2002) Sequence-Dependent DNA Curvature and Flexibility from Scanning Force Microscopy Images. **83**, 2408–2418.

66. Milani, P., Marilley, M. and Rocca-Serra, J. (2007) TBP binding capacity of the TATA box is associated with specific structural properties: AFM study of the IL-2R alpha gene promoter. *Biochimie*, **89**, 528–533.

67. Wiggins, P.A., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A., Phillips, R., Widom, J., Dekker, C. and Nelson, P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.*, **1**, 137–141.

68. Billingsley, D.J., Bonass, W.A., Crampton, N., Kirkham, J. and Thomson, N.H. (2012) Single-molecule studies of DNA transcription using atomic force microscopy. *Phys. Biol.*, **9**, 1–15.

69. Demurtas, D., Amzallag, A., Rawdon, E.J., Maddocks, J.H., Dubochet, J. and Stasiak, A. (2009) Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucleic Acids Res.*, **37**, 2882–2893.

70. Cloutier, T.E. and Widom, J. (2004) Spontaneous sharp bending of double-stranded DNA. *Mol. Cell*, **14**, 355–362.

71. Lionberger, T.A., Demurtas, D., Witz, G., Dorier, J., Lillian, T., Meyhöfer, E. and Stasiak, A. (2011) Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res.*, **39**, 9820–9832.

72. Johnson, S., Chen, Y.J. and Phillips, R. (2013) Poly(dA:dT)-Rich DNAs Are Highly Flexible in the Context of DNA Looping. *PLoS One*, **8**, e75799.

73. Salerno, D., Tempestini, A., Mai, I., Brogioli, D., Ziano, R., Cassina, V. and Mantegazza, F. (2012) Single-Molecule Study of the DNA Denaturation Phase Transition in the Force-Torsion Space. *Phys. Rev. Lett.*, **109**, 118303.

74. Randall, G.L., Zechiedrich, L. and Pettitt, B.M. (2009) In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res.*, **37**, 5568–77.

75. Matek, C., Ouldridge, T.E., Doye, J.P.K. and Louis, A.A. (2014) Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA. *Nat. Sci. Reports*, **5**, srep07655.

76. Shimada, J. and Yamakawa, H. (1984) Ring-Closure Probabilities for Twisted Wormlike Chains. Application t o DNA. *Macromolecules*, **17**, 689–698.

77. Du, Q., Vologodskaia, M., Kuhn, H., Frank-Kamenetskii, M. and Vologodskii, A. (2005) Gapped DNA and cyclization of short DNA fragments. *Biophys. J.*, **88**, 4137–4145.

78. Vafabakhsh, R. and Ha, T. (2012) Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science*, **337**, 1097–1101.

79. Le, T.T. and Kim, H.D. (2013) Measuring shape-dependent looping probability of DNA. *Biophys. J.*, **104**, 2068–2076.

80. Lankas, F., Lavery, R. and Maddocks, J.H. (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure*, **14**, 1527–1534.

81. Anonymous (2009) Gene therapy deserves a fresh chance. *Nature*, **461**, 1173.

82. Shibata, Y., Kumar, P., Layer, R., Willcox, S., Gagan, J.R., Griffith, J.D. and Dutta, A. (2012) Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science*, **336**, 82–86.

83. Pardi, a, Hare, D.R. and Wang, C. (1988) Determination of DNA structures by NMR and distance geometry techniques: a computer simulation. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 8785–8789.

84. Campagne, S., Gervais, V. and Milon, A. (2011) Nuclear magnetic resonance analysis of protein − DNA interactions Nuclear magnetic resonance analysis of protein – DNA interactions. *J. R. Soc. Interface*, doi: 10.1098/rsif.2010.0543.

85. Galindo-Murillo, R., Roe, D.R. and Cheatham, T.E. (2014) On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nat. Commun.*, **5**, 5152.

86. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. and Shaw, D.E. (2012) Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.*, **41**, 429–452.

87. Leach, A.R. (2001) Molecular Modelling: Principles and Applications.

88. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

89. Verlet, L. (1967) Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, **159**, 98–103.

90. Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. and Berman, H.M. (1996) Geometric Parameters in Nucleic Acids : Nitrogenous Bases. *J. Am. Chem. Soc.*, **118**, 509–518.

91. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., et al. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.

92. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.

93. Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., et al. (2010) Amber 11. *Univ. California, San Fr.*

94. Zimmermann, K. (1991) ORAL: All purpose molecular mechanics simulator and energy minimizer. *J. Comput. Chem.*, **12**, 310–319.

95. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, **23**, 327–341.

96. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.

97. Weiser, J., Shenkin, P.S. and Still, W.C. (1999) Approximate Atomic Surfaces from Linear Combinations of Pairwise. *J. Comput. Chem.*, **20**, 217–230.

98. Hawkins, G.D., Cramer, C.J. and Truhlar, D.G. (1996) Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.*, **100**, 19824–19839.

99. Zgarbova, M., Luque, F.J., Jir, S., Cheatham, T.E., Otyepka, M. and Jurec, P. (2013) Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.

100. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.

101. Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Cheatham III, T.E., Sponer, J. and Otyepka, M. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations : Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.*, **6**, 3836–3849.

102. Beveridge, D.L., Cheatham, T.E. and Mezei, M. (2012) The ABCs of molecular dynamics simulations on B-DNA, circa 2012. *J. Biosci.*, **37**, 379–397.

103. Pasi, M., Maddocks, J.H., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T., Dans, P.D., Jayaram, B., Lankas, F., Laughton, C., et al. (2014) μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.

104. Dans, P.D., Pérez, A., Faustino, I., Lavery, R. and Orozco, M. (2012) Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.*, **40**, 10668–10678.

105. Dans, P.D., Faustino, I., Battistini, F., Zakrzewska, K., Lavery, R. and Orozco, M. (2014) Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.

106. Deniz, Ö., Flores, O., Battistini, F., Pérez, A., Soler-López, M. and Orozco, M. (2011) Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics*, **12**, 489.

107. Curuksu, J., Zacharias, M., Lavery, R. and Zakrzewska, K. (2009) Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.*, **37**, 3766–3773.

108. Kannan, S., Kohlhoff, K. and Zacharias, M. (2006) B-DNA under stress: over- and untwisting of DNA during molecular dynamics simulations. *Biophys. J.*, **91**, 2956–2965.

109. Wereszczynski, J. and Andricioaei, I. (2006) On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 16200–16205.

110. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M. and Vologodskii, A. (2005) Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 5397–5402.

111. Šulc, P., Romano, F., Ouldridge, T.E., Rovigatti, L., Doye, J.P.K. and Louis, A.A. (2012) Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.*, **137**, 135101.

112. Wang, Q. and Pettitt, B.M. (2014) Modeling DNA thermodynamics under torsional stress. *Biophys. J.*, **106**, 1182–1193.

113. Kowalski, D., Natale, D.A. and Eddy, M.J. (1988) Stable DNA unwinding, not 'breathing,' accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 9464–9468.

114. Ouldridge, T.E., Louis, A.A. and Doye, J.P.K. (2011) Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, **134**, 085101.

115. SantaLucia, J. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

116. Ouldridge, T.E., Hoare, R.L., Louis, A.A., Doye, J.P.K., Bath, J. and Turberfield, A.J. (2013) Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA Walker. *ACS. Nano.*, **7**, 2479–2490.

117. Romano, F., Hudson, A., Doye, J.P.K., Ouldridge, T.E. and Louis, A.A. (2012) The effect of topology on the structure and free energy landscape of DNA kissing complexes. *J. Chem. Phys.*, **136**, 215102.

118. Savelyev, A. and Papoian, G.A. (2010) Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 20340–20345.

119. Dans, P.D., Zeida, A., MacHado, M.R. and Pantano, S. (2010) A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J. Chem. Theory Comput.*, **6**, 1711–1725.

120. Demille, R.C., Cheatham, T.E. and Molinero, V. (2011) A Coarse-Grained Model For Explicit Solvation Of DNA By Water And Ions. *J. Phys. Chem. B*, **115**, 132–142.

121. Ramachandran, A., Guo, Q., Iqbal, S.M. and Liu, Y. (2011) Coarse-grained molecular dynamics simulation of DNA translocation in chemically modified nanopores. *J. Phys. Chem. B*, **115**, 6138–6148.

122. Korolev, N., Luo, D., Lyubartsev, A.P. and Nordenskiöld, L. (2014) A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers (Basel).*, **6**, 1655–1675.

123. Freeman, G.S., Hinckley, D.M., Lequieu, J.P., Whitmer, J.K. and de Pablo, J.J. (2014) Coarse-grained Modeling of DNA Curvature. *J. Chem. Phys.*, **141**, 165103.

124. Marko, J.F. and Siggia, E.D. (1995) Stretching DNA. *Macromolecules*, **28**, 8759–8770.

125. Bouchiat, C., Wang, M.D., Allemand, J., Strick, T., Block, S.M. and Croquette, V. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, **76**, 409–413.

126. Swigon, D., Coleman, B.D. and Tobias, I. (1998) The elastic rod model for DNA and its application to the tertiary structure of DNA minicircles in mononucleosomes. *Biophys. J.*, **74**, 2515–2530.

127. Liverpool, T.B., Harris, S.A. and Laughton, C.A. (2008) Supercoiling and denaturation of DNA loops. *Phys. Rev. Lett.*, **100**, 238103.

128. Zheng, X. and Vologodskii, A. (2009) Theoretical analysis of disruptions in DNA minicircles. *Biophys. J.*, **96**, 1341–1349.

129. Giovan, S.M., Scharein, R.G., Hanke, A. and Levene, S.D. (2014) Free-energy calculations for semi-flexible macromolecules: Applications to DNA knotting and looping. *J. Chem. Phys.*, **141**, 174902.

130. Wang, H., Kaloper, M. and Benham, C.J. (2006) SIDDBASE: a database containing the stress-induced DNA duplex destabilization (SIDD) profiles of complete microbial genomes. *Nucleic Acids Res.*, **34**, D373–D378.

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

132. Blanchet, C., Pasi, M., Zakrzewska, K. and Lavery, R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.

133. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T., Dixit, S., Jayaram, B., Lankas, F., Laughton, C., et al. (2009) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.

134. Mitchell, J.S. (2011) A Molecular Dynamics Study of Supercoiled DNA Minicircles. University of Leeds, Leeds.

135. Johnson, S., Linden, M. and Phillips, R. (2012) Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.*, **40**, 7728–7738.

136. Bates, A.D., Noy, A., Piperakis, M.M., Harris, S.A. and Maxwell, A. (2013) Small DNA circles as probes of DNA topology. *Biochem. Soc. Trans.*, **41**, 565–570.

137. Fuller, F.B. (1971) The Writhing Number of a Space Curve. *Proc. Natl. Acad. Sci.*, **68**, 815–819.

138. White, J.H. (1969) Self-linking and the gauss integral in higher dimensions. *Am. J. Math.*, **91**, 693–728.

139. Lu, X.-J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.

140. Redinbo, M.R. (1998) Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA. *Science*, **279**, 1504–1513.

141. Shao, J., Tanner, S.W., Thompson, N. and Cheatham, T.E. (2007) Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.*, **3**, 2312–2334.

142. Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J. and Hol, W.G.J. (1998) Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA. *Science*, **279**, 1504–1513.

143. Stewart, L. (1998) A Model for the Mechanism of Human Topoisomerase I. *Science*, **279**, 1534–1541.

144. Tsui, V. and Case, D.A. (2000) Theory and Applications of the Generalized Born Solvation Model in. *Biopolymers*, **56**, 275–291.

Appendix A
DNA Sequences

A.1) Sequences of the small DNA minicircles in Chapter 3

A.1.1) 102 bp minicircles at ΔLk = -0.5 (Lk = 9)

RANDOM sequence:

TTGCGGCAGT TAATCGAACA AGACCCGTGC AATGCTATCG ACATCAAGGC CTATCGTATT
ACGGGGTTGG GAGTCAATGG GTTCAGGATG CAGGTGAGGA TC

FUSE-embedded sequence:

TATATTTAAT ATATAATGTA TATTCCCTCG GGATTTTTTA TTTTGTGTTA TTTATCGCTA
TTACGGGGTT GGGAGTCAAT GGGTTCAGGA TGCAGGTGAA TC

DESIGNED sequence:

^{1}CGCGCGCGCGCGCGC ^{16}**TATATATATATA** ^{28}CGCGCGCGCGCGC ^{41}**ATATATATATAT**
^{53}CGCGCGCGCGCGC ^{66}**AAAAAAAAAAAA** ^{78}CGCGCGCGCGCGC ^{91}**CACACACACACA**

A.1.2) 108 bp minicircles at ΔLk = 0 (Lk = 10) and ΔLk = -1 (Lk = 9)

RANDOM sequence:

TTTGCGGCAG TTAATCGAAC AAGACCCGTG CAATGCTATC GACATCAAGG CCTATCGCTA
TTACGGGGTT GGGAGTCAAT GGGTTCAGGA TGCAGGTGAG GATATATC

FUSE-embedded sequence:

TATATTTAAT ATATAATGTA TATTCCCTCG GGATTTTTTA TTTTGTGTTA TTTATCGCTA
TTACGGGGTT GGGAGTCAAT GGGTTCAGGA TGCAGGTGAG GATATATC

DESIGNED sequence:

^{1}CGCGCGCGCGCGCGC ^{16}**TATATATATATA** ^{28}CGCGCGCGCGCGCGC ^{43}**ATATATATATAT**
^{55}CGCGCGCGCGCGCGC ^{70}**AAAAAAAAAAAA** ^{82}CGCGCGCGCGCGCGC ^{97}**CACACACACACA**

A.2) Sequences of the larger DNA minicircles in Chapter 4,5

A.2.1) 260 bp minicircles in Chapter 4

TTTATACTAA CTTGGAGCGA AACGGGAAGG TAAAAAGACC AAAAAGTTGT TTTTAATACC
TTTAAGACGC TCTCCCTTAT GCGGACTCCT GCGATATCGC CCTCGGCTCT GTTACAGGTC
ACTAATACCA TCTAAGTTAG TTGATTCATA GTGACTGCCA TATGTTGTGT TTTACAGGTA
TTATGTAGTC TGTTTTTTAT GCAAAATCCT AATTTAATAT ATTGATATTT ATATCATTTT
ACGTTTCCTC GTTCAGCTTT

A.2.2) 240 bp minicircles in Chapter 5

AGACTATACG AAGTTATTAG GTCTGAAGAG GAGTTTACGT CCAGCCAAGC TTCAAGTCGG
TAGTTTATCA CAGTTAAATT GCTAACGCAG TCAGGCACCG TGTATGAAAT CTAACAATGC
GCTCACCCGT TCTCGGAGAC TTGTCCGACC GCTTTGGCCG CCGCCCAGTC CTGCTGCCTT
CGCTACTTGG AGCCACTATC GACTACGCGA TCATGGCGAC CACACCCGTC CTGTGGATCC

A.2.3) 336 bp minicircles in Chapter 4 and 5

AAACGCGCGA GGCAGCTGTA TGGCATGAAA GAGTTCTTCC CGGAAAACGC GGTGGAATAT
TTCGTTTCCT ACTACGACTA CTATCAGCCG GAAGCCTATG TACCGAGTTC CGACACTTTC
ATTGAGAAAG ATGCCTCAGC TCTGTTACAG GTCACTAATA CCATCTAAGT AGTTGATTCA
TAGTGACTGC ATATGTTGTG TTTTACAGTA TTATGTAGTC TGTTTTTTAT GCAAAATCTA
ATTTAATATA TTGATATTTA TATCATTTTA CGTTTCTCGT TCAGCTTTTT TATACTAACT
TGAGCGAAAC GGGAAGGGTT TTCACCGATA TCACCG

Appendix B
Detailed Simulation Protocols

B.1) Explicit solvent simulations protocols and AMBER input files

Stages	Restraints	
Mininisation 1	k = 500.0 kCal/mol/Å^2	10000 cycles
Mininisation 2	k = 50.0 kCal/mol/Å^2	10000 cycles
Mininisation 3	k = 25.0 kCal/mol/Å^2	10000 cycles
Mininisation 4	No restraints	10000 cycles
Equilibration 1	k = 500.0 kCal/mol/Å^2, T = 100K	10 ps
Equilibration 2	k = 50.0 kCal/mol/Å^2, T=100-300K	10 ps
Equilibration 3	k = 50.0 kCal/mol/Å^2, T = 300K	10 ps
Equilibration 4	k = 25.0 kCal/mol/Å^2, T = 300K	10 ps
Equilibration 5	k = 10.0 kCal/mol/Å^2, T = 300K	10 ps
Equilibration 6	k = 5.0 kCal/mol/Å^2, T = 300K	10 ps
Equilibration 7	k = 2.5 kCal/mol/Å^2, T = 300K	10 ps
Equilibration 8	k = 1.0 kCal/mol/Å^2, T = 300K	10 ps
Productive Runs	(none)	10-100 ns

```
MINIMIZATION

::::::::::::::
min1.in
::::::::::::::
initial minimization solvent + ions
 &cntrl
  imin   = 1,
  maxcyc = 1000,
  ncyc   = 500,
  ntb    = 1,
  ntr    = 1,
  cut    = 12.0
 /
Hold the DNA fixed
500.0
RES 1 216
END
END
```

```
::::::::::::::
min2.in
::::::::::::::
initial minimization solvent + ions
 &cntrl
  imin   = 1,
  maxcyc = 10000,
  ncyc   = 5000,
  ntb    = 1,
  ntr    = 1,
  cut    = 12.0
 /
Hold the DNA fixed
50.0
RES 1 216
END
END
```

```
::::::::::::::
min3.in
::::::::::::::
initial minimization solvent + ions
 &cntrl
  imin   = 1,
  maxcyc = 10000,
  ncyc   = 5000,
  ntb    = 1,
  ntr    = 1,
  cut    = 12.0
 /
Hold the DNA fixed
25.0
RES 1 86
END
END
```

```
::::::::::::::
min4.in
::::::::::::::
initial minimization whole system
 &cntrl
  imin   = 1,
  maxcyc = 10000,
  ncyc   = 5000,
  ntb    = 2,
  ntp    = 1
  ntr    = 0,
  cut    = 12.0
 /
```

MOLECULAR DYNAMICS

```
::::::::::::::
md1.in
::::::::::::::
MD on the water and ions about the DNA
 &cntrl
       ntf=2, ntb=2, cut=12.0,
       nstlim=5000, dt=0.002,
       temp0=100.0, ntt=1,
       ntp=1,
       ntc=2,
       ntr=1,
 &end
Restrain DNA
500.0
RES 1 208
END
END
```

```
::::::::::::::
md2.in
::::::::::::::
MD on water and ions around DNA raising T
to 300K
 &cntrl
       irest=1, ntx=7,
       ntf=2, ntb=2, cut=12.0,
       ntr=1,
       nstlim=5000, dt=0.002,
       tempi=100.0 temp0=300.0, ntt=1,
       ntp=1,
       ntc=2,
 &end
Restrain DNA
50.0
RES 1 208
END
END
```

```
::::::::::::::
md3.in
::::::::::::::
MD on the water and K about DNA
 &cntrl
       irest=1, ntx=7,
       ntf=2, ntb=2, cut=12.0,
       ntr=1,
       nstlim=5000, dt=0.002,
       temp0=300.0, ntt=1,
       ntp=1,
       ntc=2,
 &end
Restrain DNA
50.0
RES 1 208
END
END
```

```
::::::::::::::
md4.in
::::::::::::::
MD on the water and K about DNA
 &cntrl
       irest=1, ntx=7,
       ntf=2, ntb=2, cut=12.0,
       ntr=1,
       nstlim=5000, dt=0.002,
       temp0=300.0, ntt=1,
       ntp=1,
       ntc=2,
 &end
Restrain DNA
25.0
RES 1 208
END
END
```

```
::::::::::::::
md5.in
::::::::::::::
MD on the water and K about DNA
 &cntrl
       irest=1, ntx=7,
       ntf=2, ntb=2, cut=12.0,
       ntr=1,
       nstlim=5000, dt=0.002,
       temp0=300.0, ntt=1,
       ntp=1,
       ntc=2,
 &end
Restrain DNA
10.0
RES 1 208
END
END
```

```
::::::::::::::
md6.in
::::::::::::::
MD on the water and Na about DNA
 &cntrl
       irest=1, ntx=7,
       ntf=2, ntb=2, cut=12.0,
       ntr=1,
       nstlim=5000, dt=0.002,
       temp0=300.0, ntt=1,
       ntp=1,
       ntc=2,
  &end
Restrain DNA
5.0
RES 1 208
END
END
```

```
::::::::::::::
md7.in
::::::::::::::
MD on the water and Na about DNA
 &cntrl
        irest=1, ntx=7,
        ntf=2, ntb=2, cut=12.0,
        ntr=1,
        nstlim=5000, dt=0.002,
        temp0=300.0, ntt=1,
        ntp=1,
        ntc=2,
 &end
Restrain DNA
2.5
RES 1 208
END
END
```

```
::::::::::::::
md8.in
::::::::::::::
MD on the water and Na about DNA
 &cntrl
        irest=1, ntx=7,
        ntf=2, ntb=2, cut=12.0,
        ntr=1,
        nstlim=5000, dt=0.002,
        temp0=300.0, ntt=1,
        ntp=1,
        ntc=2,
 &end
Restrain DNA
1.0
RES 1 208
END
END
```

```
::::::::::::::
md9.in
::::::::::::::
unrestrained MD
 &cntrl
      irest=1, ntx=7,
        ntf=2, ntb=2, cut=12.0,
        nstlim=50000000, dt=0.002,
        temp0=300.0, ntt=1, iwrap=1,
        ntp=1,
        ntc=2,
      ntwx=500,
 /
```

B.2) Implicit solvent simulations protocols and AMBER input files

Stages	Restraints	
Mininisation 1	k = 50.0 kCal/mol/Å^2	10000 cycles
Mininisation 2	No restraints	10000 cycles
Equilibration 1	k = 50.0 kCal/mol/Å^2, T=100-300K	10 ps
Equilibration 2	k = 10.0 kCal/mol/Å^2, T = 300K	100 ps
Equilibration 3	k = 1.0 kCal/mol/Å^2, T = 300K	200 ps
Productive Runs	***NMR***, k = 1.0 kCal/mol/Å^2	10-100 ns

MINIMIZATION

```
::::::::::::::
min1.in
::::::::::::::
initial minimization solvent + ions
 &cntrl
  imin=1,
  maxcyc=1000,
  ncyc=500,
  ntb=0,
  ntr=1,
  igb=1,
  cut=100.0,
  restraint_wt=500.0,
  restraintmask=':1-672'
 /
```

```
::::::::::::::
min2.in
::::::::::::::
initial minimization whole system
 &cntrl
  imin=1,
  maxcyc=10000,
  ncyc=2500,
  ntb=0,
  ntp=0,
  ntr=0,
  igb=1,
  cut=100.0,
 /
```

MD SIMULATIONS

```
::::::::::::::
md1.in
::::::::::::::
MD on the water and ions about the DNA
 &cntrl
        ntf=2, ntc=2, ntb=0, cut=100.0,
        nstlim=5000, dt=0.002,
        tempi=100.0, temp0=300.0, ntt=1,
        imin=0, irest=0, ntx=1,
        igb=1, gbsa=1, saltcon=0.1,
        ntr=1,
        restraint_wt=5.0,
        restraintmask=':1-672'
 &end
```

```
::::::::::::::
md2.in
::::::::::::::
MD on the water and ions about the DNA
 &cntrl
        ntf=2, ntc=2, ntb=0, cut=100.0,
        nstlim=50000, dt=0.002, ntpr=500,
        temp0=300.0, ntt=1,
        imin=0, irest=1, ntx=5,
        igb=1, gbsa=1, saltcon=0.1,
        ntr=1,
        restraint_wt=10.0,
        restraintmask=':1-672'
 &end
```

```
::::::::::::::
md3.in
::::::::::::::
MD on the water and ions about the DNA
 &cntrl
        ntf=2, ntc=2, ntb=0, cut=100.0,
        nstlim=100000, dt=0.002,
ntpr=500,
        temp0=300.0, ntt=1,
        imin=0, irest=1, ntx=5,
        igb=1, gbsa=1, saltcon=0.1,
        ntr=1,
        restraint_wt=1.0,
        restraintmask=':1-672'
 &end
```

```
::::::::::::::
md4_RST.in
::::::::::::::
MD on the water and ions about the DNA
 &cntrl
        ntf=2, ntc=2, ntb=0, cut=100.0,
        nstlim=50000000, dt=0.002,
ntpr=500, ntwx=500, nscm = 1000,
        temp0=300.0, ntt=1,
        imin=0, irest=1, ntx=5,
      igb=1, gbsa=1, saltcon=0.1,
        ntr=0, nmropt=1, pencut=-0.001,
 &end
 &wt type='REST', istep1=0,
istep2=500000, value1=0.05, value2=0.05,
/
 &wt type='REST', istep1=500001,
istep2=50000000, value1=0.05,
value2=0.05, /
 &wt type='END' /
LISTOUT=POUT
DISANG=RST
```

Appendix C Defects in ΔLk = -1 DESIGNED Minicircles

NAME	BLOCK	POSITION	REGION(S)
AA-Major_r1	MIN / MIN	25 / 85	TA / CG
AA-Major_r2	MIN / MAJ	21,22 / 70-77	TA / AA
AA-Major_r3	MAJ	65,66,67	CG
AA-Major_r4	MIN	106, 107	CA
AA-Major_r5	MIN	79,80,81	CG / AA
AA-Major_r6	MAJ	55	CG
AA-Major_r7	MIN	19,20	TA
AA-Major_r8	MAJ	84,85	CG
AA-Major_r9	MAJ	26,27	TA
AA-Major_r10	MIN	79,80,81	CG / AA
AA-Major_r11	MIN	44,45,46	AT
AA-Major_r12	MIN / MIN	44 / 103, 104	AT / CA
AA-Major_r13	MAJ	63,64	CG
AA-Major_r14	MIN	45	TA
AA-Major_r15	MAJ / MIN	49-52 / 94,95	TA / CA
TA-Major_r1	MAJ	57,58	CG
TA-Major_r2	MAJ	68,69	CG
TA-Major_r3	MIN	100	CA
TA-Major_r4	MAJ	7,8	CG
TA-Major_r5	MAJ	77,78,79,80	AA
TA-Major_r6	MAJ	80	AA
TA-Major_r7	MIN	50,51	AT
TA-Major_r8	MAJ / MAJ	8,9 / 68	CG / CG
TA-Major_r9	MIN / MIN	63,64 / 101	CG / CA
TA-Major_r10	MAJ	10	CG
TA-Major_r11	MAJ	79,80,81,82	AA / CG
TA-Major_r12	MAJ	9,10	CG
TA-Major_r13	MIN	108, 1	CA / CG
TA-Major_r14	MAJ	92,93	CG / CA
TA-Major_r15	MAJ	44-52	AT
CA-Minor_r1	MIN	30,31	CG
CA-Minor_r2	MAJ	61,62,63	CG / AA
CA-Minor_r3	MIN	102, 103	CA
CA-Minor_r4	MAJ	35,36,37	CG
CA-Minor_r5	MAJ	25,26,27	AT
CA-Minor_r6	MIN	18,19,20	CA
CA-Minor_r7	MIN	104	CA
CA-Minor_r8	MAJ	106, 107, 108	CA
CA-Minor_r9	MAJ / MAJ	36,37 / 85	CG / CG
CA-Minor_r10	MAJ	99, 100, 101	CA
CA-Minor_r11	MAJ	98,99,100,101	CA
CA-Minor_r12	MAJ	87	CG
CA-Minor_r13	MIN / MIN	45,46/105,106	AT / CA
CA-Minor_r14	MAJ	86 87	CG
CA-Minor_r15	MAJ	61,62,63	CG