Structure and Dynamics of DNA Mini-Circles and Super Coiled Plasmids a Project Submitted to the DST-Inspire Division as a Part of the Research
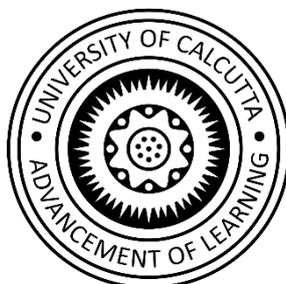
Submitted by

ASHIQ MAHAMMAD

IVR NO-**202100026966**

C.U Reg.no – **115-1115-0547-21**

Under the guidance of
DR. ANSUMAN LAHIRI
PROFESSOR

**DEPARTMENT OF BIOPHYSICS, MOLECULAR BIOLOGY, AND BIOINFORMATICS
RASHBEHARI SIKSHA PRANGAN, UNIVERSITY OF CALCUTTA**

**92, ACHARYA PRAFULLA CHANDRA ROAD, RAJA BAZAR, KOLKATA-700009, WEST
BENGAL**

ACKNOWLEDGMENT

INDEX

ABSTRACT

The traditional computational modeling of biomolecular structure, dynamics, and interactions remains difficult for many biomolecular systems. One possible way for overcoming the size problem of biologically important macromolecules is through coarse-graining, in which, instead of describing the molecules in their atomic details, groups of atoms in the molecules are merged into a "super-atomic units" and then compute the structure and dynamics of the molecules in terms of the motion and interactions of units. In the present report, we present results from our coarse-grained molecular dynamics simulation studies on some microbiologically important supercoiled DNA molecules, namely pBR322 plasmid and phix174 RF I (replicative form I) DNA along with a few smaller DNA mini-circles with random sequences. The coarse-grained description of the structure and the force field were obtained from oxDNA. We show that the structures, energies, and base-pairing propensities of the molecules in the coarse-grained representation follow experimentally observed trends.

AIM/OBJECTIVES

Our aim is to study the structure and dynamics of closed circular DNA molecules and supercoiling using molecular dynamics simulations.

We intended to study two different DNA systems in their relaxed and supercoiled form to observe the effect of supercoiling on the structure, dynamics, and energetics of the DNA molecules

- DNA minicircle of 340 base-pairs
- The double-stranded phix174 DNA having 5385 base-pairs

Since the systems are large, we plan to use a coarse-graining approximation for the DNA molecules.

CHAPTER ONE

INTRODUCTION

By dynamic changes in the structure resulting from DNA-protein interactions and stresses given by the structural characteristics of the twin helix, , chromatin adjusts and regulates different DNA-dependent processes. All DNA transactions (like transcription,DNA replication and chromosomal segregation) are necessarily linked to strong alterations in the topological state of the double helix called supercoiling.

DNA supercoiling plays a crucial role in efficiently packing genetic material within cells. Since the length of DNA is significantly greater than that of a cell, it presents a challenging task to fit this genetic information into the cell or nucleus, particularly in eukaryotes. Supercoiling of DNA serves to condense the DNA, creating more space and enabling the packaging of a larger amount of DNA. [Explanation]: - Audience: Knowledgeable - The paraphrased text is written for readers who are familiar with the subject matter. - Formality: Neutral - The paraphrased text maintains a neutral tone, avoiding slang and colloquialisms. - Domain: General - The paraphrased text provides a general description of a scientific phenomenon. - Intent: Describe - The paraphrased text effectively describes the importance and function of DNA supercoiling.

Topoisomerases constitute a family of essential enzymes for normal functioning of living systems. They can relax supercoiled DNA to remove helical constraints that can otherwise hinder DNA replication and transcription and thus block cell growth. Such enzymes can also introduce negative supercoils.

Essentially, DNA supercoiling refers to how much twist a particular DNA strand has, which determines its strain. Positively supercoiled strands are more tightly wound than negatively supercoiled strands. Supercoiling affects several biological processes, including compacting DNA and regulating access to the genetic code (which affects DNA metabolism and possibly gene expression). Mathematical formulas describe the amount of supercoiling in a given strand in comparison with a reference state known as "relaxed B-form" DNA.

Supercoiling can be described numerically by changes in the linking number Lk. The linking number is the most descriptive property of supercoiled DNA. $Lk0$, the number of turns in the relaxed (B type) DNA molecule, is determined by dividing the total number of base pairs of the molecule by the relaxed bp/turn which is 10.4.

$$Lk0 = bp/10.4 \qquad\qquad (1)$$

It is often visualized as the number of Watson-Crick twists found in a circular chromosome in a planar projection (usually imaginary). It cannot be altered without breaking strands at the moment of covalent closure of the chromosome.

An equation below describes the topology of DNA, in which the linking number equals the sum of Tw, the number of twists and turns of the double helix, and Wr, the number of coils. As long as the DNA molecule is closed, the sum of Tw and Wr, or the linking number, remains the same. There may, however, be complementary changes in Tw and Wr without affecting their sum.

$$Lk = Tw + Wr \qquad\qquad (2)$$

Tw, called "twist," is the number of Watson–Crick twists in the chromosome when it is not constrained to lie in a plane. Wr, called "writhe," is the number of superhelical twists. Since biological circular DNA is usually underwound, Lk will generally be less than Tw, which means that Wr will typically be negative.

The change in the linking number, ΔLk, is the actual number of turns in the DNA molecule, Lk, minus the number of turns in the relaxed molecule Lk0

$$\Delta Lk = Lk - Lk0 \qquad\qquad (3)$$

If the DNA is negatively supercoiled, $\Delta Lk < 0$. The negative supercoiling implies that the DNA is underwound. Negative supercoils favour the local unwinding of the DNA, allowing processes such as transcription, DNA replication, and recombination. Negative supercoiling is also thought to favour the transition between B-DNA and Z-DNA and moderate the interactions of DNA-binding proteins involved in gene regulation.

Nanotechnology in DNA and RNA has advanced to full-fledged applications in biology, medicine, and material science. These applications utilize the essential programmability of nucleic acids to generate nano- and even micro-scale structures with proper properties. However, the design of the DNA/RNA sequences that self-assemble into a desired structure is not straightforward and often relies on expensive trial-and-error experimental protocol. A complementary approach is provided by computer-based molecular modeling and simulations, which can model biomacromolecules at different levels of detail. Molecular modeling encompasses all methods, theoretical and computational, used to model or mimic the behavior of molecules. The methods are used in the fields of computational chemistry, drug design, and materials science to study molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modeling of any reasonably sized system.

Molecular dynamics is a technique that allows the simulation of DNA movement, including its folding into double, triple, or quadruple strands, and even its interaction with proteins and drugs. Molecular dynamics is used to address the processes that occur over time scales ranging from picoseconds to minutes and it can be used for molecular systems of various sizes, from a few nanometres to a meter.

CHAPTER TWO

THEORETICAL FRAMEWORK OF TOPIC

Since supercoiled DNA molecules are as a rule quite large in size, carrying out molecular dynamics simulations to study their structure and dynamics is computationally very demanding. To reduce the computational requirement, coarse-graining approximation is invoked in which, instead of considering all the atoms in the DNA molecule explicitly, it treats the DNA molecule as a string of rigid nucleotides. These nucleotides then interact by means of potential energies of interaction that depend on their position and orientations.

One such coarse-grain model for DNA is provided by oxDNA (Ouldridge et al. 2011). In the oxDNA model, the main interactions are sugar-phosphate backbone connectivity, excluded volume, hydrogen bonding, nearest-neighbor stacking, cross-stacking between base-pair steps in a duplex and coaxial stacking. It is also a simulation code that implements molecular dynamics and Monte Carlo techniques. Moreover, a web server version of the code utilising the coarse-grained DNA model is available at oxDNA.org.

OxDNA.org automatically generates simulation files, including a multi-phase relaxation protocol for structures exported into non-physical states from DNA/RNA design tools. When simulation is complete, oxDNA.org offers an interactive visualization and analysis interface wherethe mean values and rmsd/rmsf, distance, energy and occupation of the links can be calculated.

In this project, we use the oxDNA server (https://oxdna.org) for molecular dynamic simulations of relaxed and supercoiled DNA minicircles and double-stranded phix174 DNA.

CHAPTER THREE

PROFILE OF RESEARCH LAB

Most of the naturally occurring RNA modifications are introduced post-transcriptionally. Improved understanding of the properties of RNA with modified residues has allowed us to explore their many-fold use.

This laboratory has been studying the conformational and dynamical context of the naturally occurring modifications that help in regulating the biological functions of RNA. A molecular modeling protocol has been established and validated for the accurate modeling of the structural and dynamic effects of several modifications.

This lab studies a number of these important modifications, theoretically and computationally, and validates against experimental data for the new residues and subsequently, after improvements if needed, uses molecular modeling tools to study the effect of this modification on the structure and dynamics of biologically important RNA molecules.

Apart from this major interest, this lab also studies the conformation and dynamics of DNA protein complexes. Such complexes form during many crucial biological processes, including transcription in which transcription factors bind to their cognate sequences in DNA.

CHAPTER FOUR

METHODOLOGY FOLLOWED

A. Tacox DNA

The tacoxDNA web server (http://tacoxdna.sissa.it/) offers a simple interface to the tacoxDNA suite, which is a collection of programs for converting various common formats of DNA structures and for setting up molecular dynamics simulations. The source code of tacoxDNA is freely available. tacoxDNA allows users to produce complex DNA geometries with or without supercoiling by simply providing an XYZ coordinate file of the DNA center line or by using blueprints generated with the cadnano, CanDo, Tiamat, and vHelix tools. It can also assist in the conversion to and from all-atom or oxDNA representations (Suma et al., 2019).

The modules that are made available in the current implementation are:
• XYZ → oxDNA
• cadnano → oxDNA
• Tiamat → oxDNA
• CanDo → oxDNA
• LAMMPS ⇆oxDNA
• PDB ⇆oxDNA
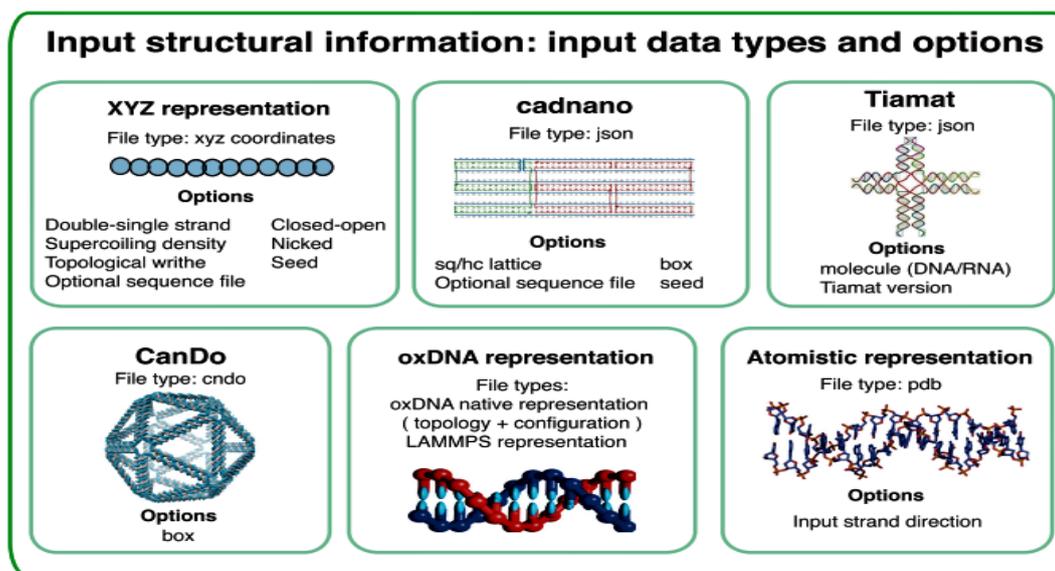
Fig. 1: Input data types for the tacoxDNA web server

B. DNA minicircles

The 340 bp DNA minicircles with random sequences consisting of 50% AT and 50% GC base pairs were constructed using the NAB program of the AmberTools21. The linking number differences were chosen as 0, +5 and -5 to generate a relaxed, positively supercoiled and negatively supercoiled minicircle, respectively. NAB generated all-atom DNA structures in the PDB format.

We took the PDB file created in this manner as an input in the PDB ⇆ oxDNA module of the tacoxDNA server. The PDB → oxDNA module takes a parameter that specifies whether the nucleotides in the PDB file are listed in the 5' → 3' (default) or 3' → 5' directions. In addition, users can optionally choose to omit hydrogen atoms from the output. The PDB ⇆oxDNA module returns the corresponding oxDNA configuration and topology file, ready to use in MD simulations.

The resulting oxDNA configuration file and the topology file suitably modified to reflect the closed circular nature of the DNA was submitted to the oxDNA webserver for molecular dynamics simulations using default setup parameters.

oxDNA analysis utilities were subsequently used to analyse the trajectories for structural and dynamical properties of the molecules.
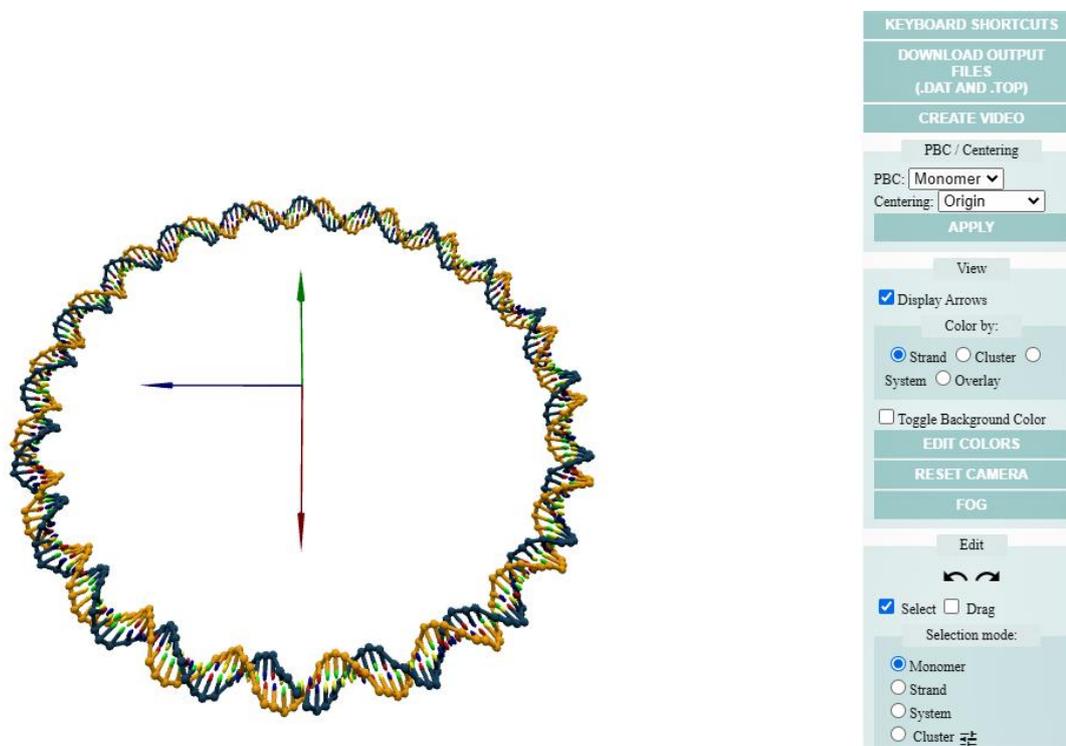
Fig. 2: Circular DNA in oxDNA configuration as viewed in oxView

C. PhiX174

The NAB utility in AmberTools generated DNA molecules with random sequences. To generate a double stranded DNA molecule with a specific sequence like phix174 RF I, we used the XYZ → oxDNA utility of the tacoxDNA webserver.

The use of the above-mentioned utility required the construction of the centerline coordinate file for the coordinates of the centers of each base pair of the DNA. To generate the centerline coordinates, we wrote a program in fortran90 language which is given below

D. Fortran code

```
program cc_centreline

! This program prints coordinates of regularly spaced points on a circle in the X-Y plane implicit none
integer nbp                                    ! number of  base pairs
integer k
real, dimension(1:100000) :: xcoor, ycoor, theta    ! X, Y coordinates and polar angle!
real, dimension(1:100000) :: ycoor
real :: r, pi = 3.14159

print *, ' Number of base pairs'
read *, nbp

r = 3.4*nbp/(2*pi)
```

```
theta(1) = 0.0
xcoor(1) = r
ycoor(1) = 0.0

print *, xcoor(1), ycoor(1), 0.0

do k=1, nbp - 1
    theta(k+1) = theta(k) + 2*pi/nbp
    xcoor(k+1) = r*cos(theta(k+1))
    ycoor(k+1) = r*sin(theta(k+1))
    print *, xcoor(k+1), ycoor(k+1), 0.0
enddo

end program
```

The centerline coordinated file generated by this program can be used with the tacoxDNA utility to generate a supercoiled DNA with a given sequence in the FASTA format and a given linking number difference. The phix174 DNA sequence was downloaded from NCBI and used to construct a relaxed DNA and a molecule with a physiological amount of negative supercoil. The calculation was carried out as follows

E. Calculation of the linking number difference of PhiX174

The natural superhelical density ofPhiX174 (σ) is approximately -0.06

(σ)=LK-LK0 /LK0
Or, (σ)= ΔLK/LK0

Here, ΔLK = σ * LK0

For relaxed molecules, Wr = 0 & LK0= Tw

Here total nucleotide of PhiX174 is =5385
Therefore, LK0=5385/10.5 (For B-DNA 10.5 number of base pair turn)
 =512.86
 $\approx$513

Hence, ΔLK = σ * LK0
   ΔLK = -0.06*513
      = -30.78
      $\approx$ -31

F. INPUT FILES FOR THE SERVER

To start an oxDNA simulation, we need three files: topology (.top) and configuration (.conf or .dat) that describe the structure, and the input file that describes the parameters of the simulation (such as temperature, duration of the simulation, salt concentration). The server automatically generates the input file for all, but one needs to provide the .top and .dat files for the simulation e-server that describes the configuration one wants to simulate.
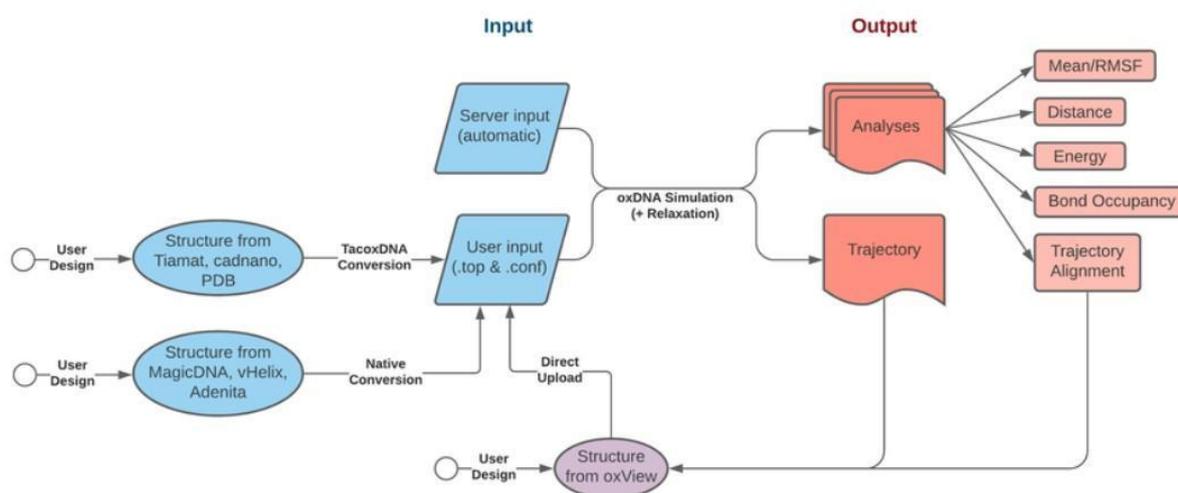
Fig. 3: Workflow of the oxDNA.org pipeline

G. SERVER OUTPUT RESULTS

There are two pages of results of the simulated tasks performed on the server. The first is a summary table of all the work performed by the user, which contains the links to download the initial configuration and topology files submitted, as well as the simulation input file and job logs generated by the server. In addition, the user can view or download the last configuration output generated by the simulation, as well as a zip archive containing the whole simulation trajectory.

Users can post-process their simulation trajectory by clicking on the job name, which takes them to the analysis page.

- Mean and RMSF
- Align Trajectory
- Distance
- Energy
- Bond Occupancy
- Duplex Angles

H. PDB file view from RasMol

Biological macromolecules, by their nature, have complex 3D structures which cannot always be easily appreciated from a 2D picture. A far better way of getting a feel and understanding of such a structure is to manipulate it interactively using a molecular graphics program.

RasMol is a molecular graphics program intended for the visualization of proteins, nucleic acids, and small molecules. RasMol is very easy to use in Windows, Linux, and Unix. Its can rotate protein structures, render them in different ways, and zoom in on them, label atoms & residues. One of the powerful features of RasMol is the use of consistent, highly informative color schemes.

Fig. 4: RasMol view of mini-circles DNA LK +5

CHAPTER FIVE

ANALYSIS AND INTERPRETATION OF RESEARCH PROJECT

➢ *Setup of current simulations*

We run three types of mini-circles DNA with the same number of base pairs(nbp=340) but different linking (dlk = 5, dlk= -5 &dlk= 0). And run the double-stranded phix174 DNA same base pairs number (nbp=5385) but a different linking number(dlk=0 &dlk=-31).Total we constructed 3 mini-circles DNA & 2 phix174 DNA. After simulating these three DNA minicircle structures and two types of phix174 DNA we compared the RMSD and energy files, obtained from the oxDNA server, for the three mini-circles DNAs & two phix174 DNA.

➢ *After simulation in oxdna*
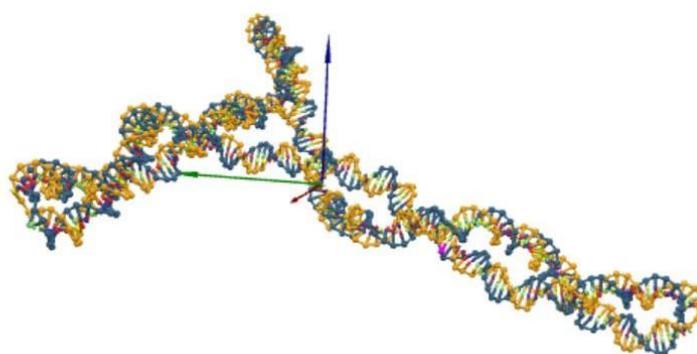
Fig. 5: Mini-Circles DNA dlk=+5 view from Oxview
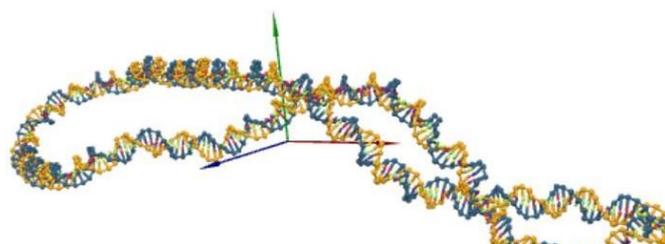
Fig. 6: Mini-Circles dlk=-5 view from Oxview

Fig. 7: PhiX174 dlk= 0 view from Oxview
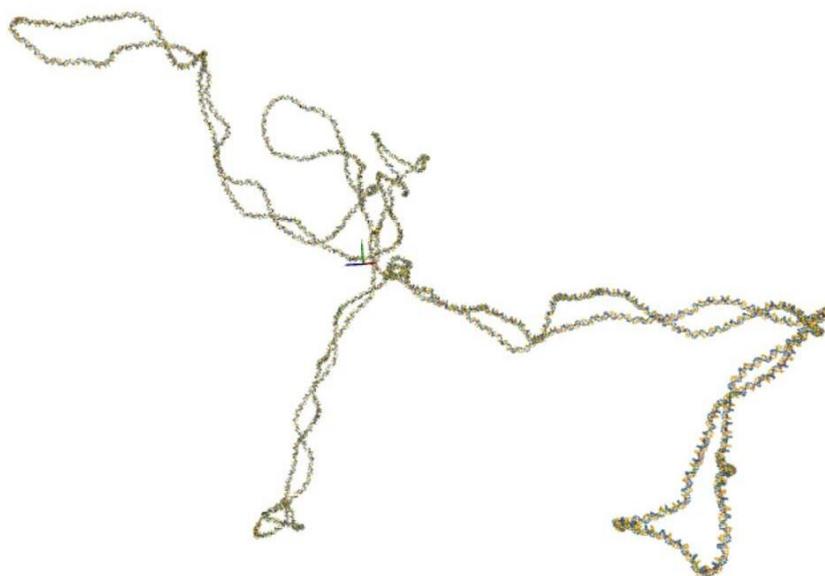
Fig. 8: PhiX174 dlk= -31 view from Oxview

CHAPTER SIX

CONCLUSION & SUGGESTION OF RESEARCH PROJECT

In the present report, we present results from our coarse-grained molecular dynamics simulation studies on some DNA minicircles and some microbiologically important supercoiled DNA molecule phix174 RF I (replicative form I) DNA. The coarse-grained description of the structure and the force field were obtained from oxDNA. We show that the structures, energies, and base-pairing propensities of the molecules in the coarse-grained representation follow experimentally observed trends.

The current project can be extended in the future to study in more detail the structural and dynamical consequences of supercoiling under different sequence contexts and under different environmental conditions.

CHAPTER SEVEN

RESULT(S) ACHIEVED

A. *Minicircles*

In Figure 9, we have plotted the root-mean-square deviation from the initial structure as a function of the frames in the molecular dynamics trajectory of the 340 bp DNA minicircles. The left panel shows the result for linking difference 0 (relaxed), the middle panel is for lining difference +5 (positively supercoiled) and the right panel shows the result for linking difference -5 (negatively supercoiled).

It is clear from the plots that the mean RMSD (shown as a red line) and its variation are lowest for the relaxed minicircle. This implies that during the dynamics, the relaxed structure does not change much with time. However, the RMSD values in the middle and the right panels are much larger, showing that, as a result of supercoiling-driven deformation, the DNA minicircles change their structure from the initial planar circular structure by quite a large amount.
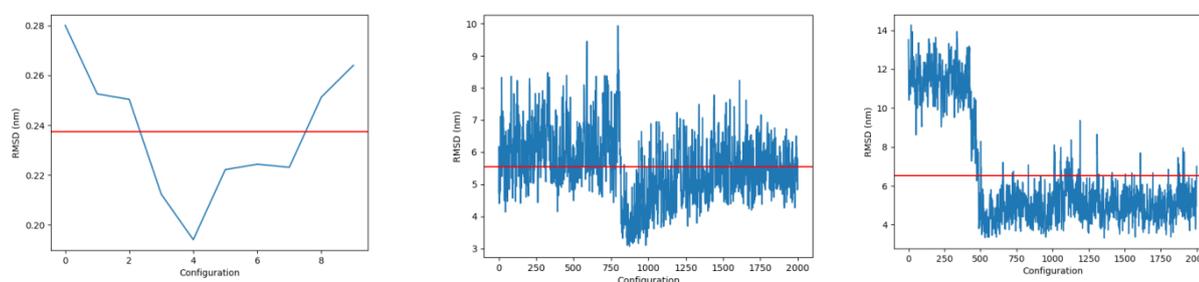
Fig. 9: The time evolution of the RMSD of DNA minicircles (340 bp) and the mean RMSD (in red). The left panel is for the relaxed minicircle, the centre for the positively supercoiled (Delta Lk = +5), and the right one is for the negatively supercoiled (Delta Lk = -5) molecules.

In figure 10, we have plotted the distribution of the total energy of the minicircles. Since a significant component of the total energy of the minicircles arise from the supercoiling-driven deformation of the DNA structure, we see a shift of the maximum probable energy (the peaks of the histograms) towards higher values in the cases of positively and negatively supercoiled molecules compared to the relaxed minicircle. Also, the increase in energy is approximately the same for the positively and negatively supercoiled molecules which corresponds to the theoretical result that the supercoiling energy is proportional to the square of the linking difference.
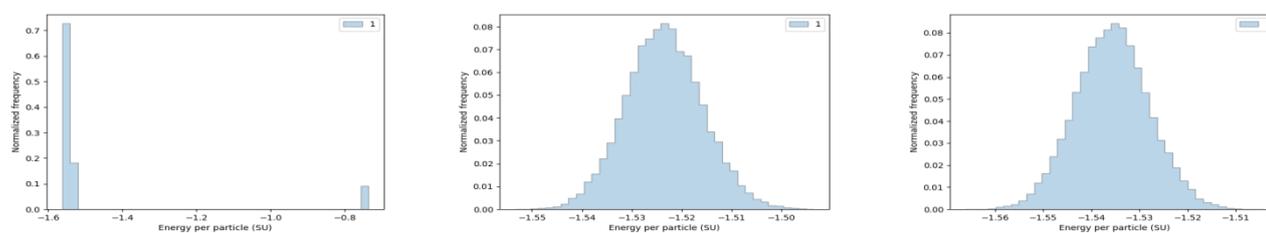
Fig. 10: Energy histogram for the DNA minicircles. The left panel is for the relaxed minicircle, the central one for the positively supercoiled (Delta Lk = 5) and the right one for the negatively supercoiled (Delta Lk = -5).

B. PhiX174

The phiX 174 bacteriophage is a single-stranded DNA (ssDNA) virus that infects *Escherichia coli* and is the first DNA-based genome to be sequenced. Its capsid structure is icosahedral (i.e. spherical) and its genome size ranges from 4.6 kb to 6.1 kb. Microviridae provides the first evidence of overlapping genes. Studies on the replication of these phages led to the discovery of rolling circle replication. PhiX174 has been used in many landmark experiments because of its small genome size (5,385 nucleotides) and nonpathogenic status. The capsid of phiX174 is designed to find bacterial cells, and then infect them with its DNA. Sixty copies of the capsid protein form a spherical shell around the DNA, and the spike proteins form 12 pentagonal spikes on the surface. The DNA is thought to be ejected through the middle of the spikes when the virus infects an *Escherichia coli* cell. The DNA itself encodes 11 genes. To fit into this tiny protein shell, however, the DNA is so short that the genes must actually overlap.

It has been used as a model organism in many evolution experiments. phiX174 is regularly used as a positive control in DNA sequencing due to its relatively small genome size in comparison to other organisms, and its fairly balanced nucleotide content. phiX174 is also used to test the resistance of personal protective equipment to blood-borne viruses. phiX174 has also been modified to enable peptide display (phage display) from the viral capsid G protein. The ΦX174 genome was the first phage to be cloned in yeast, which provides a convenient dry dock for genome modifications. Soon after making their discovery, Twort and d'Hérelle began to use phages in treating (phage therapy) human bacterial diseases such as bubonic plague and cholera.
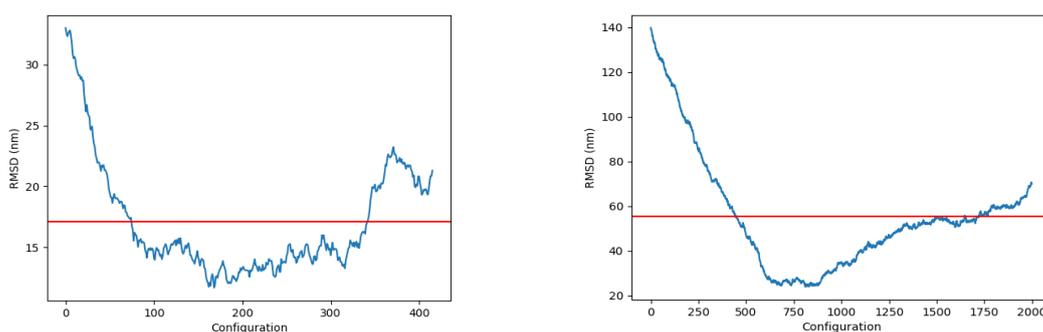
Fig. 11: RMSD value of Phix174 (Delta lk 0 & -31) for the measure of the average distance between the atoms (usually the backbone atoms) .The left panel is for the lower RMSD (Delta lk = 0). And the right panel for the higher RMSD (Delta lk = -31).

C. ENERGY HISTOGRAM OF PhiX174 LK (0 & -31)
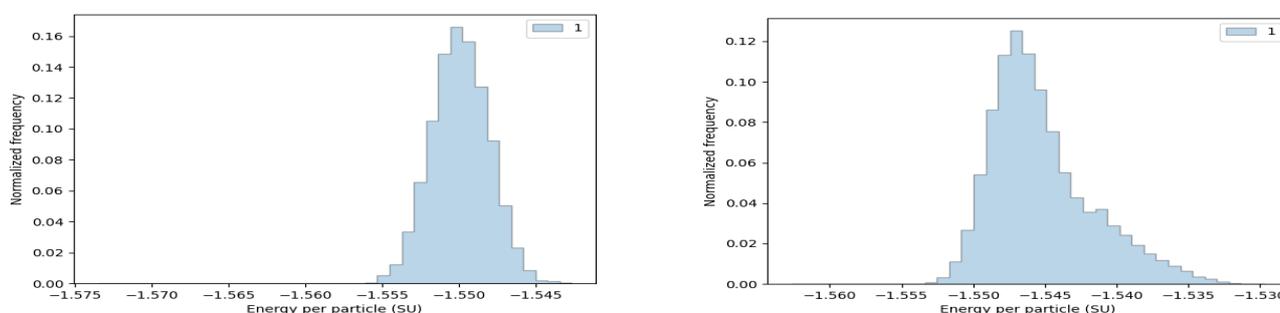
Fig. 12: The energy histogram of Phix174 (Delta lk 0 & -31) plots the potential energy of the structure & produces a histogram. The left panel is for relaxed Phix174(Delta lk 0) and the right panel is for the negatively supercoiled Phix174(Delta lk -31).

In Figure 12 we plotted the distribution of the total energy of the phix174 (Delta lk = 0 & -31). Since a significant component of the total energy of the Phix174 arise from the supercoiling-driven deformation of the DNA structure, we see the maximum probable energy (the peaks of the histogram) towards a lower value in Phix174 (Delta lk = 0) compared to the Phix174(Delta lk = -31). The increase of energy depends on the linking difference.

D. *The average energy of Figure 10*

Delta lk	Average energy (arbitrary unit)
0	-1.547732
-5	-1.535718
+5	-1.523517

E. *The average energy of Figure 12*

Delta lk	Average energy (arbitrary unit)
0	-1.549801
-31	-1.545224

F. *Comparison of energy of Figure 12*

	-1.575	-1.57	-1.565	-1.56	-1.555	-1.55	-1.545	-1.54	-1.535	-1.53
LK 0	0	0	0	0	0.018	0.164	0.01	0	0	0
LK-31	0	0	0	0	0	0.03	0.095	0.026	0.008	0
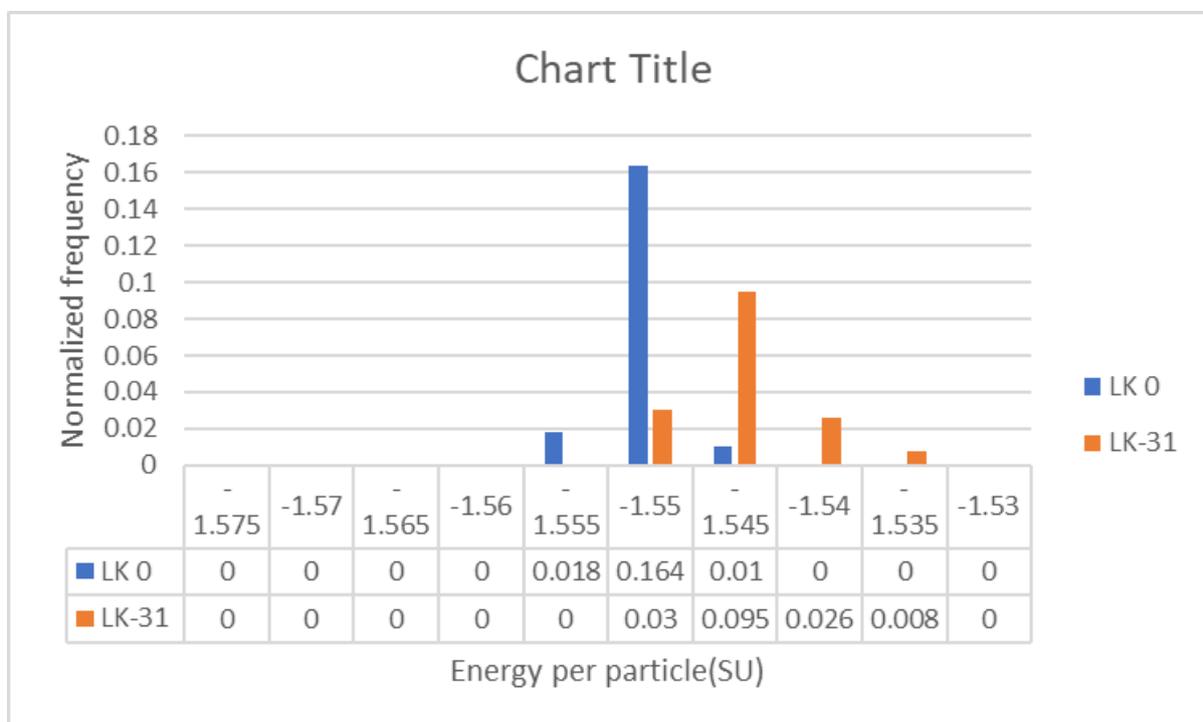
Fig. 13: Comparison of energy of Phix174 (Delta lk = 0 & -31).

REFERENCES

[1.] T. E. Ouldridge, A. A. Louis and J. P. K. Doye, J. Chem. Phys, 134, 085101 (2011)

[2.] Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (July 2016)."Coarse-Grained Protein Models and Their Applications". Chemical Reviews. 116 (14):7898–936. doi:10.1021/acs.chemrev.6b00163. PMID 27333362

[3.] Thomas J. Macke , W. A. Svrcek-seiler , Russell A. Brown. NAB Users ' Manual Version 5.1 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.460.7407

[4.] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. An overview of the Amber biomolecular simulation package https://doi.org/10.1002/wcms.1121

[5.] A. Suma, E. Poppleton, M. Matthies, P. Šulc, F. Romano, A. A. Louis, J. P. K. Doye, C.Micheletti, and L. Rovigatti, TacoxDNA: A user-friendly web server for simulations ofcomplex DNA structures, from single strands to origami, J.Comput. Chem. 40, 2586(2019)

[6.] Erik Poppleton, Roger Romero, Aatmik Mallya, Lorenzo Rovigatti, Petr Šulc, OxDNA.org: a public webserver for coarse-grained simulations of DNA and RNA nanostructures, *Nucleic Acids Research*, Volume 49, Issue W1, 2 July 2021, Pages W491–W498, https://doi.org/10.1093/nar/gkab324

[7.] Bansal M, Bhattacharyya D, Ravi B. NUPARM, and NUCGEN: software for analysis and generation of sequence-dependent nucleic acid structures. Comput Appl Biosci. 1995Jun;11(3):281-7. doi 10.1093/bioinformatics/11.3.281. PMID: 7583696.

[8.] Baron R, Trzesniak D, de Vries AH, Elsener A, Marrink SJ, van Gunsteren WF.Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. ChemPhysChem. 2007 Feb 19;8(3):452-61. doi 10.1002/cphc.200600658. PMID: 17290360.