Coarse-grained modelling of a DNA origami force sensor

Navoneel Sen

Lady Margaret Hall

University of Oxford



A Thesis Submitted for the degree of

Master of Science (by Research) in Physical and Theoretical Chemistry

Trinity 2021

To Ma, Baba and Tua.

Because you always understood.

Abstract

We have simulated large-scale DNA nanostructures using the oxDNA software which uses coarse-graining at the nucleotide level. These simulations can complement experiments as they typically allow greater structural resolution and give access to a range of properties that are otherwise hard to measure (e.g. internal forces).

To address the challenges associated with parallelised force spectroscopy, we report the development of a bottom-up approach for manipulating the mechanical forces experienced by target biomolecules. We have used molecular dynamics simulations to study the mechanical unfolding of DNA hairpins using an origami-polymer force clamp which allows ensemble scale parallelisation. The target molecules are sandwiched between a rigid DNA origami and a thermo-responsive polymer particle that shrinks upon a rise in temperature. This technique does not need any dedicated single-molecule force spectroscopy instruments or surface tethering. The key features of the origami-polymer force clamp are discussed along with the characterisation of the mechanical unfolding of DNA hairpins, and how these simulations aid in the interpretation of experimental results from these devices.

In eukaryotic cells, the basic structural unit into which DNA is packaged is a nucleosome. We have developed models of the nucleosome in its fully wrapped and partially unwrapped states to enable the design of DNA origami force clamps that will allow the structure of the nucleosome to be determined by cryoEM while subject to force. Our work focuses on molecular dynamics simulations of the nucleosome core

particle obtained from a species of African clawed frogs (*Xenopus laevis*). We have looked at the response of our nucleosome models to external force and determined the parameters needed to simulate it in a partially unravelled state. The next step in this study would be to optimise clamp designs that utilise the entropic spring behaviour of ssDNA to exert forces that stabilise the nucleosome in the partially unwrapped state. The ultimate goal of this project is to develop a DNA origami based force-sensor that can exert the appropriate force to cause unfolding of nucleosomes.

The equilibrium structure of a DNA origami can be manipulated in a controlled way using engineered defects within the design. We have explored how to use the insertion or deletion of base pairs to generate complex structures out of a 6 helix-bundle DNA origami. The particular target is a three-dimensional trefoil knot. The designs considered involve a complex interplay between twist and curvature and we report the progress we have made towards that target.

Acknowledgements

To begin with, I would like to thank my supervisor, Professor Jonathan Doye, for not just being an exemplary mentor but for being extremely supportive and understanding at all times. I am grateful for his inexhaustible reserve of ideas and his ability to inspire and enthuse. I am deeply indebted to Marco Wong for all his suggestions and valuable inputs. My experimental collaborators Hanquan Su, Joshua Brockman and Khalid Salaita have been extremely helpful and patient and I must thank them for their invaluable support. I have been so lucky to receive tidings of warmth from Hemani Chhabra to whom I must extend my deepest gratitude. While life in the theory wing of the PTCL is presumed by most to be extremely uneventful and monotonous, Michael Selby and Rory Giddins did not step back from proving the contrary by keeping me busy in games of Monopoly Deal while never letting me focus too much on work in the office. It was a pleasure to share the same office with Hannah Fowler, Ian Seet, Chris Mingard, Don Danilov, Monty Sant, Chuyan Tang, Jess Mitchell, Cintia Pulido Lamas and Jon (the panda) who were all very kind and welcoming.

A special mention for a special person: Lavanya Singh, a truly remarkable human whose charming company has brightened up quite a few gloomy afternoons in Oxford. My heart goes out to my wonderful friends and flatmates Gayatri and Arya who never shied away from showering me with compliments. I cannot begin to describe how much I appreciate your support and kindness. The two of you helped

me find home and happiness when I was about to give up on both. I am thankful to Ashwin Jainarayanan for helping me realise the importance of friendship, or when it was necessary: the lack of it. I must thank my friends from different corners of the world, including but not limited to: Peter Pathikrit, Kathamrita Mukherjee, Yash Maheswari, Aishwarya Mukhopadhyay, Shivani Suresh, Jaswanth Dadi, Olana Peters, Ruth Parry, Caren Crofton Home, Petunia Fernandez and Vipin Sabu for never letting me down.

Finally, I must thank my family for everything that they have done for me. I would not be the person I am today, without their love, support and sacrifice.

My journey was not a perfectly smooth ride and I just want to take a moment to say that I am proud of myself for finishing it. What got me going through thick and thin were the wonderful clouds atop Port Meadow and plentiful bowls of Kothu Roti from the Sri Lankan stall at Gloucester Green Market. The western world is not necessarily kind and welcoming to people like myself and I want to thank everyone who is actively working to make it more inclusive and tolerant. I remember promising someone that if I am able to complete this degree, I will start believing in magic. I am expecting my letter from Hogwarts any day now.

List of abbreviations

The following abbreviations have been used in this thesis:

A	adenine
AFM	atomic force microscopy
AMBER	assisted model building with energy refinement
ANM	anisotropic network model
bp	base pair
C	cytosine
CHARMM	chemistry at Harvard macromolecular mechanics
cryoEM	cryogenic electron microscopy
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
FENE	finitely extensible non-linear elastic
FRET	Förster resonance energy transfer
G	guanine
LAMMPS	large-scale atomic/molecular massively parallel simulator
MC	Monte Carlo
MD	molecular dynamics

NCP	nucleosome core particle
nt	nucleotide
OPFC	origami polymer force clamp
PDB	protein data bank
RNA	ribonucleic acid
SCOMP	smoothly cutoff Morse potential
SMFS	single molecular force spectroscopy
ssDNA	single-stranded DNA
T	thymine
3SPN	3 site per nucleotide
6HB	6 helix bundle

Table of Contents

List of Figures **ix**

List of Tables **xv**

1 Introduction **1**

1.1 Chemical and structural properties of Nucleic Acids 2
1.2 Biological functions of nucleic acids 5
1.2.1 Inheritance 5
1.2.2 Gene expression 5
1.2.3 DNA topology 7
1.3 DNA Nanotechnology 8
1.4 Studies in DNA Nanotechnology 10
1.4.1 Theoretical Models of DNA 10
1.5 Single-molecule manipulation methods 14
1.5.1 Atomic Force Microscopy 15
1.5.2 Optical and Magnetic tweezers 16
1.6 Fluorescent Techniques 19

2 Methods **22**

2.1 Molecular Dynamics 22
2.2 The oxDNA Model 25
2.2.1 Coarse-graining 25
2.2.2 OxDNA and oxDNA2 models 27
2.2.3 Conversion of design to oxDNA format 31
2.2.4 Relaxation of initial geometry 32
2.2.5 Simulation 34
2.3 ANM-oxDNA: A nucleic acid-protein hybrid model 35
2.3.1 Protein Model 36
2.3.2 Extensions to the model 38
2.3.3 Protein-nucleic acid interactions 38

3 DNA-Polymer force clamp **39**

3.1 DNA Origami substrate 40
3.2 Design of hairpins and linkers 43

3.3 System setup . . . 44
3.4 Force calculations . . . 50
3.5 Angle of bending . . . 54
3.6 Elastic Properties of the origami rod . . . 58
3.7 Additional simulations . . . 60
3.8 Comparison with experimental results on the force clamp . . . 64

4 Structure of the nucleosome core particle under the influence of external force 68
4.1 Simulating a nucleosome in ANM-oxDNA . . . 71
4.2 Nucleosome under constant force . . . 72
4.2.1 Effect of trap stiffness . . . 74
4.2.2 Effect of the number of traps . . . 74
4.3 Force-sensor calibration . . . 76
4.4 Force actuating device strategy . . . 77

5 DNA origami trefoil knots 81
5.1 DNA-origami six-helix-bundles . . . 83
5.2 Twisted and bent nanotubes . . . 84
5.2.1 Completely untwisted 6HB . . . 85
5.3 Calculation of curvature and torsion . . . 86
5.4 Assembling 6HBs . . . 90
5.4.1 Individual monomer designs . . . 92
5.4.2 Kabsch algorithm for positioning 6HBs . . . 92
5.5 6-particle trefoil . . . 97

6 Conclusions and Future Work 98
6.1 Conclusions . . . 98
6.2 Future Work . . . 99

List of Figures

1.1 Molecular structure of DNA 4
1.2 The DNA replication process by which each DNA strand produces two identical copies of itself. 6
1.3 Electron micrograph of DNA isolated from E.coli. Linear DNA (arrowheads), a plasmid (box) and a rosette-like structure (arrow) were observed. Figure reproduced from Ref. [38]. 8
1.4 The self-assembly of a DNA origami. The picture shows a scaffold strand (black) with the red parts denoting regions complementary with the corresponding (red) staple strand that forces the scaffold regions to come close. 10
1.5 A few examples of different DNA nanostructures. (a) Some building blocks of nanostructures including (i) double-crossover (ii) triple-crossover (iii) 4x4 and (iv) three-point start tiles. (b) Bottom-up assembly of building block leads to higher order structures such as (i) DNA cube (ii) DNA tetrahedron (iii) DNA polyhedron (iv) DNA octahedron and (v) DNA dendrimer. (c) Structures made using DNA origami. (d) A few other DNA nanostructures like (i) DNA nanohydrogels (ii) DNA nanoflowers and (iii) metal-induced DNA assembly. Figure reproduced with permission from Ref. [49]. 11
1.6 A schematic showing the operation of an Atomic Force Microscope. The cantilever (orange) exerts minute forces on the molecule, whose response leads to a change in the position of the cantilever that is recorded when a laser beam (red) is reflected onto a quadrant photodiode (blue). (a) A topographical image is obtained when the cantilever scans the target molecule (b) A DNA molecule tethered to a glass coverslip. Figure reproduced from Ref. [100]. 16

1.7 Two classes of molecular tweezers. (a) Examples of a typical optical tweezer setup. A DNA molecule (blue/green) is chained between an optically trapped dielectric bead (violet) and either the glass substrate (top) or another bead held on a micropipette tip (middle) or a second optical trap (bottom). The black arrow shows the direction of the force acting on the molecule. (b) A typical magnetic tweezer experiment showing a magnetic bead trapped between a pair of permanent magnets. The solid arrow represents the direction of the force (F) on the bead and the dashed arrow indicates the direction of the induced magnetic moment (m_0). The induced moment can provide the torque required for supercoiling resulting in plectonemic loops. Figure reproduced from Ref. [100]. 17

2.1 (a) A representation of a rigid oxDNA nucleotide. (b) A 12 base pair DNA double helix as represented by oxDNA. The inset shows the different type of interactions that are accounted for in the oxDNA model, namely, hydrogen-bonding between complementary bases, (coaxial) stacking interactions between bases, cross-stacking interactions between diagonally opposite bases in a double helix, and a FENE potential between backbone sites, as well as excluded-volume repulsion associated with the backbone and base sites. Figure adapted from Ref. [147] 26

2.2 Snapshots of a DNA duplex in oxDNA1 (top) and oxDNA2 (bottom), including a cross sectional view of a base pair. Distinct major and minor grooves are apparent in oxDNA2 as compared to oxDNA1 where the grooves are symmetric. Figure reproduced with permission from Ref. [149] 28

2.3 (a) The FENE backbone potential and its modified form for parameter values $F_{\max} = 10$; $A = 1$ (values in simulation units; equivalent to 486 pN and 48.6 pN). (b) The forces due to these potentials. . . . 33

2.4 ANM representation of a protein. The blue sphere encloses all residues within the cutoff distance ($r_{\max}$) from the central residue (black circle). The interacting residues are shown in in blue whereas the non-interacting residues are shown in red. Figure adapted with permission from Ref. [162] 36

3.1 A schematic comparing conventional force spectroscopy with our approach that can allow ensemble-level parallelisation. Whereas conventional approaches may use an electric or a magnetic field to exert force, the OPFC design exerts force due to the decrease in size of a polymer sphere with an increase in temperature. Figure adapted with permission from Ref. [174]. 41
3.2 (a) caDNAno design of the 16 helix bundle origami rod and (b) its crosssection 44
3.3 caDNAno representation of the same rod as **Figure 3.3** but with hairpins and linkers drawn. The inset shows three hairpins as single strands on virtual helices 16 and 20 as well as three linkers on virtual helix 17. 45
3.4 Binding of hairpins to complementary strands on the origami and the polymer sphere in a typical experiment. 45
3.5 Schematic showing how the parameters for modelling the polymer sphere in the oxDNA simulations were chosen (for clarity only four of the six hairpins are shown). The relative sizes of the objects are not to scale. 47
3.6 Simulation snapshots of systems **C1**, **C6** and Δ32 at low temperature showing their respective structural features. The hairpins have been coloured blue while the linkers have been coloured green and orange according to the same scheme used in experiments (**Figure 3.4**) . . . 48
3.7 Variation of the hydrodynamic diameter of the sphere with temperature. 49
3.8 The average force acting on all hairpins for each system at low and high temperature. The dashed black line represents the $F_{\frac{1}{2}}$ of the hairpin. The green dot represents the scenario explained in **Section 3.7** where the hairpins were deliberately kept open. 51
3.9 Force equilibration plot for one of the hairpins in the **C6** system at high temperature. 53
3.10 Distributions of the force going through each hairpin in the **C6** system at high temperature. 53
3.11 Fluctuations in the angle of bending for the **Δ32** system at low and high temperature. 55
3.12 The average angle of bending for each system at low and high temperature. The green dot represents the scenario explained in **Section 3.7** where the hairpins were deliberately kept open. 56

3.13 Plot of the tangent-tangent correlation function for the 16HB origami rod. The periodic trend is observed every 33 base pairs which roughly corresponds to the frequency at which junctions occur in the structure of the origami. 59
3.14 Results from individual constant force simulations of the hairpin. . . 61
3.15 Fraction of time spent by the hairpin in each state as a function of the force applied to its ends. The solid dashed lines show the moving average for each state. 62
3.16 Snapshots of the hairpin held at a constant force of (a) 2.0 pN (b) 15.5 pN (c) 16.5 pN (d) 20.0 pN. 62
3.17 Schematic showing only the 10 linkers in the **C6** system. 64
3.18 **C6** system with the hairpins kept deliberately open. 64
3.19 Temperature-dependent fluorescence measurements of the (a) **C6** (also known as $\mathbf{L_{110nm}}$) (b) Δ32 and (c) **C1** (also known as $\mathbf{L_{10nm}}$) systems. The plots in the right column are zoomed-in versions of the corresponding plots in the left column. 65

4.1 Electron micrographs of NCPs in ice. White and black arrows indicate bent DNA and straight DNA respectively. Figure reproduced with permission from Ref. [191] 69
4.2 Snapshots of the nucleosome **3LZ0** as visualised on oxView from two different angles. The colour of the amino acid residues represented as balls is arbitrary. 71
4.3 The force extension curve of a spool model of a nucleosome. Theoretical curves are drawn for completely wrapped (purple), partially unwrapped (blue) and completely unwrapped (green) states along with experimental data for increasing (black) and decreasing (red) tension. The solid part of the curve indicates that the state is thermodynamically preferred and the dashed part indicates a state that is stable but not preferred. Figure reproduced with permission from Ref. [213]. 73
4.4 Dependence of nucleosome extension on trap stiffness under a constant external force of 4 pN. Snapshots show configurations of the nucleosome at a particular stiffness value. The extension is measured relative to the completely wrapped state of the nucleosome when simulated using ANM-oxDNA. 75

4.5 Dependence of nucleosome extension on the number of traps under a constant external force of 4 pN. Snapshots show configurations of the nucleosome with a particular number of traps. The extension is measured relative to the completely wrapped state of the nucleosome when simulated using ANM-oxDNA. 76
4.6 Simulation snapshots of the force sensor developed by Nickels *et al.* in the (a) off and (b) on states. 78
4.7 End-to-end distance distribution for scaffold nucleotides on either side of the central gap in the Nickels force clamp. 78
4.8 A schematic showing the strategy that is going to be used to generate force in ssDNA springs (shown in brown). The two states of the system before and after hybridisation are shown separately in (a) and (b), respectively. The orange line represents the extra strand that initiates the hybridisation process by binding to complementary regions on the ssDNA (shown as green lines). The lengths of various sections (in nucleotides) are labelled. The values of l and g can be changed keeping the length $l + 2d + 2g$ constant, to change the force. 79
4.9 Simulation snapshots of the nucleosome in (a) the completely wrapped state at low force and (b) the partially unwrapped state where the force acting on the nucleosome is a result of ssDNA strand shortening (as described in **Section 4.4**). For the system shown here, $l = 80$ nt, $g = 5$ nt and $d = 15$ nt. 80

5.1 RecA coated DNA trefoil. Figure reproduced with permission from Ref. [216]. 81
5.2 Two-dimensional gel electrophoresis of purified DNA obtained from P4 Phage particles. Knots are identified by comparing the gel velocity of twist knots with known relative migration distances (center) with that of individual knot populations resolved by electrophoresis (right). Figure reproduced with permission from Ref. [227]. 82
5.3 The cross section of a honeycomb-like lattice in caDNAno. 84

5.4 Controlling the twist and curvature in DNA origami helix bundles. (a) Helices run parallel in a honeycomb-like lattice where crossovers between neighbouring helices occur every 7 bp. The crossover planes divide the helix bundle to small sections out of which one is highlighted. (b) A section of a helix with 7 base pairs between two consecutive crossover planes, exerting no stress on its neighbours. (c) Above, a section with 5 base pairs exerting a left-handed torque on its neighbours. Below, a section with 9 base pairs exerting a right-handed torque on its neighbours. (d) Deletion/insertion of base pairs (shown in orange/blue) in appropriate positions leads to global left-/right-handed twisting, respectively. (e) A combination of appropriately placed deletions (shown in orange) and insertions (shown in blue) can be used to induce global bending. Figure adapted from Ref. [46]. 85
5.5 A measure of twist as a function of the number of planes of deletions in a 6HB origami. 86
5.6 A snapshot of the 6HB origami with 6 planes of deletions. Helices diagonally opposite to each other in the hexagonal lattice are coloured the same. 87
5.7 (a) Curvature and (b) torsion of a trefoil knot as a function of the arc length. 89
5.8 (a) Curvature and (b) torsion of a 6HB monomer (Monomer C, as described later in **Section 5.4.1**). 91
5.9 Visualisation of each monomer from two different angles. 93
5.10 A snapshot of three monomers, each oriented by a rotation matrix calculated from the KABSCH algorithm. 95
5.11 The monomers in **Figure 5.10** after re-positioning in oxView. 96
5.12 Snapshots of the final configuration of the trimer after an attempt was made to assemble it from individual 6HB monomers. 96
5.13 Snapshots of a monomer for a 6-particle trefoil knot composed of two origamis with different handedness of twist. 97

List of Tables

3.1 Average force along hairpins at low and high temperatures. 52
3.2 Average angle of bending at low and high temperatures. 57
3.3 Snapshots of **C1**-**C6** and **Δ32** at low and high temperatures. At high temperature, a strong bend is visible in all systems from **C1** to **C6** while **Δ32** is kinked. 57

5.1 Details of monomer design. 92

Chapter 1

Introduction

In this thesis, the behaviour of DNA when subjected to mechanical stress is studied using a coarse-grained model called oxDNA[1]. The response of nucleic acids to such external forces is extremely important to understand their physical properties and the results of many single-molecule experiments. It also throws light on the behaviour of these molecules in a biological context. The thesis is structured as follows:

- In **Chapter 1**, fundamental properties of nucleic acids and their biological roles will be introduced along with a brief review of recently developed experimental techniques that allow us to study these properties on a single-molecule level.

- In **Chapter 2** the various computational tools used in this study along with their most relevant details are introduced. This includes details of the models used and the programs written thereafter to analyse the data.

- In **Chapter 3** we discuss the design and implementation of an origami-polymer force clamp.

- In **Chapter 4** we develop models of a nucleosome core particle in different force-induced states.

- In **Chapter 5** we consider twisted and bent DNA origami rods and their potential to form a 3D trefoil knot.

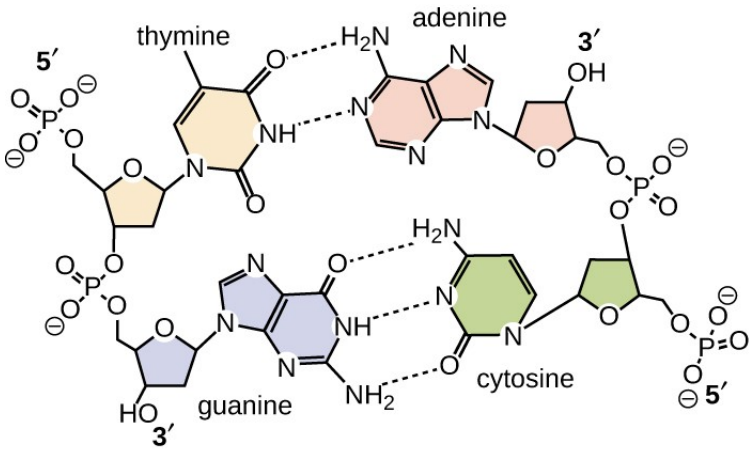
1.1 Chemical and structural properties of Nucleic Acids

One of the greatest discoveries of modern-day biology is the ability of nucleic acids to store genetic information. In 1869, the physiologist Miescher was able to isolate a substance from human white blood cells, which he later called 'nuclein'[2]. He analysed the substance chemically to find that it was rich in phosphorus but did not contain any sulfur, thus concluding that it was not a protein. Building on Miescher's work, the scientist Kossel showed that 'nuclein' is actually made up of a protein part and a non-protein part. He isolated the non-protein part ('nucleic acid') and by 1901, he had shown that nucleic acids are made up of five bases and a sugar[3]. It was Levene who first identified that components of nucleic acids were linked together in the order phosphate-sugar-base to form units that he called 'nucleotides'[4] (see **Figure 1.1a**) and recognised DNA and RNA to be two different types of nucleic acids based on the different sugars present in their structures. However, Levene wrongly hypothesised that DNA contained equal amounts of adenine (A), guanine (G), cytosine (C) and thymine (T) and cannot store genetic information because it was chemically too simple. This hypothesis was proved to be incorrect by Chargaff, who, along with his coworkers discovered that the relative amount of A and T as well as that of C and G were equal in DNA[5]. With the emergence of sophisticated X-ray diffraction techniques, Astbury carried out detailed structural studies on nucleic acids in the 1930s[6]. Although these studies were successful in suggesting that

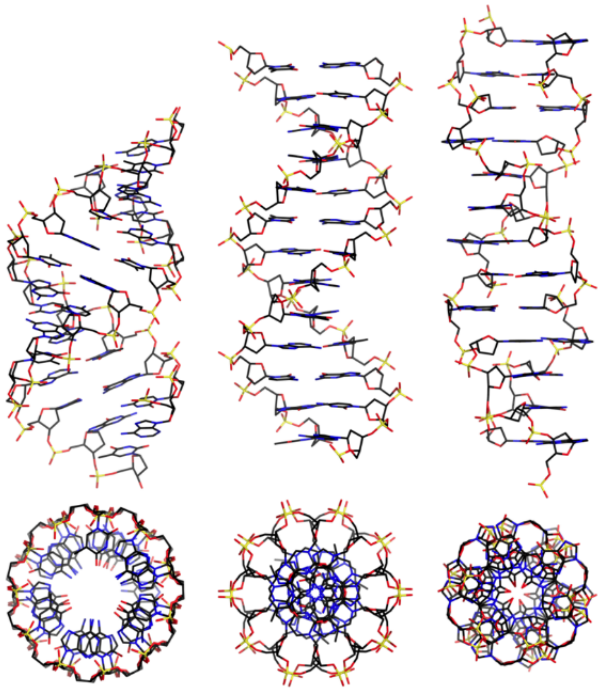
DNA is periodically organised, the interpretation of the diffraction data was only possible when Franklin discovered that DNA existed in two allomorphic forms: the 'A' and the 'B' forms[7] (see **Figure 1.1b**). All studies carried out to determine the structure of DNA culminated in 1953 when Watson and Crick published their model of DNA as a right-handed double helix of nucleotides[8] (**Figure 1.1b**). The model not only showed that individual nucleotides on a strand are linked by means of covalent bonds between the sugar and phosphate but it also included canonical base pairing mediated by two hydrogen bonds in A-T base pairs and three hydrogen bonds in G-C base pairs. Although non-canonical base pairs such as Hoogsten base pairs[9,10] are known to exist in DNA, the most important contribution to the secondary structure of the DNA molecule comes from A-T and G-C base pairs as illustrated in **Figure 1.1**.

The three most common double helical tertiary structures of DNA in a biological context are shown in **Figure 1.1b**. While B-DNA is the most prevalent conformation *in vivo*, Z-DNA is thought to have a specific biological function when the DNA is highly supercoiled, for example, in the case of transcription[12] and some studies have linked the presence of Z-DNA in the hippocampus of the human brain to the progression of Alzheimer's disease[13]. It has been suggested that with A-DNA being shorter than B-DNA, it is able to act as an intermediate during the packaging of dsDNA in a bacteriophage[14–16]. Apart from these, many other tertiary structures have been reported[17].

In the case of RNA, there is a hydroxyl group attached at the C-2 position of the ribose ring instead of a single hydrogen and the nucleobase thymine (T) is replaced by uracil (U). Although DNA and RNA are structurally very similar, the latter



(a) Chemical structure of double-stranded DNA showing the canonical base pairing between A-T and G-C. The dashed lines represent hydrogen bonds whereas the solid lines represent covalent bonds.



(b) A representation of the most common DNA tertiary structures: A-DNA*(left)*, B-DNA*(centre)* and Z-DNA*(right)* and their projections on a plane perpendicular to the helical axis. Figure adapted from Ref. [11]

Figure 1.1: Molecular structure of DNA

has a more diverse biological function. Generally, RNA occurring in cells adopts a single-stranded configuration but they can also exist in an A-helical double stranded form[10,18]. Some studies have hinted at RNA-DNA hybrid structures adopting an A-helical conformation[19,20].

1.2 Biological functions of nucleic acids

1.2.1 Inheritance

Theories of inheritance have existed since ancient times[21] but it was Mendel who first laid out the foundations of genetics[22]. Experiments by Avery *et al.*[23] showed that DNA was responsible for the induction of bacterial virulence thus implying its role in inheritance. Taking inspiration from Schrödinger's theoretical prediction that the carrier of genetic information should be an 'aperiodic crystal'[24], Watson and Crick quickly identified that the structure of DNA was ideal for copying genetic material[8]. Because the base pairing in DNA is specific, the two strands carry complementary versions of the same sequence. A semi-conservative mechanism utilises this feature of DNA in the replication process: The two strands in the duplex are separated and each serves as a template to rebuild the respective complementary strand (see **Figure 1.2**). The enzyme DNA polymerase mediates the above process *in vivo*[25].

1.2.2 Gene expression

In the nucleus of a cell, a process called transcription converts the genetic information stored in DNA to RNA. The sequence is read by the enzyme RNA polymerase,

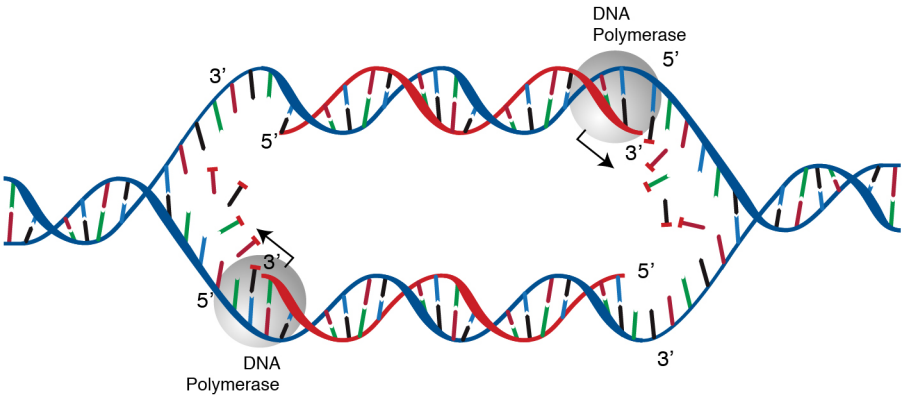


Figure 1.2: The DNA replication process by which each DNA strand produces two identical copies of itself.

which produces a complementary, antiparallel RNA strand. If the stretch of DNA is transcribed into an RNA molecule that encodes a protein, the RNA is termed messenger RNA (mRNA) which in turn serves as a blueprint for the synthesis of a polypeptide chain in ribosomes by a process called translation[25]. The polypeptide chain later folds into an active protein, that is sent out of the cell to perform structural and functional roles. RNA not only serves as a mediator between DNA and protein during gene expression but also fulfills a large number of regulatory roles in the cell[26]. Crick called the unidirectional flow of information from DNA to protein the 'Central Dogma of Molecular Biology'[27]

It was believed that the key role of DNA was to store the genetic information essential for protein synthesis, until recently, when the biological importance of non-coding sequences gained significant limelight[28]. Adding to other developments in the field, the role of chemical modifications on DNA such as methylation has been recognised[29]. These advancements imply a more multi-faceted picture of biological DNA as an active component implanted in the heterogeneous environment of the

cell, rather than a passive depot of information[30,31]. To fully appreciate the dynamical role of nucleic acids, we must extend the one-dimensional description of their structure as a sequence to include the physical context in which the sequence is stored, as polymeric DNA and RNA.

1.2.3 DNA topology

As soon as the double-helical structure of DNA was discovered, its implications for readout and replication mechanisms were studied[32,33]. One of the first questions asked was how strand separation proceeds over long distances given how intertwined the two strands are. This 'DNA replication problem' was widely discussed in the 1950s and it was presumed that melting of the double strand over large regions would exert torsional stress on adjacent parts of the strand. This would result in the system bending back on itself, building up mechanical resistance against further progression of the replication fork[34]. Although these studies were purely theoretical, they served to highlight the significance of the topological properties of DNA. Experiments performed by Vinograd *et al.* suggested that the genome of the Polyoma virus had a 'twisted circular' microscopic configuration[35], thereby implying the existence of 'topoisomers', that only differ in topological properties of the DNA strand and not in their sequence.

Many bacterial species have closed circular genomes in an underwound state. In eukaryotic cells, the genome and the nuclear matrix form many topological domains within which supercoiling is conserved[36]. Further work[36,37] has established that biological activity of DNA depends significantly on the topological properties of the genome. Many key processes like transcription, replication and recombination

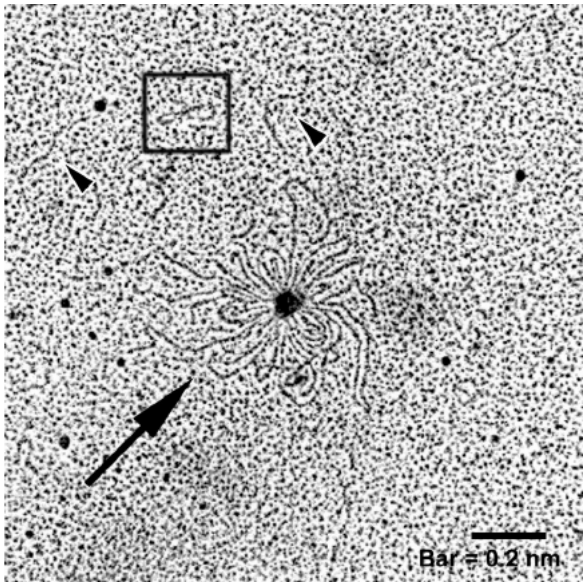


Figure 1.3: Electron micrograph of DNA isolated from E.coli. Linear DNA (arrowheads), a plasmid (box) and a rosette-like structure (arrow) were observed. Figure reproduced from Ref. [38].

in which the DNA separates into single strands, are also influenced by topological features. The stability of a duplex is depends on supercoiled it is, which also dictates the free energy barriers that need to be overcome to access the encoded sequence. It is also one of the primary factors responsible for the organisation of the genome into looped domains (see **Figure 1.3**), affecting the compaction of DNA. A class of enzymes called topoisomerases are able to control the inter-conversion between different topoisomers of DNA by over- or under-winding the strand[37], which again implies how important it is to acknowledge the topology of DNA in a biological context.

1.3 DNA Nanotechnology

The specificity of the base pairing in DNA allows us to rationally design objects made out of DNA for which the intended structure is the global minimum on the

free energy landscape. In 1982, Seeman outlined how DNA junctions could be joined to form a lattice thus giving birth to the field of DNA nanotechnology[39]. In 2006, researchers around the world started taking more interest in the field after Rothemund introduced a powerful and flexible method of assembling arbitrary two-dimensional structures by folding a long strand of DNA[40]. This approach, which Rothemund called DNA origami, was versatile and could be employed to make very large structures with thousands of DNA base pairs. Many other alternative ways to create DNA nanostructures have been established since then, including the 'DNA brick' scheme[41] and the wire-frame approach[42–44].

Because of the adaptability and robust nature of DNA origami self-assembly, it is very popular among researchers and structures produced using this technique will be a primary focus in this work. In this method, a very long (typically 7 kilo-bases) single strand of DNA, termed the *scaffold* strand folds into the desired shape assisted by many short strands, termed *staple* strands, that bind specifically to two or more non-adjacent parts of the scaffold strand (see **Figure 1.4**). Such structures are made up of anti-parallel double helices joined by *crossovers*, where staple strands cross from one region of the scaffold to another. The method has been successful in producing 2D structures[40], 3D shapes[45], systems with designated twist or curvature[46,47] and even tensegrity structures[48]. **Figure 1.5** shows some representative DNA nanostructures made using origami and non-origami methods.

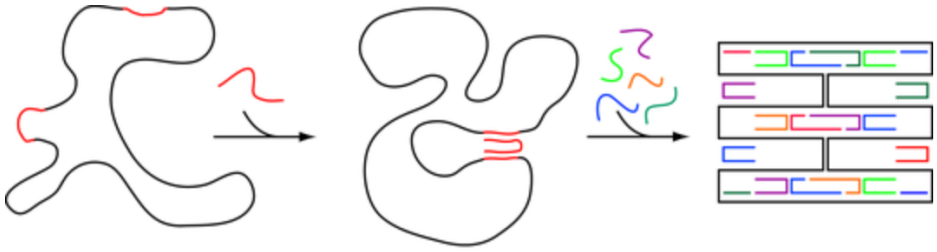


Figure 1.4: The self-assembly of a DNA origami. The picture shows a scaffold strand (black) with the red parts denoting regions complementary with the corresponding (red) staple strand that forces the scaffold regions to come close.

1.4 Studies in DNA Nanotechnology

Much progress been made in the field of DNA nanotechnology in recent years, most of which has opened new doors for further investigation. It has significant scope for development and can catalyse a better understanding of many fundamental processes underlying nucleic acid assembly. Experimental groups around the world have made efforts to carefully study DNA origami structures[46,50,51] and their assembly[52–56].

1.4.1 Theoretical Models of DNA

It has become clear that mathematical models of DNA have the potential to greatly improve our current understanding of DNA devices. The most detailed theoretical approach is to use quantum chemistry to calculate the interactions between nucleotides. Although such methods have been proposed[57–59], the high computational cost makes them unfit for large structures in solution. Classical all-atom molecular dynamics (MD) methods that have been widely used to study small DNA structures[60,61], have recently been applied to larger systems[62–65]. Since these approaches involve the modelling of every atom of DNA and the surrounding solvent, it is still a challenge to employ them to study processes that happen on scales longer

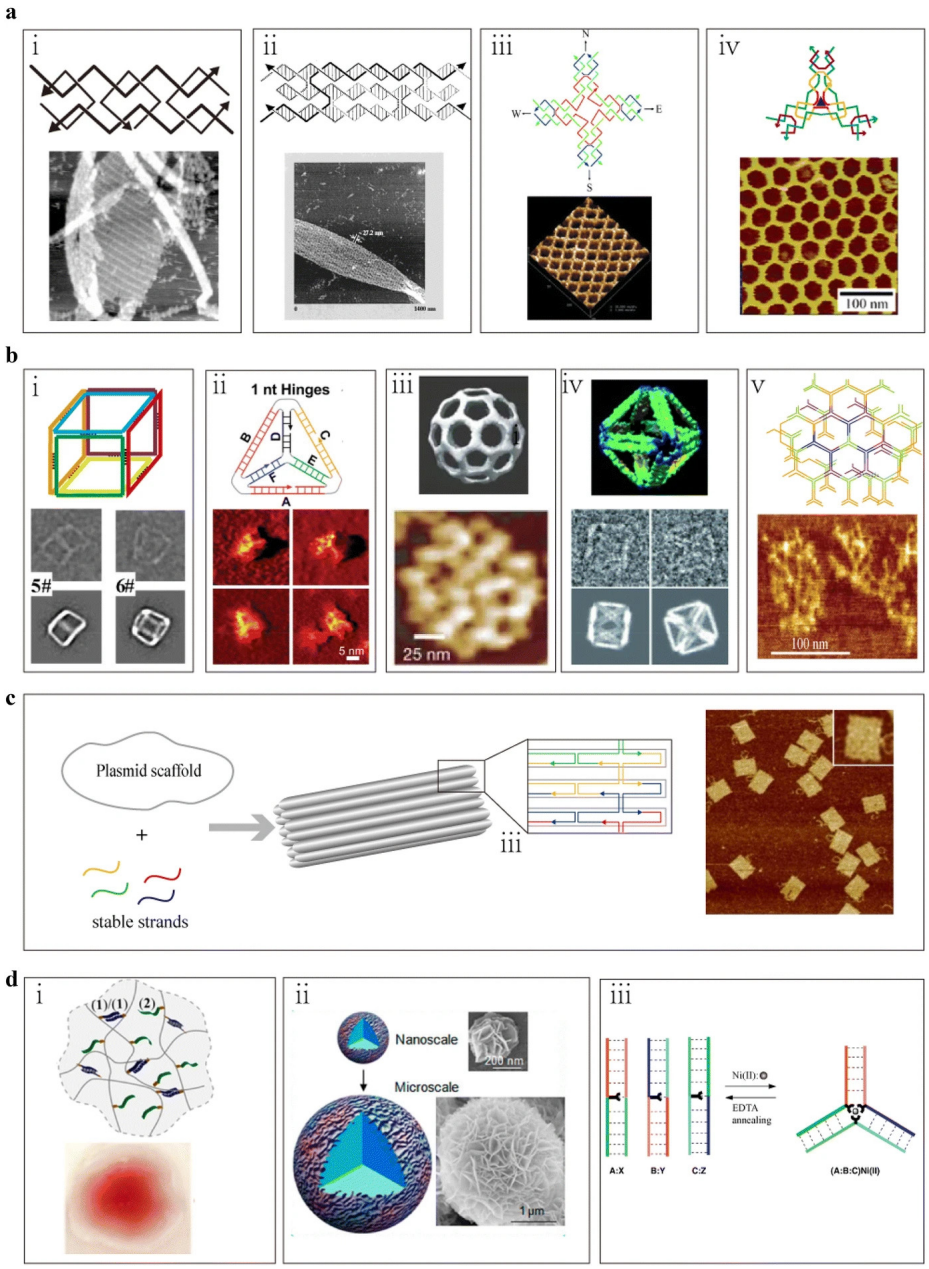


Figure 1.5: A few examples of different DNA nanostructures. (a) Some building blocks of nanostructures including (i) double-crossover (ii) triple-crossover (iii) 4x4 and (iv) three-point start tiles. (b) Bottom-up assembly of building block leads to higher order structures such as (i) DNA cube (ii) DNA tetrahedron (iii) DNA polyhedron (iv) DNA octahedron and (v) DNA dendrimer. (c) Structures made using DNA origami. (d) A few other DNA nanostructures like (i) DNA nanohydrogels (ii) DNA nanoflowers and (iii) metal-induced DNA assembly. Figure reproduced with permission from Ref. [49].

than the μs timescale[65], such as breaking or formation of base pairs. Moreover, it is still not fully clear how well these atomistic models capture the thermodynamics of DNA.

On the other hand, several theoretical techniques have been developed to investigate large-scale properties of DNA. One such model is the WLC (worm-like chain) model, that treats DNA as a continuously flexible polymer[66] characterised by a bending modulus. The model can be modified to include an extension modulus[67], a twist modulus[68] and even a coupling between the two[69,70]. Such models are useful to predict certain physical properties but of course fail to describe processes like duplex formation because the model does not resolve individual nucleotides.

'Coarse-grained' (or 'mesoscale') models represent a middle-ground between continuum and all-atom models. Such methods often integrate out many degrees of freedom of a DNA nucleotide and avoid explicit modelling of the solvent. It should be noted that such models attempt to optimise the trade-off between accuracy and computational efficiency and care must be taken when choosing an appropriate model to represent a given system. The simple picture of DNA as presented by coarse-grained approaches can be extremely useful as they not only increase the limits for time scale and number of nucleotides that can be studied but also enable us to understand the physics more easily. The process of coarse-graining can range from 'bottom-up' approaches that try to map the statistical behaviour of a more detailed model onto a coarse-grained description[71] to 'top-down' approaches like oxDNA that seek to replicate experimental values of as many relevant properties as possible[72–76].

A partially formed DNA origami can be broken up into domains where each domain is made up of a continuous double helix. Such domain-based models have

been used to probe self-assembly with reasonable success[56,77]. Another noteworthy model is CanDo[78,79] that represents DNA origami as a set of double helices treated as linear elastic rods joined by four-way junctions (acting as flexible connections between the rods).

Detailed models focusing just on B-DNA have also been proposed. Savelyev and Papoian, suggested that DNA can be treated in its double-stranded state with one bead replacing each nucleotide and each bead interacting with 11 nearest neighbours on the partner strand[80]. This model has been used to study nucleosome unravelling[81]. Members of the Nordenkiöld group have created a model where each DNA strand is represented with five beads for every two base pairs, with parameterisation achieved from all-atoms MD simulations[82]. This model has shown good agreement with experiments in reproducing the persistence length and its salt-dependence at concentrations below and equal to physiological ionic strengths. Although these models are suitable for some biophysical systems, they suffer from the caveat of only being able to study DNA in its double-stranded form which makes them inappropriate for DNA nanotechnology in which we expect to see transitions between ssDNA and dsDNA along with many more complicated structural motifs.

The Dorfman group put forth a model[83] that is capable of representing both single and double-stranded DNA, focusing particularly on Hoogsten bonding, with each nucleotide having three interaction sites. Around the same time, Araque *et al.* introduced a lattice-based, one-bead-per-nucleotide model of DNA[84] that allowed the study of hybridisation in complex systems. While the model was able to portray DNA hybridisation with decent accuracy, it compromised on structural features that are essential for the study of nanostructures.

The Plotkin group have invented a three-sites-per-nucleotide model[85] that has been employed to study some structural and mechanical properties of DNA as a function of temperature, salt concentration and sequence. Recently, the Aksimentiev group proposed a model of ssDNA, with an interaction site on each sugar/base group and phosphate group[86]. Their approach was able to reproduce the force-extension curves for ssDNA observed in experiments. Additionally, a coarse-grained RNA model called HiRe-RNA has been extended to represent DNA systems. The HiRe-RNA model uses a relatively detailed rendering with 6 to 7 interacting beads per nucleotide and was used to reproduce melting curves for short DNA duplex[87].

One of the most appropriate models put forward to investigate a broad array of DNA nanostructures is the 3SPN model[72] that represents DNA with three independently mobile interaction sites per nucleotide. This has been used to study nucleosome unwrapping[88] and duplex hybridisation[89]. A recent development in the field of coarse-grained molecular dynamics is the protein-DNA model proposed by Lu *et al.*[90] that combines the latest 3SPN model with the AWSEM protein model[91]. This model should enable the investigation of large protein-DNA constructs that are fundamental to many cellular processes. Although the latest version of the 3SPN model describes the basic properties of DNA well, it hasn't been tested on large DNA nanostructures.

1.5 Single-molecule manipulation methods

Initial attempts to decipher the structure and function of DNA used classical chemical bulk assays such as gel electrophoresis, optical absorption, electron microscopy

or X-ray diffraction. Most of these experimental methods gave information about the behaviour of an ensemble of molecules with limited details about the state of individual molecules in the population. The ensemble averaging carried out by such methods makes it tough to see individual transformations, which is specifically of interest for DNA as it occurs in small copy numbers *in vivo*. Such methods also cannot accurately predict the conformational dynamics of a biomolecule because of their inability to distinguish between static and dynamical disorder in a collection of molecules[92].

In contrast to bulk methods, single-molecule experiments provide a more detailed picture of the components of the system under study. They are able to predict statistical distributions and not just the average value of any property of the molecule. Single-molecule assays not only produce information about the structure and dynamics but also allow us to calculate mechanical properties of biomolecules *in vitro* by active manipulation. The methods of single-molecule manipulation described in the following sections will form a basis for the results presented in this thesis, as well as a guide to the design of new DNA-based force sensors.

1.5.1 Atomic Force Microscopy

Atomic Force Microscopy (AFM) was initially developed as a method to study the structure of surfaces at the atomistic level, but has been employed in the study of DNA from its early stages of development[93]. The technique is based on a cantilever with an ultra-thin tip that is scanned over a surface, with its movement monitored by the deflection of a laser beam. Since AFM is able to achieve a very high spatial resolution, defects in the DNA double-strand such as denaturation bubbles[94,95] or

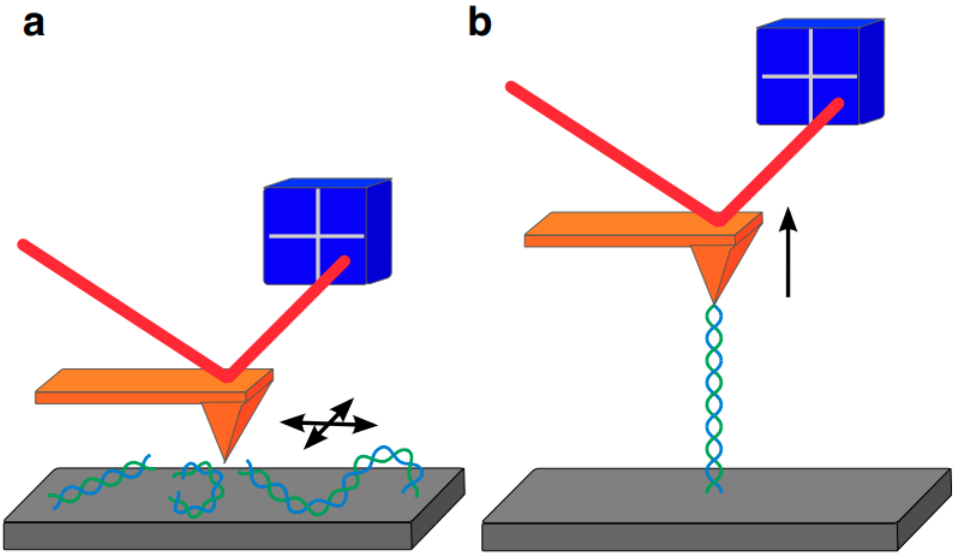


Figure 1.6: A schematic showing the operation of an Atomic Force Microscope. The cantilever (orange) exerts minute forces on the molecule, whose response leads to a change in the position of the cantilever that is recorded when a laser beam (red) is reflected onto a quadrant photodiode (blue). (a) A topographical image is obtained when the cantilever scans the target molecule (b) A DNA molecule tethered to a glass coverslip. Figure reproduced from Ref. [100].

cruciforms[96] along with the recruitment of DNA-binding proteins such as RecA[97] can be visualised directly. AFM can also be used to image DNA or RNA at the sub-nanometer scale[92,98] as well as for force spectroscopy of single molecules[92,99]. Recent advancements have been able to improve the time-resolution of AFM[98] which previously suffered from very long image acquisition times. AFM studies can be carried out in solution[92] or on molecules attached to a surface, with results being interpreted accordingly. **Figure 1.6** shows a simplified representation of AFM.

1.5.2 Optical and Magnetic tweezers

Molecular tweezers have served as an important tool for understanding the mechanical response of biomolecules to external forces or torques since the early 1990s. The most common primary designs are magnetic and optical tweezers although

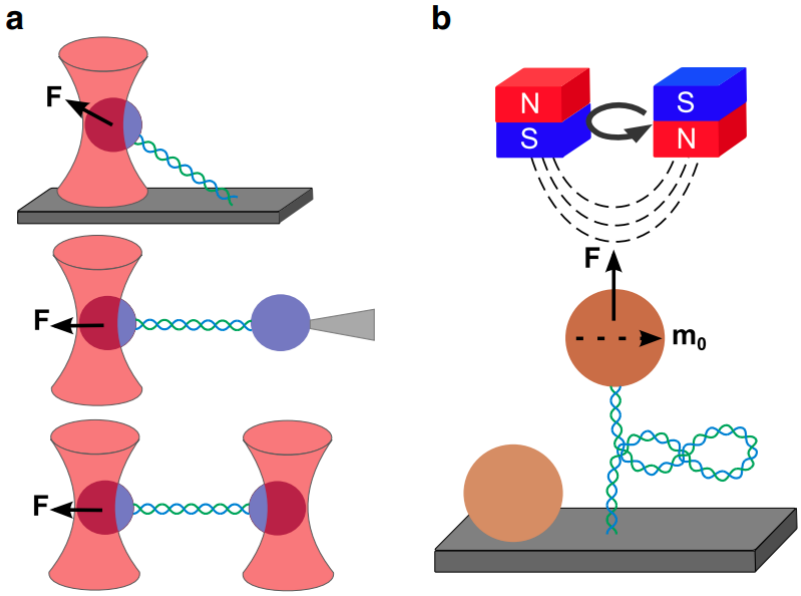


Figure 1.7: Two classes of molecular tweezers. (a) Examples of a typical optical tweezer setup. A DNA molecule (blue/green) is chained between an optically trapped dielectric bead (violet) and either the glass substrate (top) or another bead held on a micropipette tip (middle) or a second optical trap (bottom). The black arrow shows the direction of the force acting on the molecule. (b) A typical magnetic tweezer experiment showing a magnetic bead trapped between a pair of permanent magnets. The solid arrow represents the direction of the force (F) on the bead and the dashed arrow indicates the direction of the induced magnetic moment (m_0). The induced moment can provide the torque required for supercoiling resulting in plectonemic loops. Figure reproduced from Ref. [100].

for specific uses, many other alternative forms of molecular tweezers have been developed.[101–104]

Optical Tweezers

During optical tweezer experiments, the molecule of interest is clamped between dielectric beads that are then trapped in a highly focused laser beam (as shown in **Figure 1.7a**). The position of the beads and the traps are recorded and the force is calculated from the displacement of the bead from the trap center. It is possible to hold the molecule in a state of constant end-to-end distance or constant

force by shifting the focal plane of the lasers[92,101]. Optical tweezers are excellent to exert unidirectional linear forces and have been employed to study the stretching behaviour of DNA[105,106], the unfolding of secondary structures like DNA or RNA hairpins[107,108] and the denaturation behaviour of RNA[109]. A few recent modifications[110] have been made to the original design that allow the determination of torque on a molecule by angular trapping of optical beads[111].

Magnetic Tweezers

Magnetic tweezers were first used by Strick *et al.* to study the elastic response of supercoiled phage λ DNA[112–114]. Since then, they have been widely used by many to understand the elastic and mechanical properties of DNA for different strand lengths, salt concentrations and pH conditions[115–119]. In a magnetic tweezer based assay, a single molecule is tethered to a surface with one of its ends, and to a magnetic bead with the other. The magnetic bead's position can be changed by moving the magnets situated above it. The distance between the magnetic bead and the magnets determines the stretching force experienced by the molecule while the rotation of the magnets determines the torque applied on the molecule. A simplified diagram of a magnetic tweezer is shown in **Figure 1.7b**. The distance between the bead and the surface is obtained from the bead's diffraction pattern. Transverse fluctuations determine the force acting on the bead[120].

Magnetic tweezers have turned out to be a very flexible method for the study of a large array of nucleic acid systems, especially where the response to twist of interest. They have been employed to study the kinetics of cruciform formation[121], the transition of B-DNA to Z and L forms[122] and the denaturation of the double

strand[117–119]. Lipfert *et al.* managed to perform a methodical investigation of the mechanical properties of dsRNA using magnetic tweezers[123]. Magnetic tweezers have also been applied for the microscopic investigation of nucleic acid processing enzymes, such as in mechanisms of transcription[124], or the activity of helicases[125] and topoisomerases[126,127].

Molecular tweezers, although very versatile, suffer from a few limitations. As a stand-alone device, a molecular tweezer assay can only determine a global property of the molecule, such as the end-to-end extension, but don't allow the stress-induced changes to the local structure to be observed. They also suffer from a limited time resolution due to viscous drag of the solvent and the inertia of the attachments. Thus, data from a molecular tweezer experiment is often complimented with other experimental results or theoretical calculations, if local structural changes are of particular interest. Experimentally, fluorescence-based probes can provide more detailed spatial information about the molecule.

1.6 Fluorescent Techniques

Fluorescence Microscopy

Fluorescence microscopy is a common method to investigate the changes in the general molecular conformations of DNA strands. It can be integrated with single-molecule manipulation methods such as molecular tweezers (explained in **Section 1.5.2**). A typical fluorescence microscopy involves the tagging of a molecule of interest with a covalently-linked or intercalating dye that is then excited by a light source and observed with the help of a microscope. Using this technique, one is able

to monitor the overall configuration of the entire molecule as a function of time. This technique has been applied to the investigation of local denaturation observed in DNA upon overstretching[128,129], and to studies of protein-DNA interactions by fluorescent labelling. The Kowalczykowski lab was able to monitor the kinetics of RecA assembly on DNA[130] and Van Oijen *et al.* were successful in deciphering the dynamics of the bacterial replication machinery[131,132] using fluorescence miscroscopy.

FRET

FRET[133] stands for Förster Resonance Energy Transfer and provides a means to measure distances within molecules with a precision of just a few nanometers. In FRET, energy is transferred from one molecule (donor) to another (acceptor) in a non-radiative manner (dipole-dipole coupling). The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between the donor and the acceptor. This makes FRET extremely sensitive to minute changes in distance[134].

Using FRET, it is possible to understand the underlying conformational dynamics and reveal the transition between different microscopic states. Ha and colleagues studied a collection of systems using the FRET method, such as the cyclisation of DNA strands due to its intrinsic bendability[135] and the conformational diversity of human telomeric DNA[136]. FRET, combined with other single-molecule techniques, has also been used to construct the two-dimensional reaction landscape of four-way DNA junctions[137]. Cordes *et al.* have used FRET to observe the opening and closing of dsDNA during transcription[138]. The FRET methodology can be extended to track multiple concurrent fluorescent interactions[139,140], allowing one to simultan-

eously study many structural changes in the probed molecule, thus supplying more intricate local information about its conformation.

Although most work till date has used fluorescence techniques *in vitro*, they have the potential to be used in living cells[141].

Chapter 2

Methods

Computer simulation uses mathematical modelling to predict the behaviour of, or the outcome of, a physical system. Such simulations can be used to get insights into new technology and to study systems that are too complex to investigate in experiments. For example, a system with different components is much easier and more cost-effective to set up on the computer rather than in an experiment.

2.1 Molecular Dynamics

Molecular Dynamics (MD) is an N-body simulation technique which works by numerically integrating Newton's laws of motion over a small time step, using potential energies and forces calculated from a model, in this case, the oxDNA model. Initially it may seem impossible to perform MD simulations in ensembles other than the microcanonical (constant-NVE) ensemble without losing ergodicity. To get around this problem two of the paths that have been proposed are: (1) mixing Newtonian MD with certain Monte Carlo moves to dynamically simulate other ensembles and (2) reformulating the Lagrangian equations of motion of the system. Presently, the MD simulations are carried out in a canonical ensemble, conserving the number of particles (N), the volume (V) and the temperature (T). To achieve constant temperature conditions in the system, we couple it to a large heat bath. Under those conditions, the probability of finding the system in a given energy state is given

by the Boltzmann distribution and, for a classical system, the Maxwell-Boltzmann distribution in terms of the velocity v is as follows:

$$P(v) = \left(\frac{\beta m}{2\pi}\right)^{\frac{3}{2}} \exp\left[-\beta m v^2/2\right], \tag{2.1}$$

where β is the inverse thermodynamic temperature and m is the molecular mass of the particle. We then obtain the simple relation between the imposed temperature T and the (translational) kinetic energy per particle:

$$k_B T = m\left\langle v_\alpha^2 \right\rangle, \tag{2.2}$$

where k_B is the Boltzmann constant and v_α is the α^{th} component of velocity.

In the constant temperature method proposed by Andersen[142], the system is coupled to a heat bath that imposes the desired temperature. The coupling to a heat bath is represented by stochastic impulsive forces that act occasionally on randomly selected particles. These stochastic collisions with the heat bath can be considered as Monte Carlo moves that transport the system from one constant energy shell to another. All accessible constant energy shells are visited according to their Boltzmann weights during these stochastic collisions. The frequency of such collisions determines how strongly the bath is coupled to the system. Assuming that successive collisions are uncorrelated, we can use a Poisson-like function to describe the distribution of the time intervals between two successive collisions:

$$P\left(t;\nu\right) = \nu \exp[-\nu t], \tag{2.3}$$

where ν is the frequency and $P(t;\nu)dt$ is the probability that the next collision will take place in the interval [t,$t + dt$]. The velocity of the particle that is selected to undergo a collision is drawn from the Maxwell-Boltzmann distribution corresponding to any desired temperature T. All other particles are unaffected by this collision.

The thermostat used for most of the simulations carried out in this study is a Langevin thermostat where, at each time step all particles receive a random force and have their velocities and angular momenta lowered using a constant friction. The average magnitude of the random forces and the friction are related in a particular way, which guarantees that the fluctuation-dissipation theorem is obeyed, thereby guaranteeing NVT statistics. The equation of motion of the i^{th} particle is given by:

$$m\ddot{r_i} = -\nabla_i U - m\Gamma\dot{r_i} + \mathbf{W_i}(t), \tag{2.4}$$

where Γ represents the coefficient of friction with units of τ^{-1} and $\mathbf{W_i}$ is a random force acting on all particles and uncorrelated in time. As the only significant particles interactions in this system are short-range, a Verlet list is used for each particle so that only these interactions are considered.

For the ANM-oxDNA simulations, described later in **Chapter 4**, the 'John' thermostat[143] was used. In this Andersen-like algorithm, the velocity and momentum of every particle is refreshed with a given probability from the Maxwell distribution at the specified temperature. On time scales longer than the time between two consecutive refresh events, the John thermostat emulates Brownian dynamics.

In this study, only MD has been used to simulate DNA nanostructures, although

Monte Carlo (MC) simulations could also be used to investigate the equilibrium behaviour of these systems. The MD algorithm for many-particle systems is relatively trivial to parallelise as compared to the MC algorithm, leading to the former being preferred for our simulations. This parallelisation can be achieved by separately computing the forces for each particle or pair of interacting atoms (the latter is typically more efficient for our examples[144]). A GPU code that is able to run MD simulations of DNA nanostructures using the oxDNA model had already been developed; this drastically reduces the compute time for large DNA origami nanostructures, thereby allowing us to run the simulations for sufficiently long to get well-converged results.

2.2 The oxDNA Model

2.2.1 Coarse-graining

All theoretical calculations involve a compromise between detail and computational efficiency. All-atom simulations of DNA origamis have been performed[145], but they represent massive simulations and cannot be carried out routinely. On the other hand, continuum models treat the DNA duplex as a uniform medium, and can only provide information about long length-scale DNA behaviour. DNA coarse-graining is somewhere in between the two extremes and can provide information about melting of a duplex or hybridisation depending upon the level of coarse-graining.

Models other than nucleotide-level or base-pair-level models are either too computationally expensive for our task due to too little coarse-graining or insufficiently detailed due to too much coarse-graining. Rigid base-pair models suffer from the

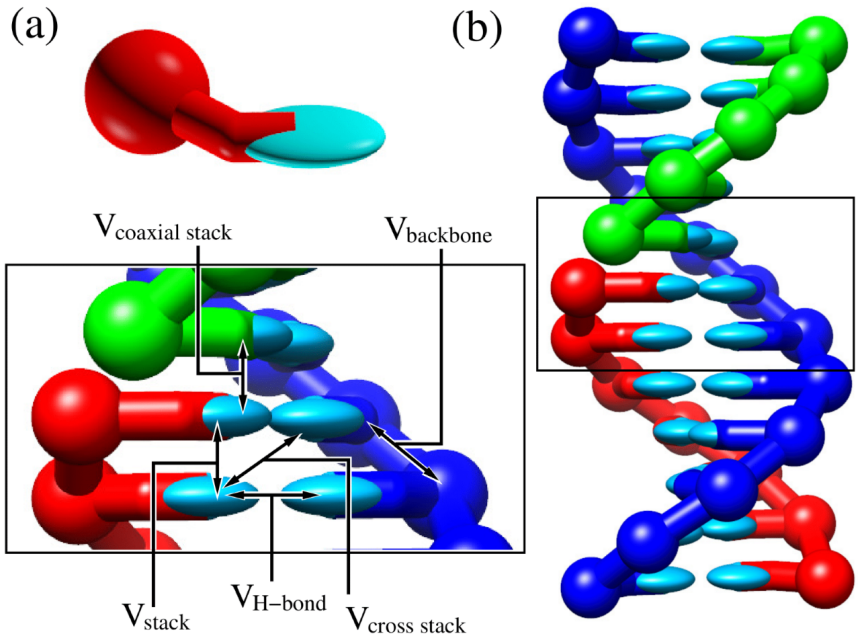


Figure 2.1: (a) A representation of a rigid oxDNA nucleotide. (b) A 12 base pair DNA double helix as represented by oxDNA. The inset shows the different type of interactions that are accounted for in the oxDNA model, namely, hydrogen-bonding between complementary bases, (coaxial) stacking interactions between bases, cross-stacking interactions between diagonally opposite bases in a double helix, and a FENE potential between backbone sites, as well as excluded-volume repulsion associated with the backbone and base sites. Figure adapted from Ref. [147]

inability to represent both single and double-stranded regions effectively, and although they are only marginally more coarse-grained than nucleotide-level models, they are unsuitable for studying junctions which contain both single and double-stranded regions. This makes a nucleotide level model the best fit for studying these structures and oxDNA is currently the best parameterised and most well-tested model for DNA nanostructures. Currently, the oxDNA model is available as its original stand-alone simulation program which can be used for both Monte Carlo and molecular dynamics, and also through LAMMPS[146], a widely-used molecular dynamics package.

2.2.2 OxDNA and oxDNA2 models

OxDNA is specifically parameterised for the study of assembly transitions, especially the thermodynamics of single-strand stacking, duplex hybridisation and hairpin formation. OxDNA thermodynamics is parameterised on the Santa Lucia model, which assumes that the stability of a base pair is dependent just upon its neighbouring base pairs. Each nucleotide, in oxDNA, is represented as a rigid body with the following interactions: backbone, base repulsion, stacking and hydrogen-bonding (as shown in **Figure 2.1**). The interactions between the nucleotides depend on their global orientation and not just the position of their interaction sites. A pair of vectors denoting the notional plane of the base and the direction of the hydrogen bonding interface are used to define the dependence of an interaction on orientation. This allows oxDNA to represent the coplanar base stacking interactions, the linear hydrogen bonding interactions and the edge-to-edge character of the Watson-Crick base pairing[148].

The oxDNA2 model provides a better description of the helix geometry, including major and minor grooving, and introduces a basic model of salt dependence based on a Debye-Hückel model, which uses a simple activity coefficient to account for the non-ideal nature of the solution. oxDNA2 also has provisions for improved large-scale structure prediction and includes reparameterisation of AA and TT stacking strengths to allow for the stronger stacking of poly-A strands and greater flexibility of poly-T strands. oxDNA2, being able to successfully predict melting temperatures and the elastic properties of DNA, can be used to perform simulations of DNA nanostructures at physiological salt concentrations as well as at the higher salt con-

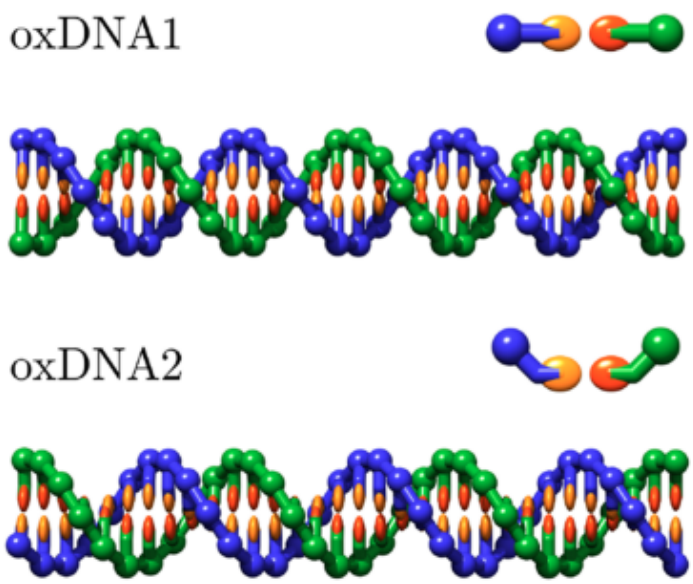


Figure 2.2: Snapshots of a DNA duplex in oxDNA1 (top) and oxDNA2 (bottom), including a cross sectional view of a base pair. Distinct major and minor grooves are apparent in oxDNA2 as compared to oxDNA1 where the grooves are symmetric. Figure reproduced with permission from Ref. [149]

centrations often used in nanotechnology experiments. At a salt concentration of 0.5 M Na^+, the pitch is set to 10.55 bp/turn and the helical rise to 0.347 nm. The correct amount of twist at nicks and junctions also required careful parameterisation to accurately reproduce the structure of DNA origamis. The potential in the oxDNA2 model can be written as:

$$V = \sum_{\langle ij \rangle} \left[V_{\text{backbone}} + V_{\text{stack}} + V_{\text{exc}}^{'} \right] + \sum_{i,j \notin \langle ij \rangle} \left[V_{\text{HB}} + V_{\text{cr.st}} + V_{\text{cx.st}} + V_{\text{exc}} + V_{\text{DH}} \right], \quad (2.5)$$

where:

- V_{backbone} is a finitely extensible non-linear elastic (FENE) spring with an equilibrium length of 0.64 nm, representing the covalent bonding along the sugar-

phosphate backbone.

- V_{stack} represents the tendency of bases to form coplanar stacks. It is a Smoothly Cut-Off Morse Potential (SCOMP) with a minimum at 0.34 nm modulated by angular terms favouring parallel normal vectors. This causes the coplanar stacking to have separations shorter than the backbone distance; this is the origin of helicity of dsDNA in oxDNA. Left-handed helices are averted by reducing this term to zero for increasing left-handed twist.

- V_{exc} and V_{exc}' quantify the excluded volumes of nucleotides by employing Lennard-Jones repulsive potentials. Bonded neighbours do not include this repulsion as their separation is regulated by the FENE spring.

- V_{HB} represents the hydrogen bonds that lead to base pairing. It is a SCOMP with angular terms which favour the colinear alignment of all four backbone and hydrogen bonding sites and anti-parallel normal vectors. It is set to zero for non-complementary base pairs.

- $V_{cr.st}$ and $V_{cx.st}$ represents cross-stacking and coaxial stacking interactions respectively. Mathematically, they are smooth quadratic wells and increase the stability of the duplex.

- V_{DH} is the Debye-Hückel term which represents the electrostatic interactions between the phosphates.

In **Equation 2.5**, the first summation is taken over all pairs of nucleotides that are nearest neighbors on the same strand and the second summation is over all remaining pairs. The oxDNA model[150] has two parameterisations: average-base

and sequence dependent. In the average-base model, interactions between bases are taken to be the average, while in the sequence-dependent model, the hydrogen bonding and stacking strengths depend on the identities of the relevant nucleotides. Solvent molecules are not explicitly considered in oxDNA but their effect is implicitly incorporated in the effective potentials of **Equation 2.5**. After careful parameterisation, oxDNA is able to capture the thermodynamics of hybridisation very accurately whilst being extremely computationally efficient.

For convenience, the oxDNA simulation code uses its own internal unit system (often referred to as simulation units). The values for parameters in the input file typically need to be in these units. Similarly, input and output configurations are also in simulation units. As with any coarse-grained model, there are some features of real DNA that cannot be fully captured by oxDNA and which can thus potentially negatively affect the results for certain properties. These include:

- Four-way DNA junctions in their stacked form have been experimentally shown to prefer a right-handed twist angle between the helices but, in oxDNA, they attain a left-handed form. This is not a big problem for origami modelling as most junctions are constrained to adopt an anti-parallel configuration.

- oxDNA gives a higher extensional modulus for DNA compared to the actual molecule. A choice was made during parameterisation of the model to prioritise the accurate modelling of bend and twist over the extensional modulus. This shortcoming may affect mechanical responses of origami that couple to the stretching of individual helices.

- The Debye-Hückel potential used to represent electrostatic interactions in the

oxDNA model with explicit dependence on the ionic strength of the solution is relatively simple and unable to capture ion-specific effects. The model has been fitted to reproduce the $[Na^+]$ dependence of the hybridisation thermodynamics for $[Na^+] > 0.1M$.

- Due to the relative lack of high-quality physical chemistry data, the parameterisation of the coaxial stacking interaction in oxDNA should be handled with care. Currently, the interaction has no sequence dependence, and its magnitude represents a compromise between experimental data that could not be simultaneously reproduced[150]; further, little is known about the orientation dependence of this interaction. This shortcoming may affect the modelling of blunt-ended stacking of helices that is being increasingly used to mediate multi-origami assembly.

2.2.3 Conversion of design to oxDNA format

Any origami design created using a computer-aided design software, needs to be converted into the oxDNA format before it can be simulated. The most common program is TacoxDNA[151] which can either be used in an online browser-based interface or through its suite of Python scripts. A slightly outdated python script called 'cadnano_interface.py' available in the UTILS folder under the oxDNA software can be used to convert designs made using caDNAno[45] to oxDNA by passing the relevant **.json** file, the type of lattice (he for helical and sq for square) and the box size (in simulation units, assuming a square box). The conversion tools generally output two files: an oxDNA configuration file (**.oxDNA**, **.dat** or **.conf**) specifying the positions and orientations of the nucleotides and an oxDNA topology file (**.top**)

specifying the sequence and the way in which nucleotides are covalently bonded to each other along the DNA backbone. The 'cadnano_interface.py' also generates a 'v2n' file which holds information about the mapping between virtual indices in caDNAno and nucleotide indices in oxDNA.

2.2.4 Relaxation of initial geometry

The configuration generated in the above manner may have nucleotides with overlapping excluding volumes or long backbone bonds which would result in extremely large forces and cause our simulation to fail if we were to start simulating immediately. To avoid such a scenario, the system needs to be 'relaxed' before we start our simulation. Relaxation is performed to get rid of any local aberrations in the system. This is usually carried out in two steps using a modified backbone potential.

The standard oxDNA backbone potential is a FENE potential which diverges approximately at 0.873 nm. The purpose of the modified potential is both to remove this divergence and to ensure that the forces resulting from stretched bonds do not damage (e.g. by breaking base pairs) the origami structure. The modified potential has the form:

$$V_{\text{mod}} = \begin{cases} V_{\text{FENE}} & \text{for } r \leq r_{\text{max}} \\ Ar + B\log r + C & \text{for } r > r_{\text{max}}. \end{cases} \tag{2.6}$$

where V_{FENE} is the original backbone potential, r_{max} is the distance at which the force due the FENE potential is equal to F_{max} (the value of F_{max} is set in the input file using **max_backbone_force**). A corresponds to the limiting value of the force at large r and is set using the variable **max_backbone_force_far** in the

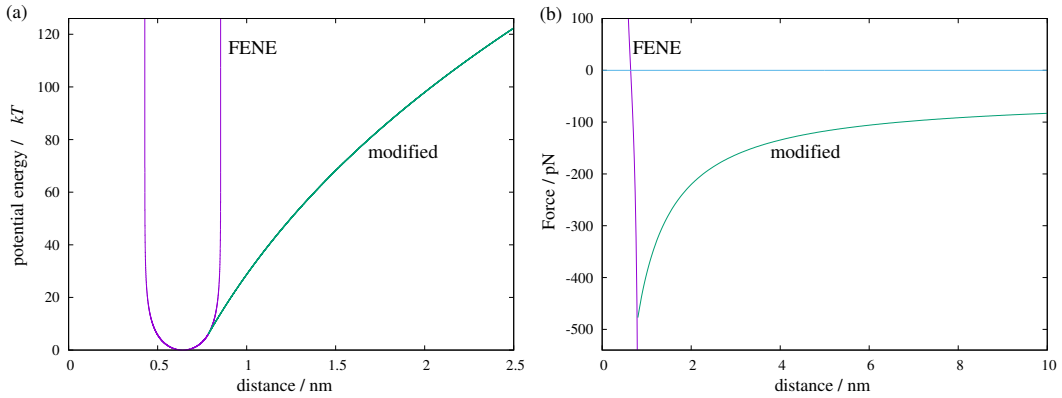


Figure 2.3: (a) The FENE backbone potential and its modified form for parameter values $F_{\max} = 10$; $A = 1$ (values in simulation units; equivalent to 486 pN and 48.6 pN). (b) The forces due to these potentials.

input file; B and C are chosen so that the modified potential is continuous and differentiable at $r_{\max}$. V_{mod} increases relatively gently as r increases with the force tending to A at sufficiently large r, as shown in **Figure 2.3**.

The first step in the relaxation process is running a minimisation calculation for a few thousand steps by specifying **sim_type = min** in the input file. This is usually sufficient to remove all particle overlaps and those overstretched bonds that are too long by only a relatively small amount (on the order of a few nanometers). It can also help nucleotides that are designed to be base paired to correctly orient themselves to fully base pair. The minimisation runs on a single CPU core but even for a full-size origami it should take no more than a few minutes.

The second stage of the relaxation process involves a standard molecular dynamics simulation with a modified backbone potential which allows for relaxations that require motion on a larger scale. While relaxing origamis, one needs to be careful to avoid topological entanglements which arise due to the layout of the structure in the design. We also need to make sure that no part of the origami suffers from base pair breaking due to transient internal stresses.

For a majority of our systems, we carried out a minimisation for 10^4 steps followed by a two-fold relaxation, first for 10^5 steps with **max_backbone_force = 5.0** and then for 10^8 steps with a **max_backbone_force = 10.0**. Following Rothemund, experiments on DNA origami often use a solution containing 12.5 mM Mg^{2+}. As noted earlier, oxDNA's description of electrostatics is simple and cannot capture ion specific effects. In terms of DNA origami structure, it has been found that those experimental solution conditions map on reasonably well to $[Na^+] = 1M$ in oxDNA. The salt concentration is always set to 1.0M Na^+ for all relaxations and simulations unless specified otherwise.

2.2.5 Simulation

To get a good picture of the origami's conformational space it is imperative that we carry out comprehensive sampling. The length of our simulation would depend upon two factors: the time required to equilibrate the system (i.e. achieve a representative starting configuration for sampling) and the time to appropriately sample the configuration space.

An important part of any simulation is an equilibration period that allows a representative configuration of the equilibrium ensemble to be achieved. Equilibration can be checked by plotting properties of the system as a function of time. An appropriate sampling time depends on the time scales associated with fluctuations in the slowest degrees of freedom (and also whether the property of interest depends on these degrees of freedom). Absolute time scales associated with coarse-grained simulations, such as those using oxDNA, should be interpreted with caution. Typically, time scales are reported using the time unit defined by the units of the basic

oxDNA interactions. However, coarse-graining reduces the time scale separation between microscopic motions and diffusion times; thus, the effective time may be considerably larger and it may be more appropriate in certain cases to report times based on a mapping of diffusion times.

2.3 ANM-oxDNA: A nucleic acid-protein hybrid model

Over the past few years, many scientists have developed particular interest in hybrid DNA-protein nanotechnology that combines the already advanced design strategies of DNA nanotechnology with functional proteins. The amalgamation of the two molecules in nanotechnology will have many new applications particularly in the field of novel bioimitec materials[152]. Example advances in the hybrid nanotechnology field include DNA-protein cages[153], a DNA nanorobot with nucleoin aptamer[154] and peptide-directed assembly of large nanostructures[155]. The oxDNA model, as introduced in **Section 2.2** can only allow the representation of nucleic acids, thus limiting its applicability in the present context. Efforts have been made to develop a coarse grained model for DNA-protein complexes[91,156–159] with particular focus on nucleosomes[160,161] but until recently there was no efficient tool at the same level of coarse-graining as oxDNA that will allow the investigation of arbitrary structures. The ANM-oxDNA (and ANM-oxRNA) hybrid model[162] is ideal for the design and study of large nanostructures made up of nucleic acids and stable folded proteins.

The ANM-oxDNA model comprises two separate coarse-grained models, oxDNA2 (see **Section 2.2.2**) for nucleic acids and an Anisotropic Network Model (ANM) for proteins. In ANM, a protein with a known structure is made up of beads (repres-

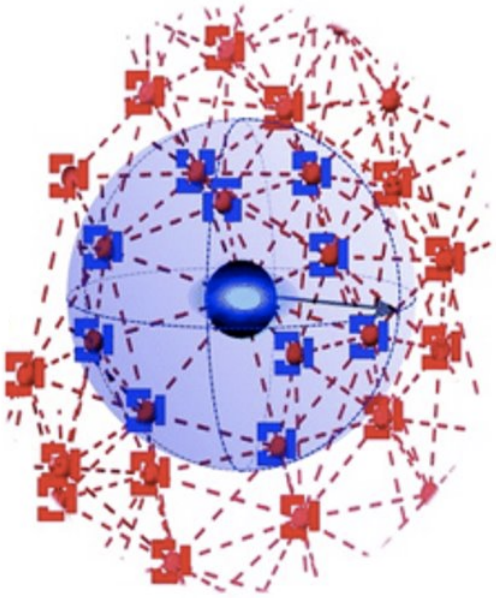


Figure 2.4: ANM representation of a protein. The blue sphere encloses all residues within the cutoff distance ($r_{\max}$) from the central residue (black circle). The interacting residues are shown in in blue whereas the non-interacting residues are shown in red. Figure adapted with permission from Ref. [162]

enting amino acid residues) connected by springs that maintain the folded shape of the protein (see **Figure 2.4**). Such a representation, while being extremely efficient and relatively simple, provides a decently unerring picture of the structure of the system.

2.3.1 Protein Model

Each residue in the protein is represented by just the position of its α-carbon. The bond between residues i and j is represented by a harmonic potential of the form:

$$V_{ij}\left(r^{ij}\right) = \frac{1}{2}\gamma\left(r^{ij} - r_0^{ij}\right)^2. \tag{2.7}$$

where r_0^{ij} is the equilibrium distance and γ is the spring constant. All residues within a certain cutoff distance $r_{\max}$ from a particular residue, are considered 'bonded' (i.e.

they interact with a potential of the form as described in **Equation 2.7**). The values of r_0^{ij} are set to the equilibrium distance between the corresponding α carbons in the PDB structure and the values of γ are chosen to best fit the B-factors of the original PDB structure. B-factors or Debye-Wallace factors are indicative of the relative vibrational motion of different parts of the structure, with more ordered atoms having low B-factors and vice versa. During simulations, the parameters file stores the values of r_0^{ij} and γ for each interacting pair of protein particles.

To account for excluded volume effects in the ANM model, a repulsive Lennard-Jones-like potential that acts on all protein-protein and protein-DNA pairs is used. Given any two particles at a distance r, their excluded volume interaction is defined as:

$$V_{\text{exc}}(r) = \begin{cases} 4\varepsilon\left(-\frac{\sigma^6}{r^6} + \frac{\sigma^{12}}{r^{12}}\right) & r \leq r^* \\ b\varepsilon\,(r - r_c)^4 & r^* < r < r_c \\ 0 & r \geq r_c. \end{cases} \tag{2.8}$$

The excluded volume diameter r_c is chosen to minimise deviations between analytical and experimental B-factors. The parameters b and r^* are calculated to make $V_{\text{exc}}(r)$ differentiable.

Like all analytical methods, the ANM model suffers from a few drawbacks, of which the most significant one is called the 'tip effect' wherein large spikes are observed in the analytically computed B-factors because of certain residues being under-constrained. Some of these peaks disappear on increasing the value of $r_{\max}$ but this affects the computational time negatively. Nevertheless, the model is able to deliver a semi-quantitative agreement with the measured data, thereby representing

a protein in a manner that is simple yet accurate enough for roughly capturing its shape and thermal fluctuations, as needed for DNA–hybrid nanotechnology systems.

2.3.2 Extensions to the model

One of the extensions to the ANM model, called ANMT, includes an extra harmonic term in the potential for torsion and bending modulations between two interacting particles. There is also a possibility of having a heterogenous ANM (hANM) representation where every pair of interacting particles has a different γ value.

2.3.3 Protein-nucleic acid interactions

In the current version of the ANM-oxDNA model, protein residues and nucleic acids interact merely by excluded volume interactions without any electrostatics. The base and backbone sites on each nucleotide interact with surrounding protein particles using separate excluded volume potentials. In reality, the DNA backbone being negatively charged can show a strong attraction for positively charged regions of a protein; this is one common source for DNA-protein interactions[163]. To mimic these interactions in the model, explicitly defined spring potentials, as described by **Equation 2.9**, may be used between designated protein particles and nucleotides.

$$V_{\text{spring}}(r) = \frac{1}{2}k\left(r - r_0\right)^2. \tag{2.9}$$

In our simulations of the nucleosome, mutual traps have been used to imitate the histone-DNA electrostatic interactions.

Chapter 3

DNA-Polymer force clamp

The task of characterizing the response of biomolecules to external mechanical forces has been made easy with the help of highly precise single-molecule techniques like AFM[164,165], optical tweezers[101,166] and biomembrane force probes[167,168]. Although highly precise, single-molecule techniques are limited by low throughput[169] and fail to provide a detailed picture of the chemical changes occurring within a target molecule as it is subjected to an external force. Moreover, the incompatibility of single-molecule force spectroscopy (SMFS) with greater than pM concentration of the target molecule makes it unfit for chemical analysis. Although it is possible to apply SMFS parallelly[170–172], these approaches are still unable to manipulate more than ∼100 molecules at a time. It is possible to further parallelise the force manipulation and to trigger the application of force for analysis using time-resolved spectroscopy[164,173]. These methods work by introducing a sudden perturbation (ns to μs), like a thermal pulse, while collecting spectroscopic signatures and are employed frequently to study the conformational dynamics of ensembles of molecules. However, the low throughput and low temporal resolution of conventional SMFS methods impede the marriage of force spectroscopy and chemical analytical tools.

In this chapter, we discus the design and properties of a DNA Origami-Polymer force clamp (OPFC). The system is made up of a DNA origami substrate onto which the target molecules (in this study, DNA hairpins) are attached and a thermo-

responsive polymer sphere, which when heated above its lower critical solution temperature (LCST) expels solvent and shrinks (as shown in **Figure 3.1**). In contrast to a conventional polymer force clamp where glass is used as a substrate, our design has DNA origamis that can bind to a polymer sphere in solution. The binding of origamis to the polymer sphere is enhanced by multiple linkers, that are nothing but 21 bp long dsDNA strands. One end of each linker is attached to the origami and the other to the polymer sphere. The shrinkage of the sphere exerts forces on the target molecules that can be monitored. In an experimental setup, each polymer sphere is able to bind to 262 ± 15 origamis and $\sim 10^{11}$ target molecules can be manipulated in a single experiment. Although these forces are quite heterogeneous, a sensible qualitative picture of the distribution of forces can be obtained by averaging over long simulations. We have also attempted to characterise some mechanical properties of the DNA origami rod and use them to understand our findings. This study was a joint collaboration with experimentalists in the group of Khalid Salaita, Emory University, Atlanta, GA, USA. The results reported are all computational findings.

3.1 DNA Origami substrate

The preliminary calculations discussed in this section were performed by members of the Salaita group at Emory University. They have been included here to act as background information.

The choice of an origami that could conceivably be used as a substrate in particle collapse-driven hairpin unfolding experiments essentially depends on the dimensions

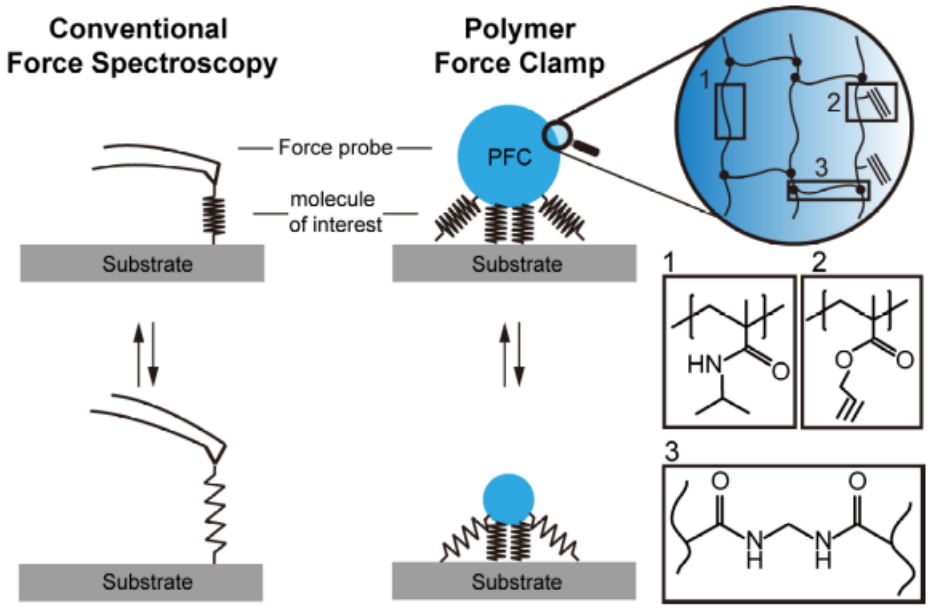


Figure 3.1: A schematic comparing conventional force spectroscopy with our approach that can allow ensemble-level parallelisation. Whereas conventional approaches may use an electric or a magnetic field to exert force, the OPFC design exerts force due to the decrease in size of a polymer sphere with an increase in temperature. Figure adapted with permission from Ref. [174].

of the rod, namely its length and thickness. Geometric calculations can be used to scan the length-thickness parameter space for a suitable choice. Fundamentally, there are four conditions the ideal design should meet:

- The polymer sphere's size can vary within a certain range in experiments and two target molecules (positioned 15 nm from each end of the origami rod) should be able to simultaneously bind to the sphere. From the force-extension curve of the hairpin under investigation (calculated using the protocol given by Zhang *et al.*[175]), the value of $F_{\frac{1}{2}}$ (the force at which the hairpin has equal probability of being in the open and closed states) was found to be ∼4.7 pN and the end-to-end extension at this force was ∼17.3 nm. Assuming a rigid sphere with an average radius of 250 nm, the maximum length of the rod above

which our hairpins will not be able to bind to the sphere can be calculated using the following equation

$$L_{\max} = 2\left(\sqrt{(17.3\text{ nm} + 250\text{ nm})^2 - (250\text{ nm})^2} + (15\text{ nm})\right). \tag{3.1}$$

Solving the above equation, $L_{\max}$ was approximately found to be 219 nm. All rods with length higher than $L_{\max}$ cannot be used in this study.

- Just like $L_{\max}$, there is a threshold value for the minimum length of the rod, below which the target molecules will not undergo any mechanical unfolding as they would be 'too close' to the polymer sphere and therefore experience much weaker forces. This $L_{\min}$ value was found to be 65 nm (assuming an end-to-end extension of 17.3 nm when the force is equal to $F_{\frac{1}{2}}$) and origami rods with lengths smaller than 65 nm were not considered.

- Particle collapse should not cause the DNA origami rod to 'buckle'. The condition for Euler buckling is:

$$F_c = \frac{\pi^2 P k_B T}{L^2}, \tag{3.2}$$

where F_c is the critical buckling force, L is the length of the rod, $k_B T$ (= 4.114 pN nm) is the thermal energy at room temperature and P is the persistence length of the rod that was estimated from the equation based on the model by Castro *et al.*[176]:

$$P = (53nm)\left(N_{\text{duplex}}\right)^{1.94}. \tag{3.3}$$

For different values of N_{duplex} and L, F_c was calculated and all (L,N_{duplex}) pairs were ruled out for which $F_c \leq 4.7$ pN ($F_{\frac{1}{2}}$). It must be noted here that although Euler buckling is a macroscopic theory, it has been shown to work well for stiff objects like DNA origami[177]. Euler buckling should be able to provide a reasonable estimate of the force at which buckling can be expected even though the geometry of the force application in the OPFC is quite different from Euler's original theory.

- The rod should have appropriate dimensions for self-assembly in solution. The number of nucleotides in a rod with a given L and N_{duplex} is roughly equal to $\frac{LN_{duplex}}{0.34\text{nm}}$ which should be less than the the number of nucleotides in a particular scaffold strand to be used in assembly. Given the fact that experimentalists have access only to a limited number of scaffold strands, all (L,N_{duplex}) pairs were discarded where the scaffold was too short to bring about the assembly of the rod.

The origami rod used as a substrate here was a 16-helix-bundle in a square lattice consisting of 13321 nucleotides (see **Figure 3.2a**).

3.2 Design of hairpins and linkers

The hairpins and linkers were all drawn on caDNAno[45]. The linker strands were necessary as they boosted the experimental yield of the OPFC assembly drastically. oxView[178], a browser-based visualisation and editing tool was used for setting appropriate sequences and also for orienting the linkers in the correct direction to prevent overlaps and entanglements during simulation. The system was then relaxed

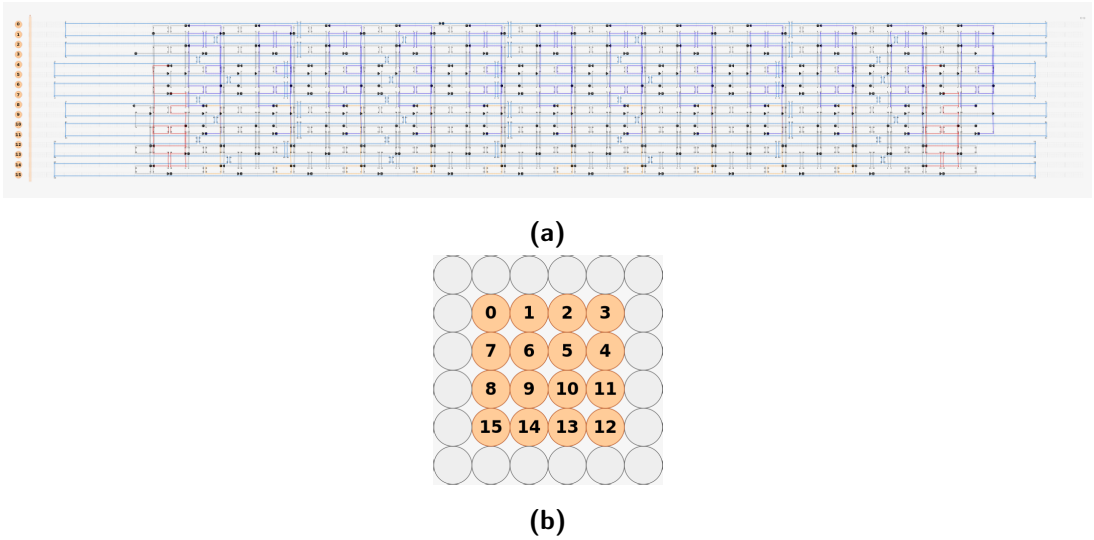


Figure 3.2: (a) caDNAno design of the 16 helix bundle origami rod and (b) its crosssection

using the method described in **Section 2.2.4**. During this relaxation, mutual traps were put between the corresponding nucleotides of the hairpins.

3.3 System setup

In the first instance, oxDNA was parameterised to capture the dynamics of only nucleic acids. Newer versions, as introduced in **Section 2.3** have incorporated nucleic acid-protein interactions as well. However, unlike the experimental setting (see **Figure 3.4**), it is still not possible to explicitly model a thermo-sensitive polymer sphere in oxDNA. Our initial plan was to use oxDNA's repulsive sphere force which encloses particles inside a sphere. It causes the particles outside the sphere to feel no force whereas those inside feel a harmonic force. This idea was dropped because movement of particles inside the sphere was not desirable. This would effectively allow the hairpins and linkers to change their point of attachment on the sphere after the collapse event. A decision was made to instead model the sphere by calculating

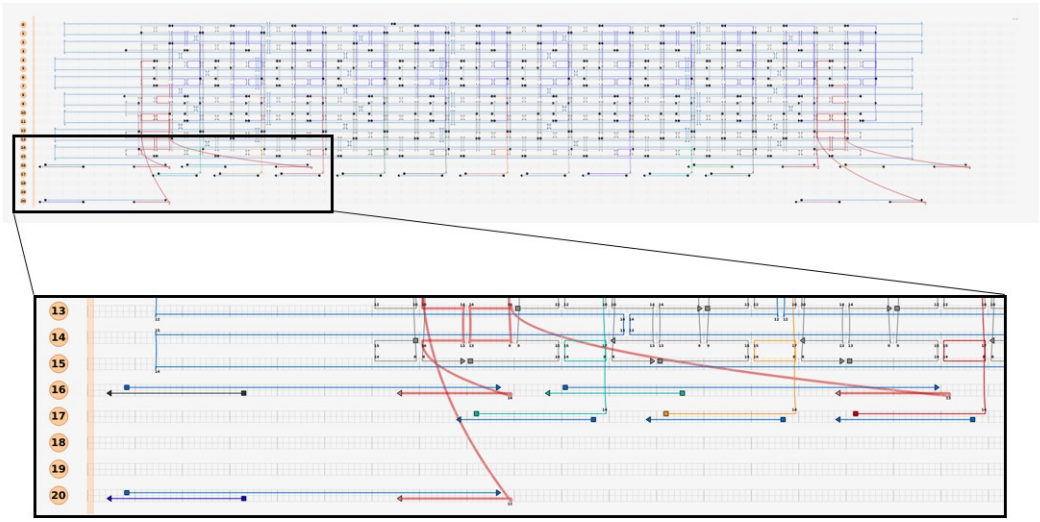


Figure 3.3: caDNAno representation of the same rod as **Figure 3.3** but with hairpins and linkers drawn. The inset shows three hairpins as single strands on virtual helices 16 and 20 as well as three linkers on virtual helix 17.

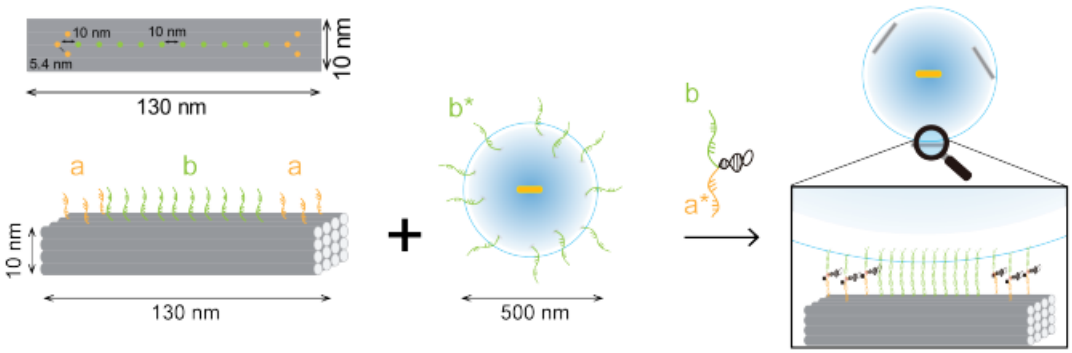


Figure 3.4: Binding of hairpins to complementary strands on the origami and the polymer sphere in a typical experiment.

the points on the sphere where the linkers and hairpins would attach themselves. This approach is illustrated in **Figure 3.5**.

The centre of mass of the origami ($\mathbf{COM_0}$), that of its top layer ($\mathbf{COM_1}$) and that of its two top-most layers ($\mathbf{COM_{12}}$) were calculated. From the centre of the top layer, a distance (d) was added followed by the radius of the sphere (r_{sphere}) in the direction given by the vector $\overrightarrow{\mathbf{V_1}} = \mathbf{COM_{12}} - \mathbf{COM_0}$ to get the centre of the sphere ($\mathbf{A}$). The positions of the points of attachment of the linkers and hairpins to the origami were extracted from the starting configuration file. Each linker/hairpin had a different point of attachment ($\mathbf{B_i}$) to the origami and another vector was defined for each linker/hairpin as $\overrightarrow{\mathbf{V_2}} = \mathbf{B_i} - \mathbf{A}$. In the direction of $\overrightarrow{\mathbf{V_2}}$, r_{sphere} was added to get the trap centre for that particular linker/hairpin on the sphere: $\mathbf{C_i}$. The distance d was just the initial distance of the sphere from the origami and it was varied in the range of 15 - 35 nm depending upon the system. After all the trap centres were calculated for all the linkers and hairpins, the external forces file was created with stationary harmonic traps of appropriate stiffness (or spring constant) and initial positions given by the positions of the trap centre. The positions $\mathbf{C_i}$ were calculated by applying this approach to the system when the polymer sphere was at low-temperature. For the high temperature $\mathbf{C_i}$ calculation, we assume that $\theta_\mathbf{i}$ remains unchanged during the sphere shrinkage. Consequently, $\mathbf{A}$, $\mathbf{B_i}$ and $\mathbf{C_i}$ may no longer remain co-linear at high temperature.

In this study we investigated seven systems **C1**,**C2**,**C3**,**C4**,**C5**,**C6** and **Δ32**, all with six hairpins. The systems **C1** to **C6** all had 10 duplexes (linkers) and differed only in the position of the hairpins along the origami. **C1** had all the hairpins concentrated at the centre of the origami (**Figure 3.6a**) while in **C6**, they were

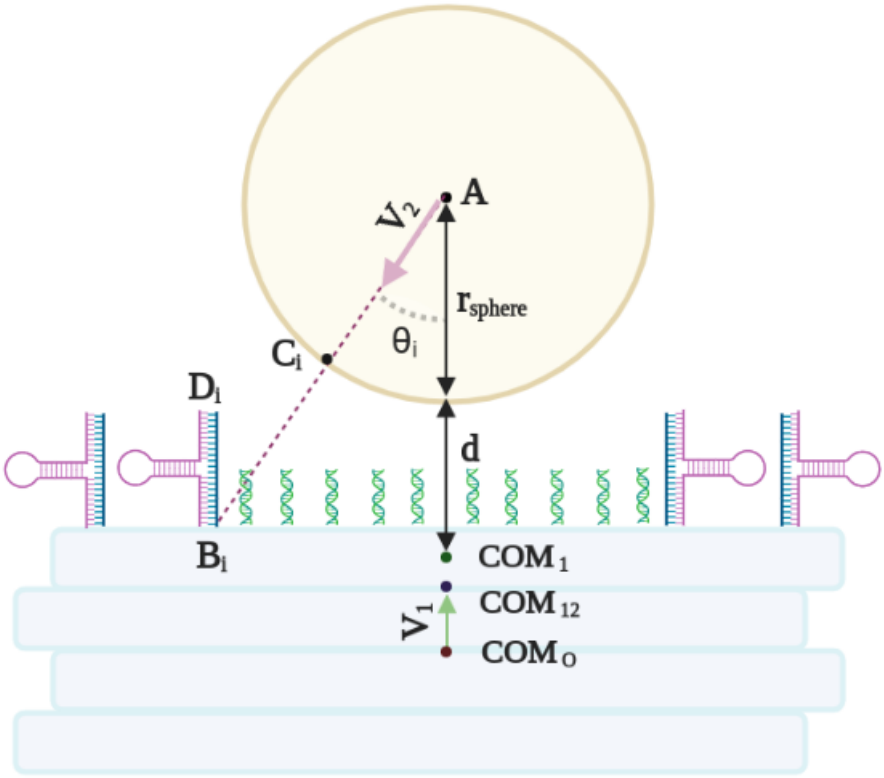
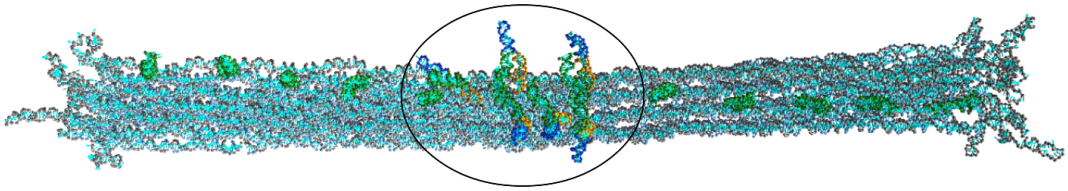


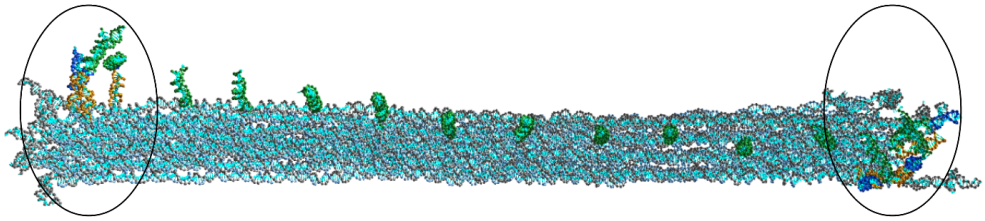
Figure 3.5: Schematic showing how the parameters for modelling the polymer sphere in the oxDNA simulations were chosen (for clarity only four of the six hairpins are shown). The relative sizes of the objects are not to scale.

at the two extremes of the origami in groups of three (**Figure 3.6b**). The **Δ32** system served as a control and had the hairpins in the same position as **C6** but lacked 16 staple strands at the centre of the origami and consequently had nine duplexes (linkers) instead of ten (**Figure 3.6c**). The absence of staple strands made **Δ32** more flexible than the other systems. From the experimental data on the dependence of the hydrodynamic diameter of the polymer sphere on temperature (see **Figure 3.7**), the sphere was assumed to have a radius of 280 nm at 313 K and 205 nm at 328 K.

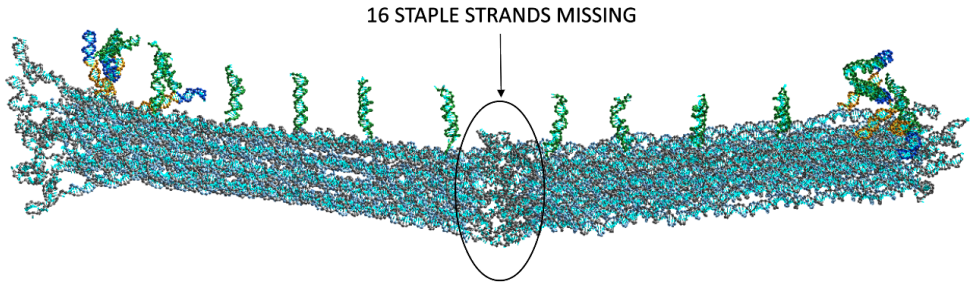
For each system, two states (at 313 K and 328 K) were simulated which shall be referred to as low and high temperature states. The simulations consisted of two



(a) The **C1** origami with all hairpins concentrated in the centre of the origami.



(b) The **C6** origami with three hairpins positioned at each end.



(c) The $\Delta 32$ origami missing a few staple strands and a linker in the highlighted region.

Figure 3.6: Simulation snapshots of systems **C1**, **C6** and $\Delta 32$ at low temperature showing their respective structural features. The hairpins have been coloured blue while the linkers have been coloured green and orange according to the same scheme used in experiments (**Figure 3.4**)

.

stages: first, the linkers and hairpins were pulled softly towards their corresponding trap centres on the sphere and then they were held firmly held there with a stiffer trap force. The simulations at low and high temperatures were run until the forces acting on the hairpins and linkers had equilibrated with a stiffness of 0.5 simulation units.

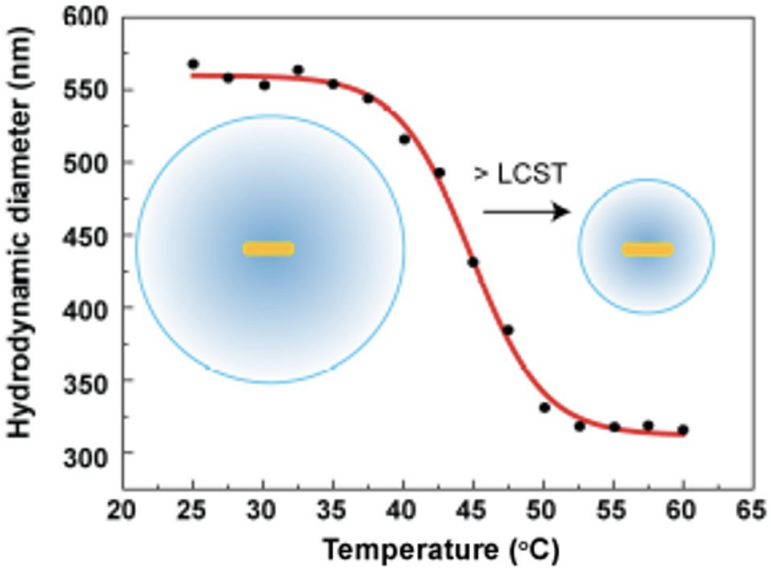


Figure 3.7: Variation of the hydrodynamic diameter of the sphere with temperature.

Assumptions

Our way of setting up the system was a simplified approach that involved many assumptions. First, we assume that the sphere was perfectly smooth and that when the hairpins and linkers attach to the sphere at low temperature, they do so to their closest point on the sphere. These were reasonable assumptions for the average point of attachment but in reality there would a distribution of such points across the ensemble of origamis. This potential heterogeneity in population of the OPFCs was not accounted for. Second, the elastic mechanical properties of the sphere were not explicitly taken into consideration. For example, in our simulations the spring constant for the harmonic traps was chosen to be in the 'stiff-sphere' limit where the resultant forces do not depend on the spring constant as the origami was more compliant than the traps. Third, all hairpins and linkers were assumed to be simultaneously attached to the sphere. If not all linkers/hairpins are bound, then the total applied force on the origami is likely to be less. In such a scenario, the

origami would be less bent which in turn could potentially lead to higher forces through the hairpins/linkers that are attached.

3.4 Force calculations

As the traps used in the simulations are of the simple harmonic type, the force experienced by the nucleotide on which the trap is acting ($\overrightarrow{\mathbf{F}}(t)$) can be easily estimated by the following equation

$$\overrightarrow{\mathbf{F}}(t) = -k\left(\overrightarrow{\mathbf{r}}_{\mathbf{trap}}(t) - \overrightarrow{\mathbf{r}}_{\mathbf{nuc}}(t)\right), \tag{3.4}$$

where k is the spring constant (or 'stiffness'), $\overrightarrow{\mathbf{r}}_{\mathbf{trap}}(t)$ and $\overrightarrow{\mathbf{r}}_{\mathbf{nuc}}(t)$ are the positions of the trap and the nucleotide, respectively, at time t. This force is a vector quantity and its projection along the hairpin must be taken to estimate the tension along the hairpin. This was done by taking a dot product. Long simulations are needed to obtain well-converged forces as the system needs to be able to diffusively explore the wide variety of all relevant configurations. The hairpins acted as our target molecules in this study and because of this, only the force along the hairpins were calculated.

If $\mathbf{F}$ is the total force acting on a nucleotide at any instance due to the trap, in order to measure the force through the hairpin, a component of this force $F_{\parallel}$ was defined which acts along the direction given by the vector $\overrightarrow{\mathbf{V_3}} = \mathbf{D_i} - \mathbf{B_i}$ where $\mathbf{D_i}$ is the position of the nucleotide on which the trap is acting (illustrated in **Figure 3.5**).

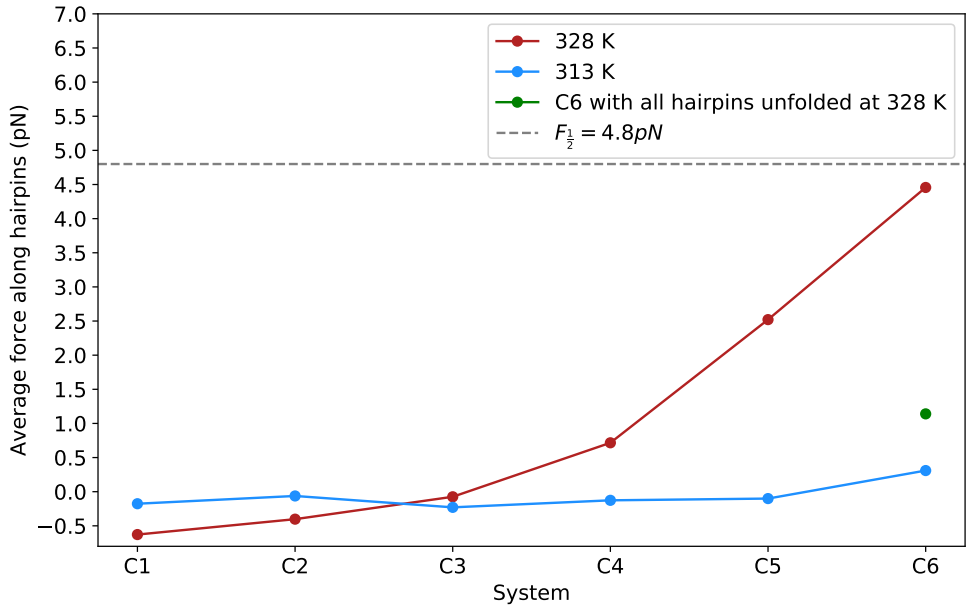


Figure 3.8: The average force acting on all hairpins for each system at low and high temperature. The dashed black line represents the $F_{\frac{1}{2}}$ of the hairpin. The green dot represents the scenario explained in **Section 3.7** where the hairpins were deliberately kept open.

So, we have:

$$F_{\parallel} = \mathbf{F} \cdot \frac{\overrightarrow{\mathbf{V_3}}}{||\overrightarrow{\mathbf{V_3}}||}. \tag{3.5}$$

From the choice of our vectors, it can be said that when $F_{\parallel} < 0$, the linker/hairpin is under compression and when $F_{\parallel} > 0$, the linker/hairpin is under extension. $\mathbf{F}$ and $F_{\parallel}$ were measured at low and high temperatures for each hairpin and were plotted along with their moving averages (calculated using 11 data points in the neighbourhood of a particular data point). A sample force versus time plot for one of the hairpins in the **C6** system is given in **Figure 3.9**. The force acting along the hairpins ($F_{\parallel}$), averaged over all configurations and all hairpins for each system from **C1** to **C6** is plotted in **Figure 3.8** (the same data is given **Table 3.1**).

It can be seen from **Figure 3.8** that at low temperature, the forces acting

Configuration	$F_{\parallel}^{\text{low}}$ (pN)	$F_{\parallel}^{\text{high}}$ (pN)
C1	-0.18 ± 0.72	-0.63 ± 0.25
C2	-0.06 ± 0.83	-0.40 ± 0.42
C3	-0.23 ± 0.64	-0.08 ± 0.59
C4	-0.13 ± 0.69	0.72 ± 0.81
C5	-0.10 ± 0.71	2.61 ± 2.04
C6	0.31 ± 0.97	4.46 ± 1.85
$\Delta 32$	-0.25 ± 0.34	1.26 ± 1.04

Table 3.1: Average force along hairpins at low and high temperatures.

along the hairpins are below their $F_{\frac{1}{2}}$ and therefore insufficient to cause unfolding. Moreover, at 313 K, the average force is close to zero irrespective of the position of the hairpins indicating that position has very little effect on the force at this temperature. At high temperature, a gradual increase in $F_{\parallel}$ is observed from **C1** to **C6**, showing that the tension in the hairpins depends on their position on the origami with the maximum tension felt when they are placed at the two ends of the origami. For the systems **C1** to **C3**, at 328 K, the sign of the force is negative which means that they are actually under compression. The tension in the stretched linkers at the extremities of the origami push the sphere and the origami together at the centre.

It is also important to realise that the force acting along a hairpin is not constant but varies significantly as a function of time. **Figure 3.10** shows the resulting distributions for the force along each hairpin of the **C6** system at high temperature. The width of these distributions are significantly larger than the average force and although the instantaneous force is often larger than $F_{\frac{1}{2}}$, it is the average force that is key in determining whether a hairpin unfolds.

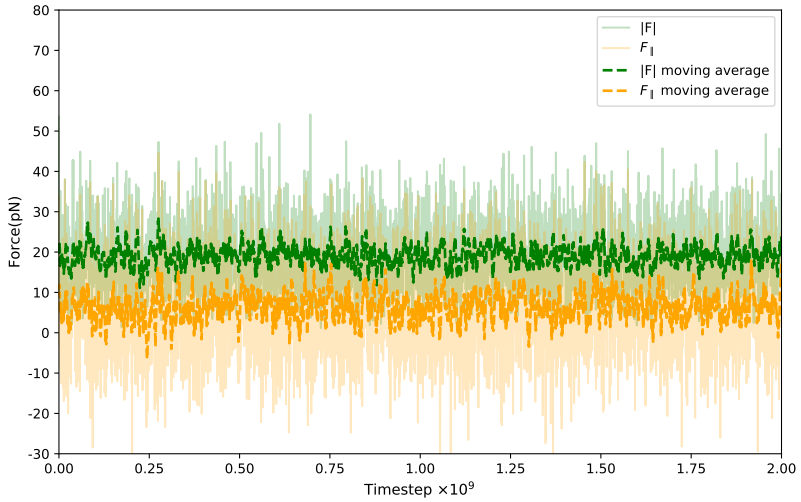


Figure 3.9: Force equilibration plot for one of the hairpins in the **C6** system at high temperature.

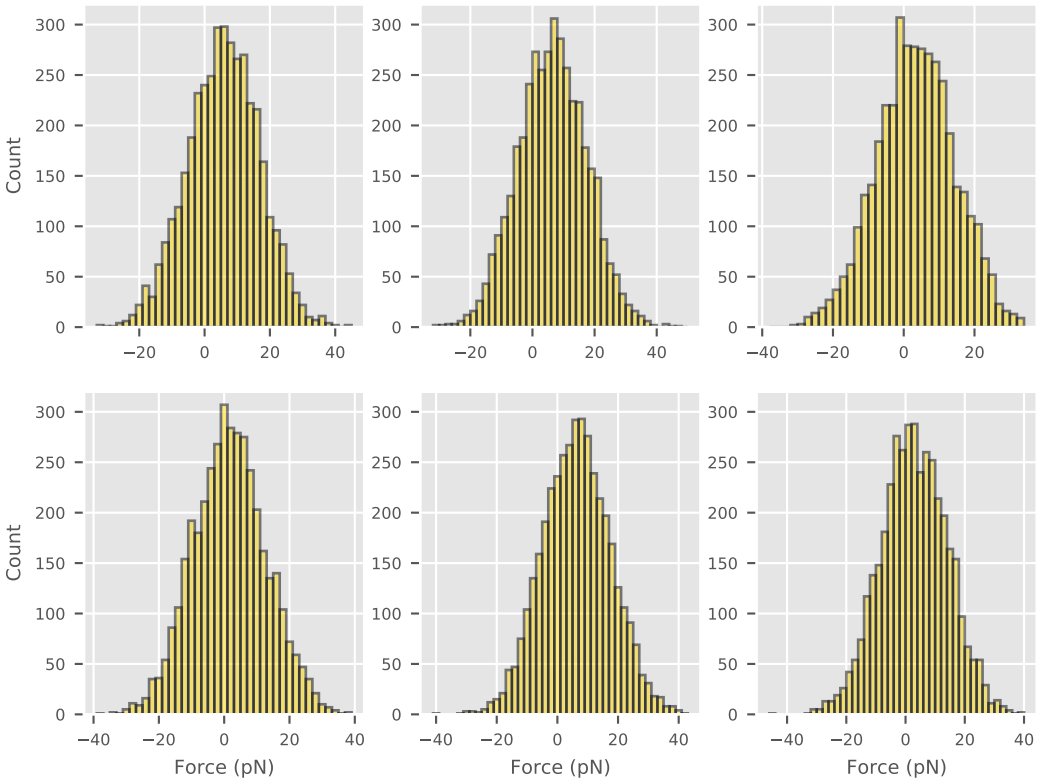


Figure 3.10: Distributions of the force going through each hairpin in the **C6** system at high temperature.

3.5 Angle of bending

When the **Δ32** system was simulated at high temperature, unsurprisingly, significant bending was observed associated with the hinge-like flexible region in the centre of the origami where the staples are missing. To calculate the angle between the two arms of the **Δ32** origami, four planes perpendicular to the long axis of the origami were defined: two at each end of the origami and two on either side of the central gap (See **Figure 3.6c**). From the centre of mass of the 4 planes, two vectors were defined each pointing to one end of the origami from the corresponding side of the central gap. The angle was then calculated using a dot product between these two vectors. This angle was plotted as a function of time for the low and high temperature simulations. Note that a straight rod correponds to an angle of 180.0° and the more bent an origami, the lower is the angle. The average angle of bending was found to be significantly smaller at high temperature ($\sim 137.4°$) than at low temperature ($\sim 166.9°$) because the flexible origami deforms to absorb much of the stress due to the particle collapse, thereby preventing hairpin unfolding (**Figure 3.11**).

The other systems under investigation, namely **C1** to **C6**, as expected, did not bend to anywhere to near the same extent but underwent uniform bending at high temperature. The parallel application of the forces through all the linkers and hairpins is the reason for this bending. The average bending angle for each system at low and high temperature was calculated and plotted as a function of the system (**Figure 3.12**). For the low temperature case, the values were quite close to 180.0° as expected but they decreased for all the systems at high temperature indicating significant bending. The origamis absorb a significant amount of the stress caused by

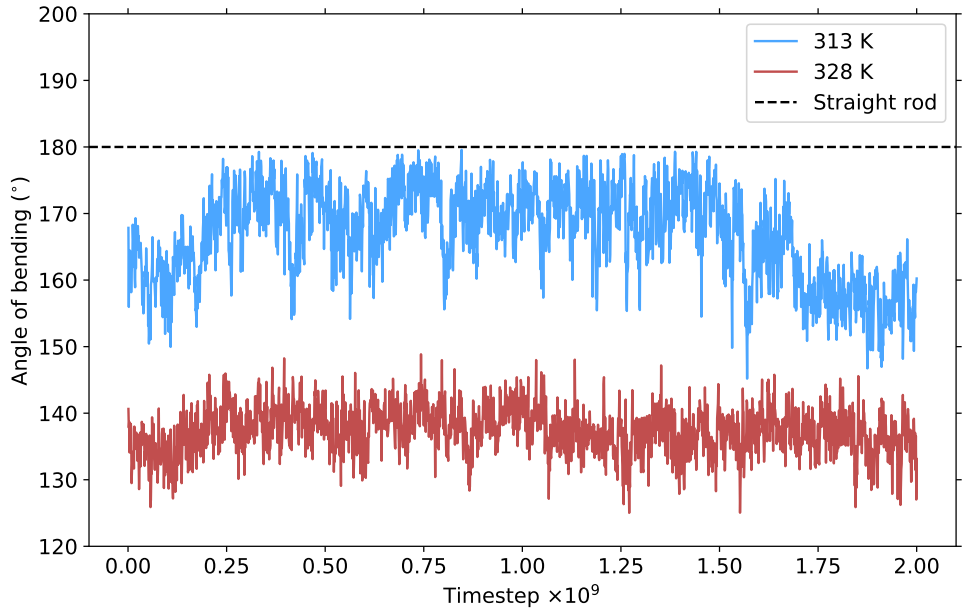


Figure 3.11: Fluctuations in the angle of bending for the **Δ32** system at low and high temperature.

the shrinkage of the sphere, and hence reduce the force felt by the hairpins compared to a perfectly rigid substrate. Although a perfect DNA origami can be considered as a rigid object, the total net force exerted on the origami in our setup, through all the six hairpins and ten linkers, is large. It is also clear that at low temperature the position of the hairpins on the origami rod had little to no effect on the bending.

The high temperature bending angle shows some variation as a function of the positions of the hairpins, albeit relatively small compared to the change in the angle of bending on going from low to high temperature. The lowest angle of bending and therefore the strongest bending deformation was observed in **C1** with an almost linear increase in the angle up to **C5**.

The precise bend angle is going to be a subtle function of the different forces mediated by each of the linkers and hairpins. The forces experienced by the linkers will be larger than a single hairpin at a given position because of their smaller

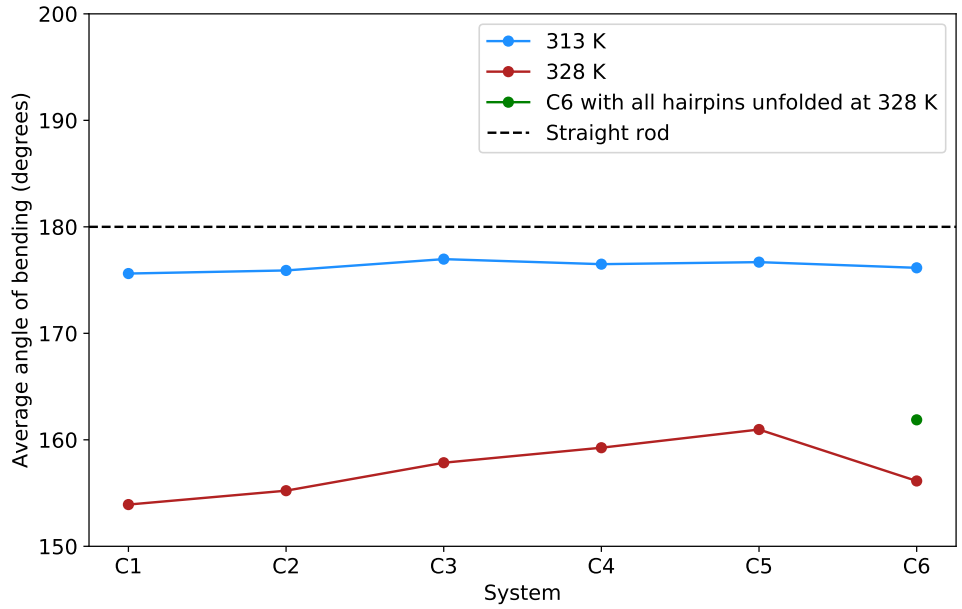


Figure 3.12: The average angle of bending for each system at low and high temperature. The green dot represents the scenario explained in **Section 3.7** where the hairpins were deliberately kept open.

effective length, but there are also three hairpins at a given position rather than a single linker. In addition, the tension in the linkers/hairpins will be greater as one moves towards the extremities of the origami, and so be more responsible for the origami bending. The increase in the bend angle on going from **C1** to **C5** is consistent with the idea that the effect of one linker is slightly more than that of three hairpins, but it is not clear why **C6** deviates from this trend. As the experiments observed some hairpin unfolding for the **C6** system, we also calculated the force in a system with the hairpins constrained to be unfolded. As this leads to a lower tensile force through these strands, the origami is less bent than when the hairpins are fully formed, as expected. **Table 3.3** shows snapshots of the final configurations of the various systems under study at low and high temperature.

Configuration	Angle of bending at low temperature (°)	Angle of bending at high temperature (°)
C1	175.61 ± 0.05	153.92 ± 0.06
C2	175.91 ± 0.06	155.22 ± 0.06
C3	176.97 ± 0.04	157.85 ± 0.07
C4	176.49 ± 0.05	159.26 ± 0.06
C5	176.69 ± 0.04	161.10 ± 0.07
C6	176.15 ± 0.05	156.14 ± 0.05
Δ32	166.93 ± 0.15	137.38 ± 0.08

Table 3.2: Average angle of bending at low and high temperatures.

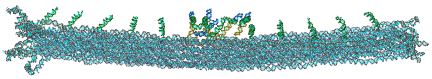
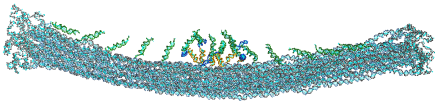
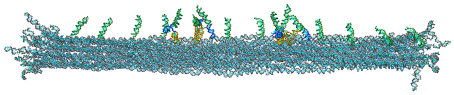
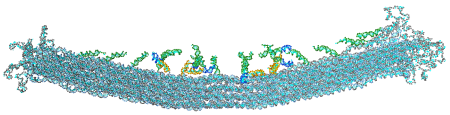
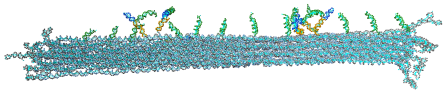
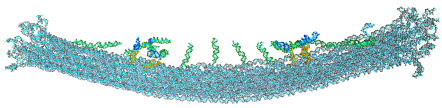
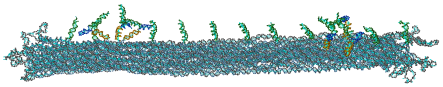
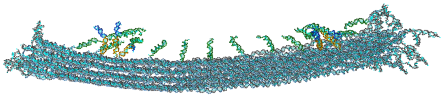
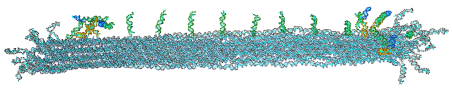
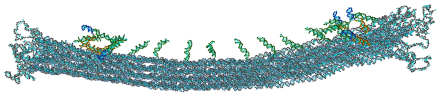
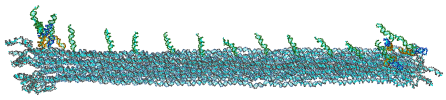
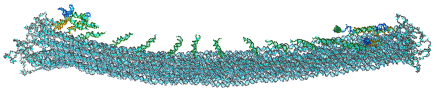
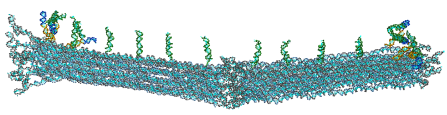
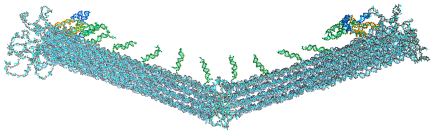
System	Low temperature	High temperature
C1		
C2		
C3		
C4		
C5		
C6		
Δ32		

Table 3.3: Snapshots of **C1**-**C6** and **Δ32** at low and high temperatures. At high temperature, a strong bend is visible in all systems from **C1** to **C6** while **Δ32** is kinked.

3.6 Elastic Properties of the origami rod

Using the protocol given by Chhabra *et al.*[179], it is possible to characterise some of the elastic mechanical properties of the 16-helix-bundle origami rod by analysing its thermal fluctuations. The method divides the origami rod into well-defined 'slices' perpendicular to the contour, each separated by a single base pair step along the direction of the rod. For each slice, an orthonormal reference frame $\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$ is calculated where $\widehat{\mathbf{e}}_3$ is tangential to the contour of the origami and $\widehat{\mathbf{e}}_1$ and $\widehat{\mathbf{e}}_2$ are orthogonal vectors lying on the plane of the slice. A very long MD simulation of just the 16-helix-bundle was performed to get a total of 2×10^3 non-correlated configurations. The elastic model assumes that the origami rod behaves like an ideal worm-like chain with an exponentially decaying tangent-tangent auto-correlation function given by **Equation 3.6**, where l_b is the bending persistence length, a is the contour length associated with a single step and m is the step size.

$$\langle \widehat{\mathbf{e}}_3(n) \cdot \widehat{\mathbf{e}}_3(n+m) \rangle = \langle \cos\theta(m) \rangle = e^{-\frac{ma}{l_b}}. \tag{3.6}$$

The method used to find the orthonormal set $\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$ involves singular value decomposition to find the orthogonal set of vectors closest to the set $\{\widehat{\mathbf{x}},\widehat{\mathbf{y}},\widehat{\mathbf{z}}\}$, where $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{y}}$ for the 16HB origami rod have been defined in **Equations 3.7 - 3.8**. The definition of $\widehat{\mathbf{z}}$ and $\mathbf{R}_{\text{duplex,i}}$ is exactly same as the original definition used in the paper[179]. The numbering of helices is done according to the virtual helix indices in

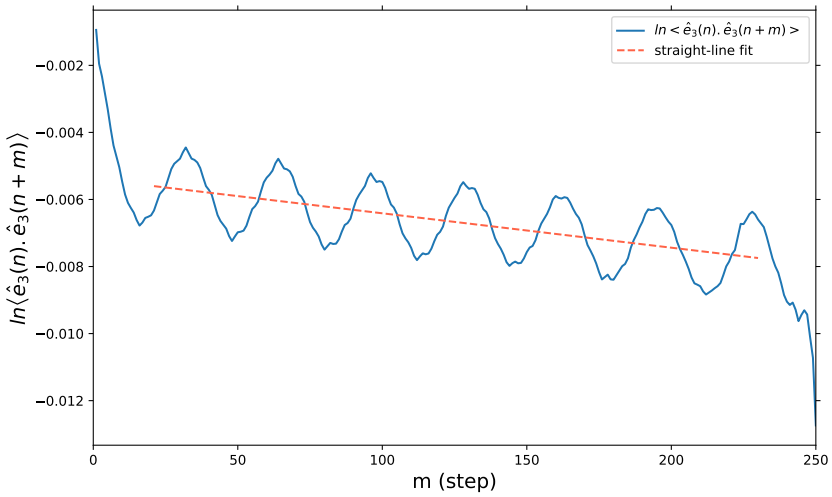


Figure 3.13: Plot of the tangent-tangent correlation function for the 16HB origami rod. The periodic trend is observed every 33 base pairs which roughly corresponds to the frequency at which junctions occur in the structure of the origami.

the caDNAno design of the rod (as shown in **Figure 3.2b**).

$$\widehat{\mathbf{x}} = \frac{\mathbf{R}_{\text{duplex},12} + \mathbf{R}_{\text{duplex},11} + \mathbf{R}_{\text{duplex},4} + \mathbf{R}_{\text{duplex},3} - \mathbf{R}_{\text{duplex},0} - \mathbf{R}_{\text{duplex},7} - \mathbf{R}_{\text{duplex},8} - \mathbf{R}_{\text{duplex},15}}{\left\|\mathbf{R}_{\text{duplex},12} + \mathbf{R}_{\text{duplex},11} + \mathbf{R}_{\text{duplex},4} + \mathbf{R}_{\text{duplex},3} - \mathbf{R}_{\text{duplex},0} - \mathbf{R}_{\text{duplex},7} - \mathbf{R}_{\text{duplex},8} - \mathbf{R}_{\text{duplex},15}\right\|} \tag{3.7}$$

$$\widehat{\mathbf{y}} = \frac{\mathbf{R}_{\text{duplex},0} + \mathbf{R}_{\text{duplex},1} + \mathbf{R}_{\text{duplex},2} + \mathbf{R}_{\text{duplex},3} - \mathbf{R}_{\text{duplex},12} - \mathbf{R}_{\text{duplex},13} - \mathbf{R}_{\text{duplex},14} - \mathbf{R}_{\text{duplex},15}}{\left\|\mathbf{R}_{\text{duplex},0} + \mathbf{R}_{\text{duplex},1} + \mathbf{R}_{\text{duplex},2} + \mathbf{R}_{\text{duplex},3} - \mathbf{R}_{\text{duplex},12} - \mathbf{R}_{\text{duplex},13} - \mathbf{R}_{\text{duplex},14} - \mathbf{R}_{\text{duplex},15}\right\|} \tag{3.8}$$

A plot of $\ln \langle \widehat{\mathbf{e}}_3(n) \cdot \widehat{\mathbf{e}}_3(n+m) \rangle$ against m, as shown in **Figure 3.13**, is expected to have a gradient of $-\frac{a}{l_b}$. This allows the value of l_b to be extracted.

To estimate the twist in the 16-helix-bundle, the average value of Θ_3, the local twist angle in the direction given by $\widehat{\mathbf{e}}_3$, was calculated to be -0.22 °. This corresponds to the origami rod having a slight left-handed twist of $\sim 84.09°$ over the length of the rod. An origami based on a square lattice without any insertions or deletions is expected to have such a twist[180], but it is not expected to significantly affect the geometric assumptions of the design. This twist is evident if one carefully

looks at the configurations in **Figure 3.6**.

The persistence length was found to be 33 µm. The persistence length is calculated using the classical WLC approximation (**Equation 3.6**). Oscillations are observed in the tangent-tangent correlation function (**Figure 3.13**). This occurs because each slice is not equivalent but can have different distances from the junctions between the helices. This lead to small variations in the correlation function which are readily apparent because the rods are so stiff and the decay of the correlation function small on the length scale of the origami. The length scale of the osciallation matches the periodicity of the pattern of junctions in the origami

3.7 Additional simulations

Constant force simulations of hairpin

To understand the dynamic behaviour of the hairpin that is acting as the target molecule in the force sensor described in **Section 3.3**, a few simulations where a constant force is applied on the first and last nucleotide of the hairpin strand were performed. This constant force was varied in the range of 0.5 - 29.0 pN and a base pair was assumed to be broken if the distance between the involved nucleotides was greater than 1.20 simulation units ($\sim$ 1.02 nm). The result of these simulations is plotted in **Figure 3.14**. Note that the results presented in this section are from short preliminary simulations to better understand the mechanism of unfolding and not to definitively locate the transition force. To do the latter, one would have to use umbrella sampling.

Three different states of the hairpin can be defined:

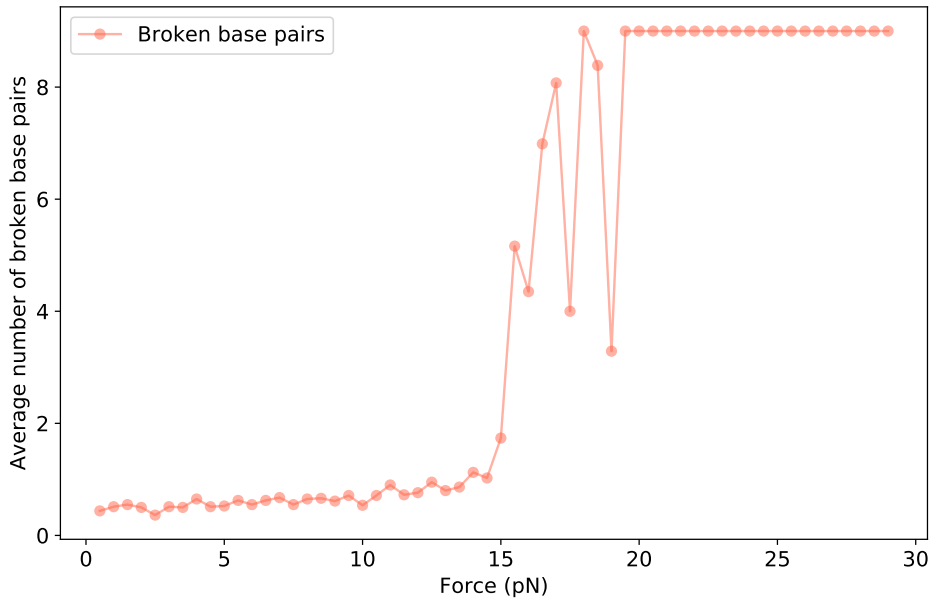


Figure 3.14: Results from individual constant force simulations of the hairpin.

- Closed, when the number of broken base pairs is 0
- Partially open, when the number of broken base pairs is between 1 to 8
- Open, when the number of broken base pairs is 9

Note that the above three states are not exactly well-defined as the ensemble is dominated by an open state and a base-paired state with some fraying at the ends which becomes stronger with an increase in force. The fraction of time spent by the hairpin in each of these states was plotted as a function of the force applied at each end in **Figure 3.15**. At low forces, the hairpin spent almost 50 - 60 % of the time in the closed state and the rest of the time in the partially open state. Between 10 and 15 pN, the partially open state became most prominent. Strong hopping was observed, particularly between the partially and completely open states when the force was between 15 and 20 pN. At forces higher than 20 pN, the open state became the most preferred state.

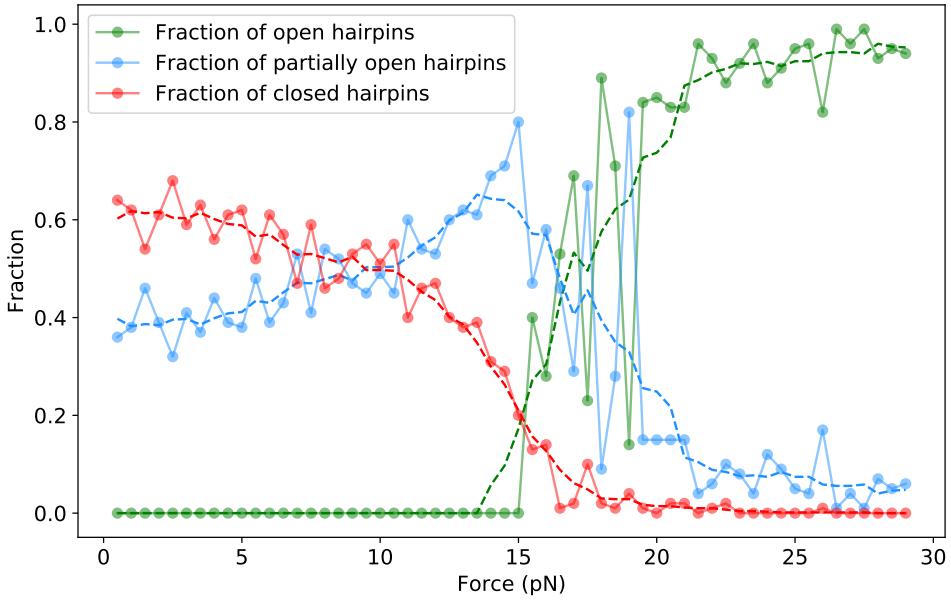


Figure 3.15: Fraction of time spent by the hairpin in each state as a function of the force applied to its ends. The solid dashed lines show the moving average for each state.

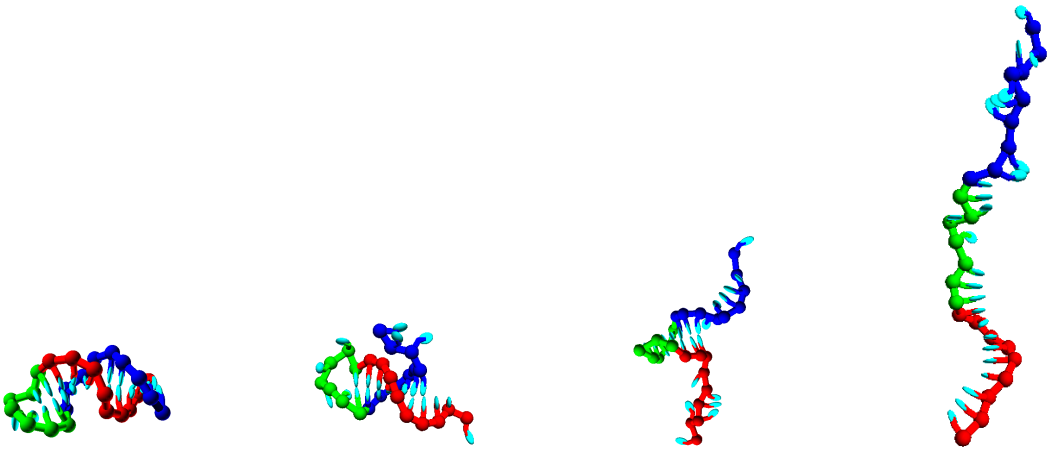


Figure 3.16: Snapshots of the hairpin held at a constant force of (a) 2.0 pN (b) 15.5 pN (c) 16.5 pN (d) 20.0 pN.

Simulations of C6 system under different conditions

Two alternative simulations of the C6 system at high temperature were performed to understand how the force exerted by the shrinking sphere was distributed between the hairpins and linkers. In the first, all hairpins were deliberately kept open by converting all the stem nucleotides to thymines. This prevented any hydrogen bonding between the hairpin nucleotides and the hairpins remained open as shown in **Figure 3.18**. Unsurprisingly, the average force ($F_{\parallel}$) going through each hairpin was much lower (1.14 ± 0.63 pN) compared to the original C6 system (4.46 pN). This proves that the stress induced by the shrinking sphere can be relieved by the unfolding of hairpins. The angle of bending for this system was calculated and plotted earlier in **Figure 3.12**.

In the second simulation, two traps acting on the duplex linkers just adjacent to the hairpins on either side (linkers L1 and L10 in **Figure 3.17**) were '*cut*' (or removed). The removal of these traps gives rise to a decrease in the number of linkers and hairpins that exert a force on the origami and hence a decrease in the origami bend angle. Consequently, the force acting along the hairpins is expected to increase. Our simulations yield an average force equal to 6.14 ± 0.61 pN, indicating that a loss of these two linkers causes the force to increase by ∼ 1.69 pN. This shows how deviations from the designed structure of the force clamp can affect the forces acting on the hairpins.

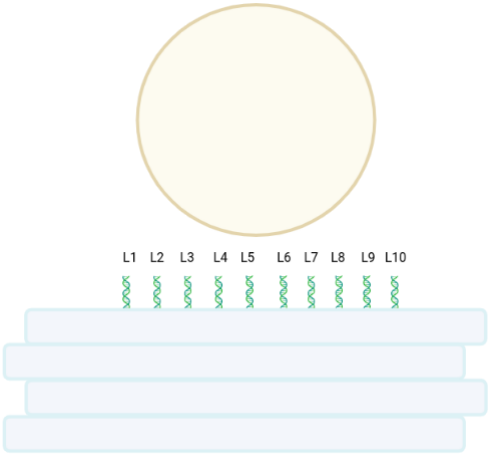


Figure 3.17: Schematic showing only the 10 linkers in the **C6** system.

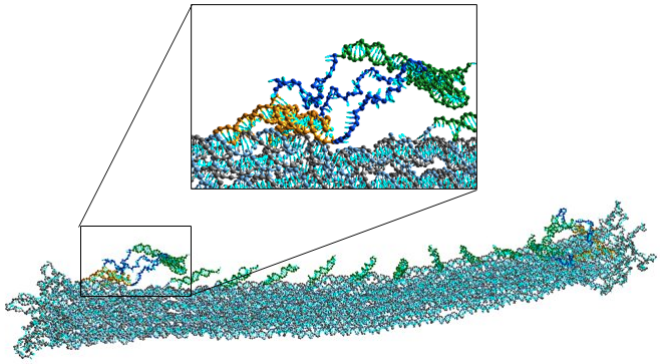


Figure 3.18: **C6** system with the hairpins kept deliberately open.

3.8 Comparison with experimental results on the force clamp

In the experiments performed on the OPFC, a fluorophore and a quencher are attached to the first and last nucleotide of each hairpin. With an increase in tension, the distance between the fluorophore and quencher increases which leads to a fluoreschence signal. In these experiments, there was a clear change in fluorescence signal on polymer sphere shrinkage (at $\sim$ 45°C), but much smaller than the jump at $\sim$ 65°C associated with the hairpin melting. This implies that a small fraction of the hairpins unfold on polymer shrinkage. The increase in normalized intensity before hairpin melting is largest for the **C6** system and almost negligible for the **C1** sys-

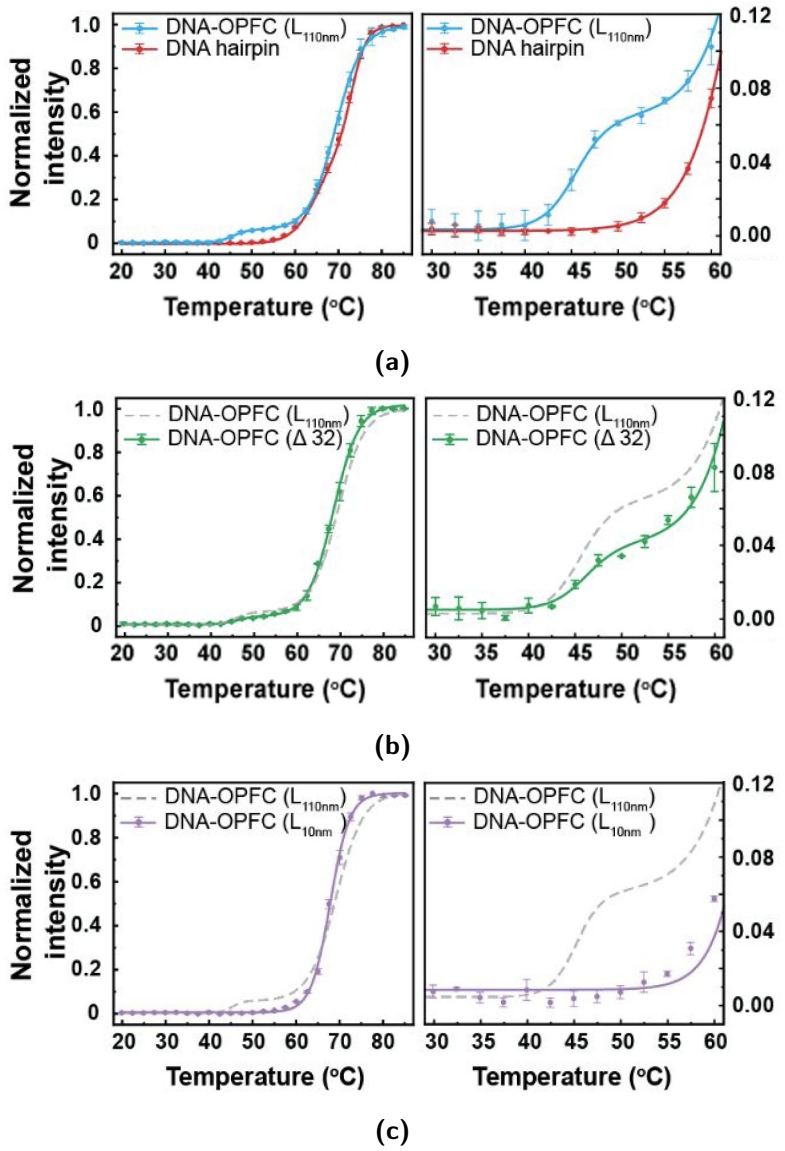


Figure 3.19: Temperature-dependent fluorescence measurements of the (a) **C6** (also known as $\mathbf{L_{110nm}}$) (b) $\Delta 32$ and (c) **C1** (also known as $\mathbf{L_{10nm}}$) systems. The plots in the right column are zoomed-in versions of the corresponding plots in the left column.

tem, with the $\mathbf{\Delta 32}$ system having a very weak signal. From this data, it is safe to conclude that in the $\mathbf{\Delta 32}$ and **C1** systems, no hairpin unfolding is observed. Note that sometimes fraying at the end of the hairpins can cause changes in the fluorescence signal, giving the impression of unfolding. Plots of the normalized fluorescence intensity against a change in temperature, as obtained from experiments is shown in **Figure 3.19**.

In simulations, the origami has a complex response to polymer shrinkage. There is a complex pattern of force transmission through the hairpins and linkers and the bending of the DNA origami rod shows that it is not a perfectly rigid substrate. This bending reduces the force experienced by hairpins compared to a completely rigid substrate. Although the force measured for the **C6** system in simulations is close to the experimental $F_{\frac{1}{2}}$, we rarely observe any unfolding. The OPFC design is not a constant force design which means that if the hairpins begin to unzip, the force through the hairpin will go down. If the OPFC is to completely unzip the hairpins, the force through the hairpins when they are all unzipped should be greater than $F_{\frac{1}{2}}$, whereas in our simulations we find that it is significantly lower, namely 1.14 pN.

Our results explain why complete hairpin unzipping is not seen for **C6** and also why no unzipping is observed for **C1** and **Δ32**. Even though the force through the hairpins is close to $F_{\frac{1}{2}}$ for **C6** at high temperature, it is not sufficient to start unzipping any of the hairpins, which is a discrepancy between the computational and experimental results. There could be a few possible reasons for this. First, the oxDNA model has certain deficiencies (for example, $F_{\frac{1}{2}}$ for hairpins in oxDNA may be greater than $F_{\frac{1}{2}}$ in experiments). Second, there is a heterogeneous population of OPFCs. It is possible that only a small fraction exert sufficient force to unfold the hairpins. There could be many sources of this heterogeneity:

- The hairpins/linkers may not bind to the nearest point on the sphere.
- Some linkers or hairpins may not be bound to the sphere (we considered an explicit example where the absence of two linkers leads to a significant increase in force).

- There may be defects in the origami, for example missing staples, that tend to enhance its flexibility[181].

Our simulations also suggest ways that the OPFC could be adapted to, for example, increase the force exerted on the hairpins. Broadly speaking our work has been successful in showing that the bending of the origami is a key factor in reducing the expected value of the force, which was not predicted in the original design. An obvious target is to reduce this bending. For example, by choosing an origami with an anisotropic cross-section one could perhaps increase the bending modulus associated with bending towards the sphere. One would probably also want to eliminate the origami twist by insertion of appropriate defects. Although the purpose of the linkers was to increase the binding of the force clamps, we have found that they also transmit a significant amount of force to the origami and hence enhance bending. If these linkers were lengthened, the origami bending could be reduced without interfering in their role of enhancing binding. Future studies can focus on modelling the location of the fluorophores as in experiments to quantify experimental signals.

Chapter 4

Structure of the nucleosome core particle under the influence of external force

This chapter will mainly focus on the implementation of a hybrid protein-DNA coarse grained model to enable the study of the structure of a nucleosome core particle under various external forces. The motivation behind this study is to establish models of the nucleosome in its fully wrapped and its force-induced partially unwound state, so that an origami force clamp can be designed that is able to actively transition the nucleosome between these two states. Note that the goal here is not to model the unwrapping of the nucleosome but to use molecular dynamics simulations to impose several geometries on the nucleosome and measure the force exerted on it in those geometries. This project is still being actively pursued in collaboration with experimental scientists from the Bustamante group at University of California, Berkeley and with the ultimate aim being to use cryoEM of nucleosomes in the force clamp to determine the structure of the partially unwrapped nucleosome.

There is enough evidence today to say that the genetic message comprises not just the sequence of DNA but also the details of how it is packaged inside the cell. In a single eukaryotic cell, the genetic information is stored in DNA molecules that are over two meters in length. Inside the nucleus of the cell, the genome is compacted to roughly a hundred-thousandth of its length by a hierarchical scheme of folding

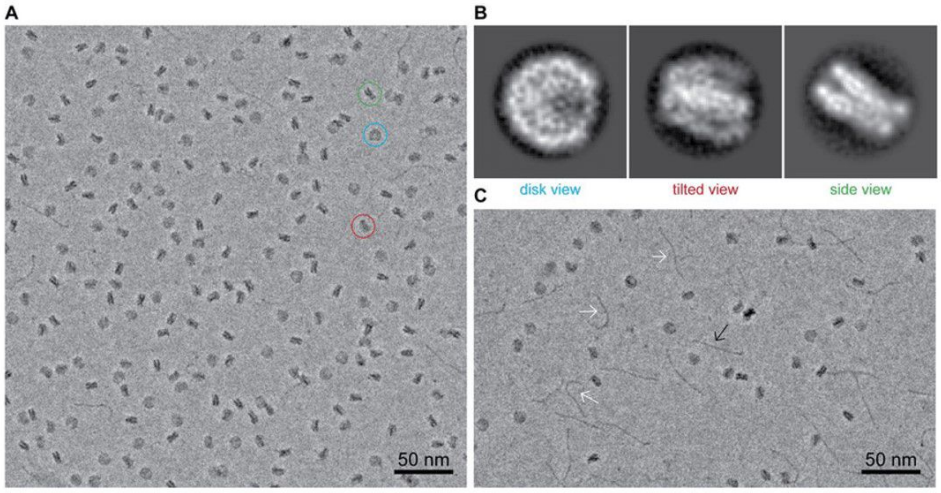


Figure 4.1: Electron micrographs of NCPs in ice. White and black arrows indicate bent DNA and straight DNA respectively. Figure reproduced with permission from Ref. [191]

around proteins giving rise to a nucleoprotein complex known as chromatin[182]. Unsurprisingly, the structure of chromatin affects DNA accessibility during key cellular processes like replication, repair and transcription[183,184]. Along a chromatin fibre, the chromosomal DNA is tightly packaged into subunits called nucleosomes[185,186]. At the first level of arrangement, a 147 bp dsDNA strand is wound ~ 1.7 times in a left-handed superhelix around a disc-like protein core, $\sim$5 nm in height and $\sim$7 nm in diameter, consisting of eight histones[187]. This organisation of the DNA strand around the protein octamer is called the 'nucleosome core particle' (NCP)[188,189]. The octameric protein core is made up of two molecules each of histones H2A, H2B, H3 and H4[188,190]. **Figure 4.1** shows NCPs imaged using electron microscopy.

The nucleosome system although extremely important, is far from fully understood. It is very interesting that the nucleosome is both highly stable and protective of the genome, while at the same time its tightly wrapped DNA can be readily

made accessible for transcription when required. Based on the persistence length of dsDNA (∼50 nm or ∼150 bp)[192–195], one would expect that significant work is be required to make it adopt such a strongly bent superhelical state. Experiments have shown that this DNA wrapping around the histone core is facilitated by strong interactions (particularly electrostatic) at various points of contact between the histones and the nucleotides[196]. This binding leads to local distortions in the dsDNA structure, for example, it is kinked in many places[197]. Interesting correlations have been found between the locations of nucleosomes and features of the DNA sequence; possible origins for this effect include stronger DNA-nucleosome interactions or lower free-energy costs associated with bending for certain sequences[198].

If the DNA is so tightly packed inside the cell, for it to serve its genetic purpose as a template in replication and transcription it must be able to be liberated from nucleosomes when required. One potential mechanism is that the DNA unwraps from the nucleosome due to some trigger, for example due to the action of protein machines that exert mechanical stresses on the DNA or due to the histones that partially disassemble. Though some biochemical studies have been successful in calculating certain properties of the nucleosome[183–185,199,200], these methods only report ensemble and time averages and do not provide any data on transient events or the intrinsic dynamics of heterogeneous biological structures. Single-molecule manipulation methods have become an increasingly useful tool for studying DNA-protein interactions on the molecular level[201–203]. In particular, they have furnished lots of new information on the structure, function and dynamics of chromatin and nucleosomes[204–209]. The Bustamante group have contributed significantly to this area[210–212]. In particular, their optical tweezer experiments[213] have identified in-

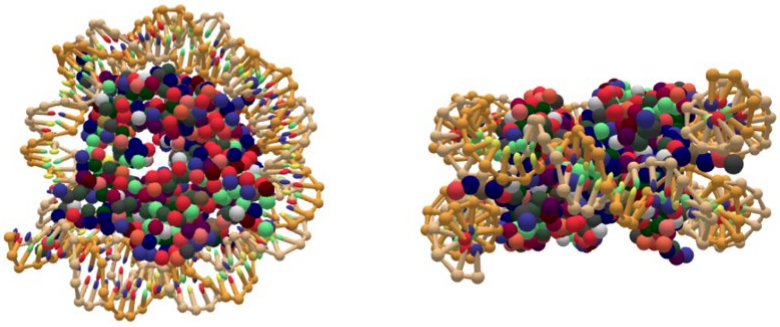


Figure 4.2: Snapshots of the nucleosome **3LZ0** as visualised on oxView from two different angles. The colour of the amino acid residues represented as balls is arbitrary.

termediates in the force-induced dissociation of dsDNA from the histone core. These experiments have characterised the force range over which these intermediates are stable and the extension released by the partial disassembly. But the detailed structure of these intermediates is unknown. The overall aim of the project to which the work described in this chapter contributes is the experimental determination of the structure of this intermediate.

4.1 Simulating a nucleosome in ANM-oxDNA

The ANM-oxDNA model as introduced in **Section 2.3** uses an anisotropic network of balls and springs to represent protein residues and bonds between them along with the standard oxDNA representation of nucleotides. To study a particular nucleosome, its .pdb file obtained from the Protein Data Bank, must be converted to ANM-oxDNA using a set of Python scripts and the tacoxDNA web server[151]. For all simulations, a cutoff distance of 18 Å was used. In the ANM-oxDNA model, the amino acid residues and the DNA nucleotides interact only via an excluded volume

potential that is not sufficient to capture the electrostatic interactions between the positively charged histone octamer and the negatively charged DNA backbone. To mimic these electrostatic forces that hold the nucleosome together, one needs to explicitly introduce mutual trap forces in the system. These traps act between protein residues and nucleotides and their number and stiffness can be varied. The equilibrium distances of these traps were set to the distance between the corresponding particles in the original PDB structure. For the initial nucleosome **5B32**, 15 mutual traps each of stiffness 5.709 pN/nm (or 0.1 simulation units) were used to simulate it in its completely wrapped state. The rest of the chapter will mainly pivot to the simulation of nucleosome **3LZ0** in its completely wrapped, partially unwrapped and fully unwrapped states. This nucleosome was chosen because it was readily available to the experimental collaborators. The initial structure of the nucleosome is shown in **Figure 4.2**.

4.2 Nucleosome under constant force

The **3LZ0** nucleosome has 751 amino acid residues and a DNA duplex consisting of 145 base pairs wound around it, making approximately 1.6 helical turns. In experiments performed by Mihardja *et al.*[213], three distinct states of the nucleosome were observed, each stable over a particular force range (as shown in **Figure 4.3**). At this state, we are interested in the transition between the completely wrapped and partially unwrapped states. For any simulation of the nucleosome under a constant force, the number of mutual traps and their stiffnesses could be varied. Simulations under a constant force have enabled us to tune the number and stiffness of the

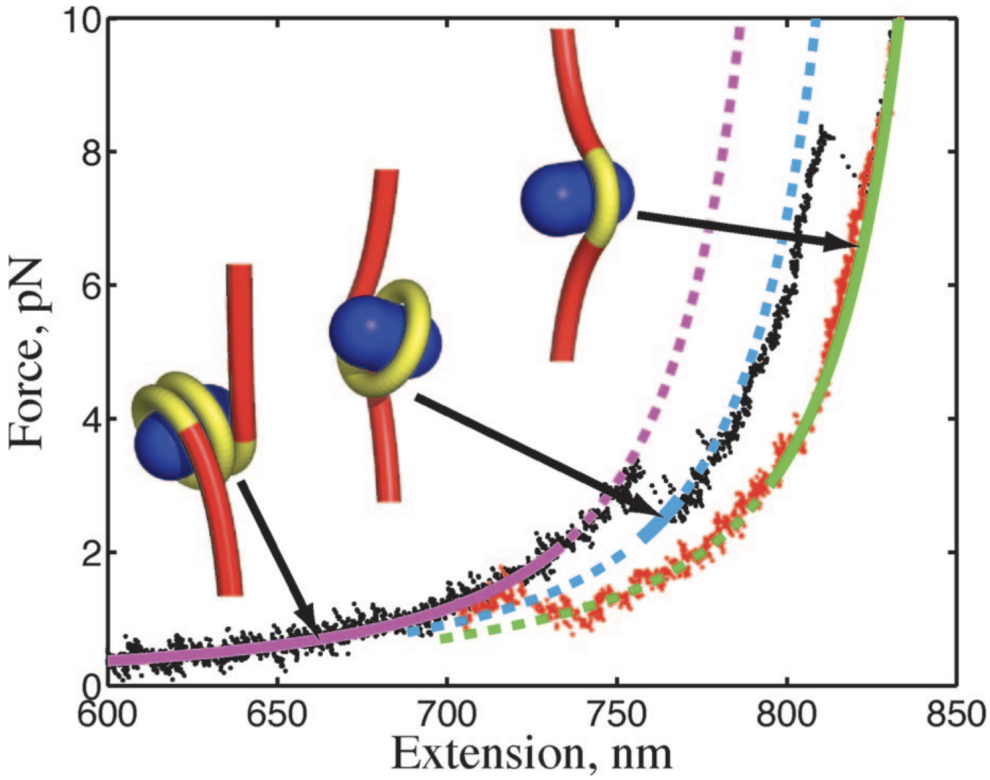


Figure 4.3: The force extension curve of a spool model of a nucleosome. Theoretical curves are drawn for completely wrapped (purple), partially unwrapped (blue) and completely unwrapped (green) states along with experimental data for increasing (black) and decreasing (red) tension. The solid part of the curve indicates that the state is thermodynamically preferred and the dashed part indicates a state that is stable but not preferred. Figure reproduced with permission from Ref. [213].

mutual traps so that two states of the nucleosome for which we wish to develop models behave in a way similar to experiments. It is known from experiments[213] that the tranistion to the partially unfolded state of the nucleosome is complete by ~ 4 pN. The change in extension associated with a transition to this state is between 10 and 20 nm. To avoid breaking base pairs of the nucleosomal DNA, three thymines were added to each end of one of the two DNA strands and force was exerted on the terminal thymine at each end. The extension was calculated from the end-to-end distance between the nucleotides on which the forces were acting.

4.2.1 Effect of trap stiffness

We first considered the behaviour of our model of the completely wrapped nucleosome under force. In particular, we focused on the effect of uniformly varying the trap stiffness of all 15 traps when under a tension of 4 pN. The dependence of the change in extension on the trap stiffness is illustrated in **Figure 4.4** along with example configurations. As expected, the extension increased as the stiffness was decreased, mediated by the unravelling of DNA from the nucleosome. Note, this unravelling is not mediated by dissociation from the traps (they are harmonic) but by extending the increasingly soft 'trap bonds'.

To model the completely wrapped nucleosome, we chose to use relatively stiff traps, namely $k = 0.3$ simulation units = 17.13 pN/nm, so that the extension of this state does not significantly change with the force. One could consider using a lower trap stiffness to capture the partially unravelled states. For example, around $k = 1$ pN/nm in **Figure 4.4**, the nucleosome has an extensional change that is consistent with experiments. However, this extensional change is likely to be too dependent on force, so instead we decided to generate a model of the partially unwrapped nucleosome by removing some of the traps.

4.2.2 Effect of the number of traps

The nucleosome can be expected to unfold if we decrease the number of mutual traps while keeping their stiffness constant. A stiffness of 0.3 simulation units was chosen which was high enough to not significantly change the nucleosome's extension under an external force. The initial structure had 15 mutual traps. The number of traps was varied in the range of 13 to 5, in steps of 2 with each step involving

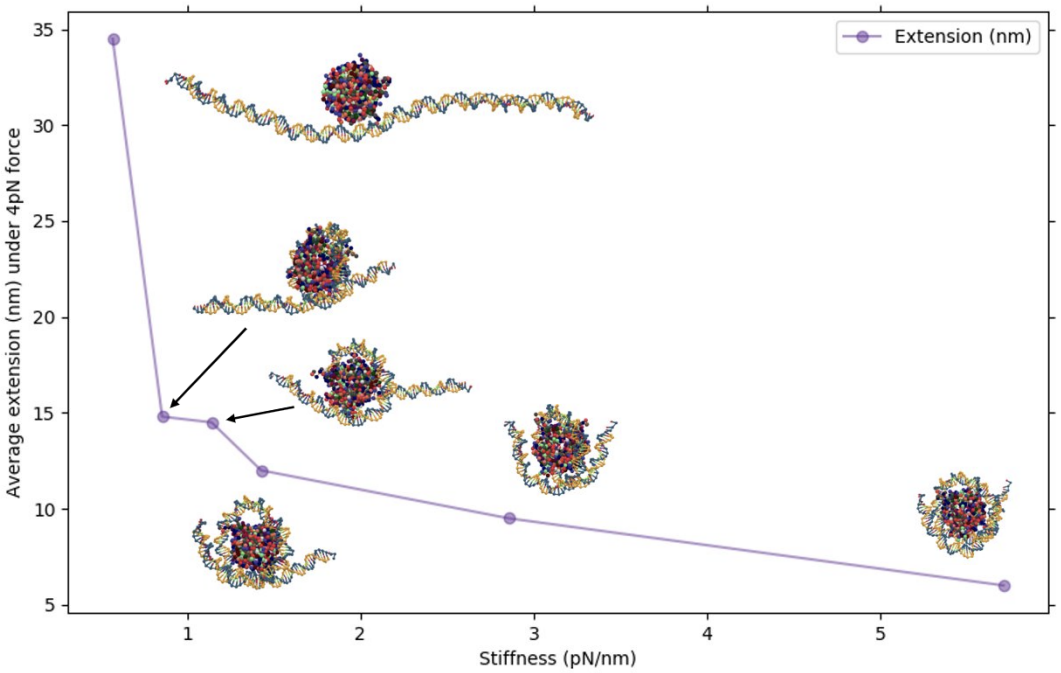


Figure 4.4: Dependence of nucleosome extension on trap stiffness under a constant external force of 4 pN. Snapshots show configurations of the nucleosome at a particular stiffness value. The extension is measured relative to the completely wrapped state of the nucleosome when simulated using ANM-oxDNA.

the removal of a trap from either end of the nucleosome symmetrically. Under a constant external force of 4 pN, the effect of the number of traps on the extension is shown in **Figure 4.5** along with illustrative configurations. As expected, the DNA unravels and the extension increases as the number of traps is decreased. Because of the geometry of the wrapped nucleosome, the extensional increase is not uniform.

From the above calculations, 9 mutual traps each of with a stiffness of 0.3 simulation units, were deemed appropriate for the model of the partially unfolded state of the nucleosome. To further check the model, the nucleosome was simulated at somewhat higher external forces (4.5 - 6.0 pN) and the extension recorded. The extension did not increase to more than 11 nm which was still within the target range (10 - 20 nm).

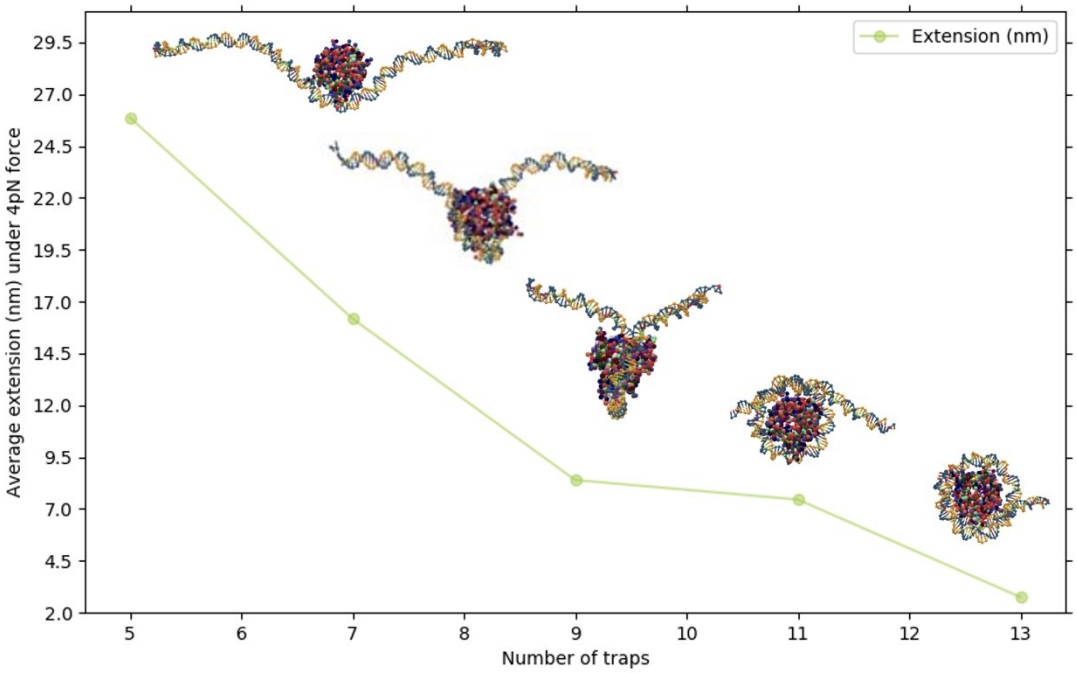


Figure 4.5: Dependence of nucleosome extension on the number of traps under a constant external force of 4 pN. Snapshots show configurations of the nucleosome with a particular number of traps. The extension is measured relative to the completely wrapped state of the nucleosome when simulated using ANM-oxDNA.

4.3 Force-sensor calibration

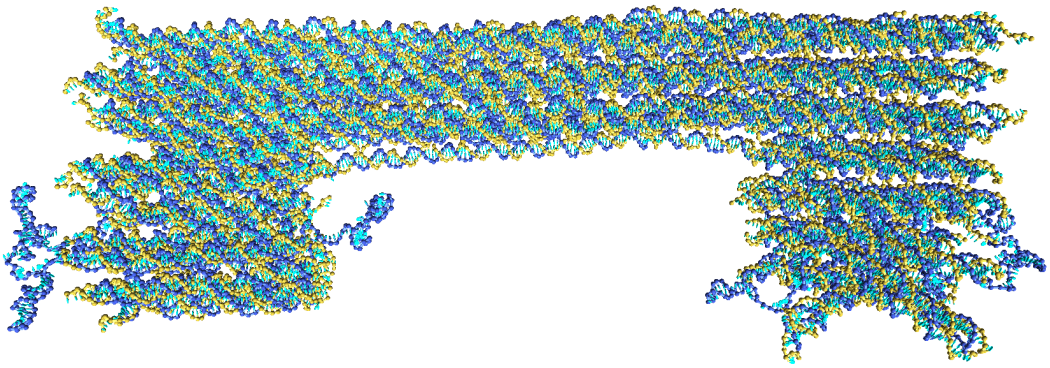
The DNA origami force clamp developed by Nickels *et al.*[214] applies force to a molecule using ssDNA entropic springs that tether the target molecule between rigid origami blocks. A series of force clamps were designed that exert different forces on the target molecule through adjustment of the length of ssDNA 'springs' in the design. This is practically done by changing the sequence of relevant staple strands in the origami so that ssDNA is transferred between a 'reservoir' and the central gap. Depending upon whether the ends of the tension-bearing strands are ligated or nicked, the sensor can be turned '*on*' and '*off*' (shown in **Figure 4.6**). We wish to adapt this force-clamp design to dynamically change the force exerted on a nucleosome attached to the DNA that bridges the gap so that it transitions

between the wrapped and the partially unwrapped states.

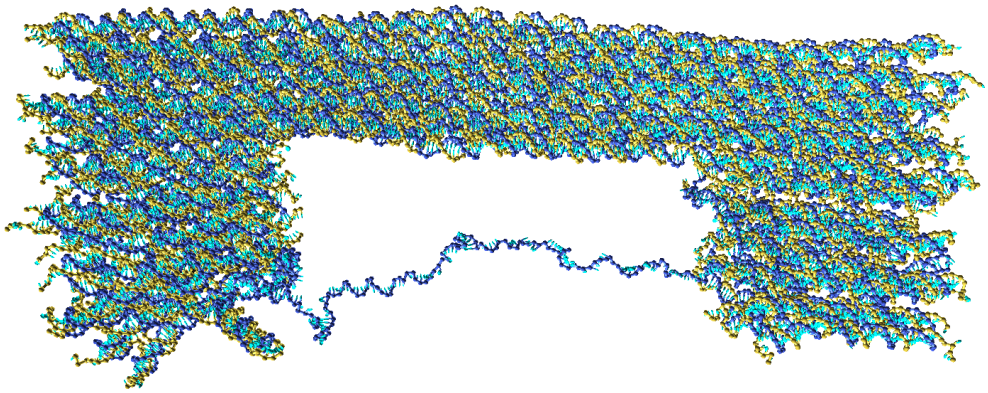
The force exerted by these origami force-clamps are not easy to estimate. However, oxDNA simulations provide a potential means to calibrate these devices, as shown by Engel *et al.*[215]. Indeed, this work suggested that the estimated forces in the original paper have substantial errors. As measuring converged internal forces directly from simulations is too computationally demanding for systems the size of DNA origamis, a two-step procedure was used. First, the constraints that the origami places on the ends of the DNA tightrope that bridges the gap were measured. The Gaussian-like distance distribution obtained in a simulation of the complete origami in the 'off' state is shown in **Figure 4.7**. Then, long simulations were performed on just the DNA tightrope subject to these constraints so that the internal force in the DNA could be measured. The constraints were reproduced by applying harmonic traps to each end of the tightrope that have a stiffness k and are separated by a distance d. Using the mean $\langle r \rangle$ and variance σ_r^2 of the distribution shown in **Figure 4.7** and the protocol outlined by Engel *et al.*[215], the effective trap stiffness k was approximately found to be 14.61 pN/nm and the trap separation d was calculated to be around 43.11 nm.

4.4 Force actuating device strategy

The planned mechanism to dynamically change the force exerted by the clamp is through the addition of a strand that on binding reduces the effective length of the DNA tightrope. Our strategy to initiate this shortening is illustrated in **Figure 4.8**. The added orange strand is designed to bind to the two green complementary regions



(a)



(b)

Figure 4.6: Simulation snapshots of the force sensor developed by Nickels *et al.* in the (a) off and (b) on states.

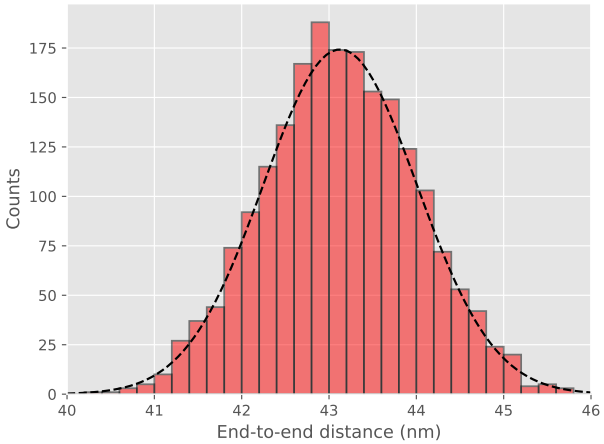


Figure 4.7: End-to-end distance distribution for scaffold nucleotides on either side of the central gap in the Nickels force clamp.

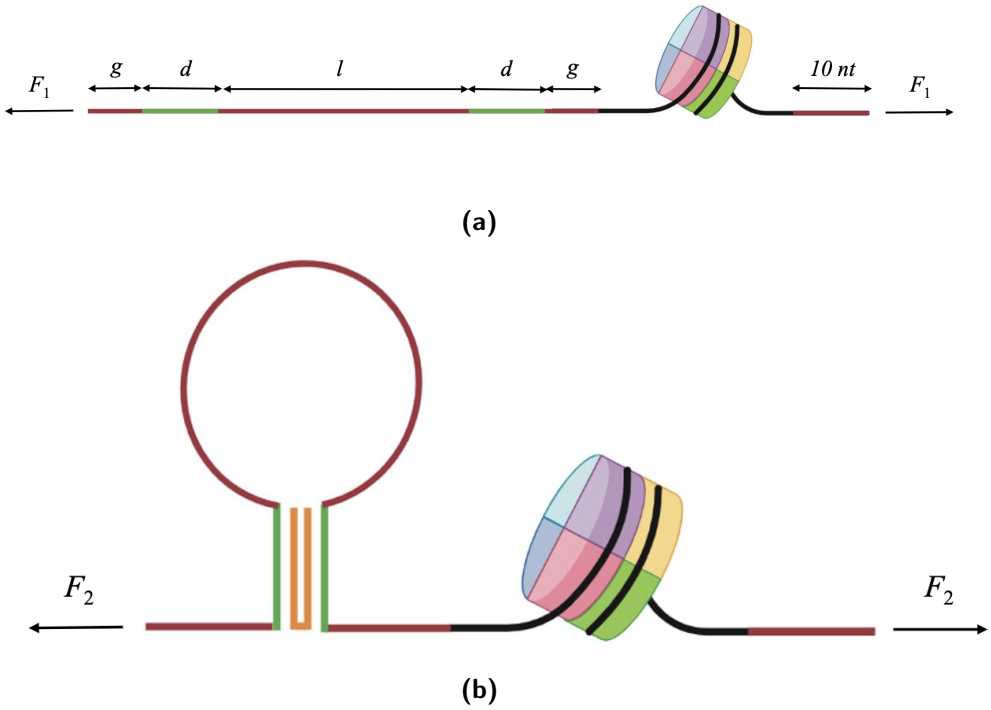
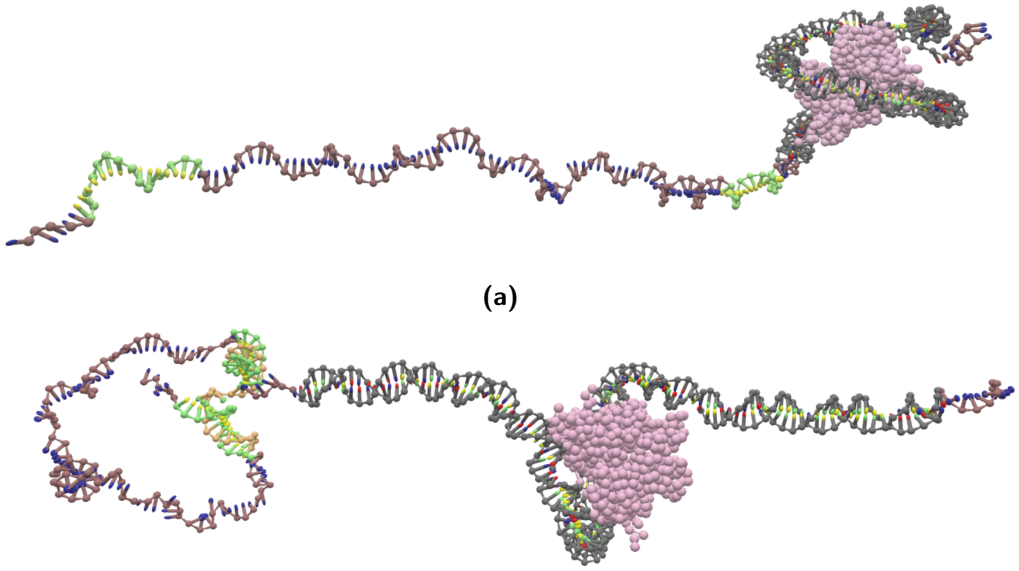


Figure 4.8: A schematic showing the strategy that is going to be used to generate force in ssDNA springs (shown in brown). The two states of the system before and after hybridisation are shown separately in (a) and (b), respectively. The orange line represents the extra strand that initiates the hybridisation process by binding to complementary regions on the ssDNA (shown as green lines). The lengths of various sections (in nucleotides) are labelled. The values of l and g can be changed keeping the length $l + 2d + 2g$ constant, to change the force.

on the ssDNA spring. The free energy gain associated with hybridisation of the orange strand needs to be sufficient to do work against the internal forces to be able to both shorten the tightrope and partially unwrap the nucleosome. This should be easy to satisfy as one can simply increase the number of nucleotides in the loop (l) until this condition is satisfied.

We require F_1 (the force prior to hybridisation of the added strand) to be low, so as not to induce unravelling. $F_1 < 2$ pN is a sensible target. We also require that F_2, the force after binding of the added strand but before any unravelling to sufficient to induce the transition. $F_2 > 4$ pN is a sensible target. But we also require that F_3



(a)

(b)

Figure 4.9: Simulation snapshots of the nucleosome in (a) the completely wrapped state at low force and (b) the partially unwrapped state where the force acting on the nucleosome is a result of ssDNA strand shortening (as described in **Section 4.4**). For the system shown here, $l = 80$ nt, $g = 5$ nt and $d = 15$ nt.

(force after the transition to the partially unravelled state is complete) is above 4 pN. The latter is a more stringent constraint as the force will decrease as nucleosome unravels. F_3 should also not be too large as to initiate a transition to the completely unwrapped state (i.e. $F_3 < 7$ pN according to **Figure 4.3**).

The main design parameter that we can use to control the magnitude force change is the effective length reduction of the ssDNA tightrope (i.e. $l + 2d$). We are currently running simulations to estimate F_1 and F_3 for different design parameters. States associated with these simulations are shown in **Figure 4.9**. For simplicity, the single-stranded regions are currently being modelled only using poly-adenines thus ruling out the formation of any secondary structure. If it is not feasible to satisfy the targets for F_1 and F_3 with the current clamp design, a new origami clamp with a larger gap could be designed.

Chapter 5

DNA origami trefoil knots

The information contained within the primary linear structure of a natural biopolymer can cause it to fold into a well-defined complex three-dimensional shape[217]. Proteins and nucleic acids, by virtue of their programmable qualities are ideal for the design and self-assembly of nanostructures with a wide range of applications[218]. One does not just need the target structure to be thermodynamically stable but to also have an energy landscape that ensures its kinetic accessibility[219]. An astonishing range of biomolecular assemblies occur in nature, out of which some of the most extraordinary and intriguing objects are knots. It has been shown that knots exist both in nucleic acids (as shown in **Figure 5.1**) and proteins[220]. In the case of proteins, knots are extremely rare and generally have a low crossing number such

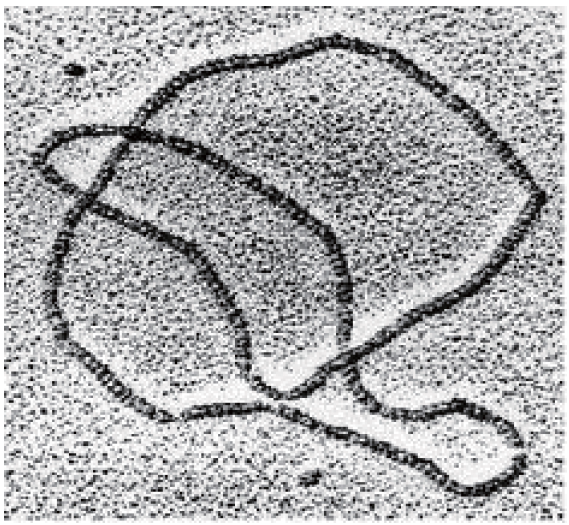


Figure 5.1: RecA coated DNA trefoil. Figure reproduced with permission from Ref. [216].

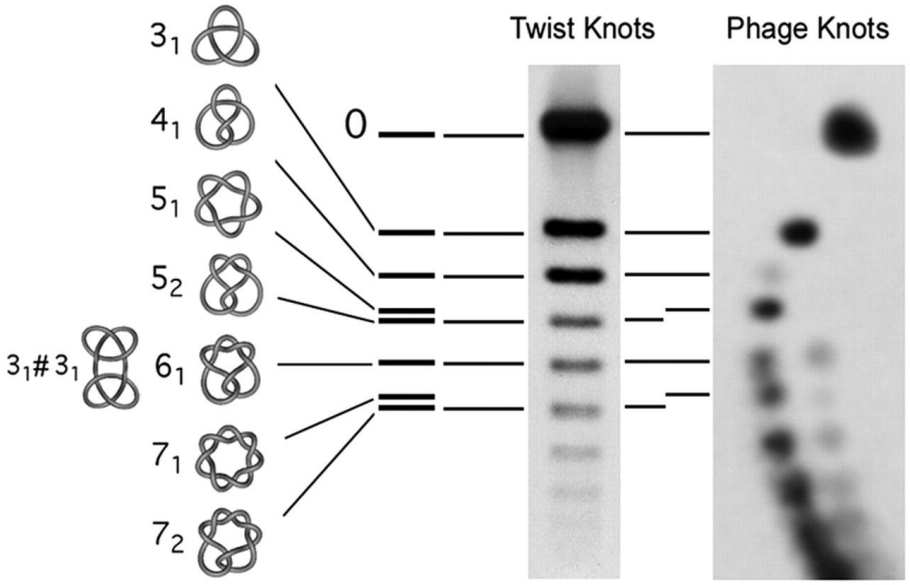


Figure 5.2: Two-dimensional gel electrophoresis of purified DNA obtained from P4 Phage particles. Knots are identified by comparing the gel velocity of twist knots with known relative migration distances (center) with that of individual knot populations resolved by electrophoresis (right). Figure reproduced with permission from Ref. [227].

as a trefoil (3_1) or a slipknot[221], although some studies have proved the occurrence of a 6_1 knot in proteins[222,223] (see **Figure 5.2**). A knot present in a protein would augment its thermal[224,225] and mechanical stability[225] but causing a protein to fold into a knotted structure remains a challenging task[226]. A particular problem in the field is to get polymers to self-assemble into knotted structures without the rupture of covalent bonds because an exact order determines the threading of the chain through loops that are already formed.

In this chapter, we have explored ways to control the detailed shape of DNA nanotubes with the aim of using them as building blocks to create toplogically complex assemblies. The fundamental aim of this project was to probe knot formation using DNA origamis in a way that is sufficiently natural that it may lead to the estimation of knot formation probabilities in experiment. Although our initial goal of achieving the assembly of a DNA origami trefoil knot has still not been been

fruitful, progress has been made in understanding how the twisting and bending of these DNA nanorods can be exploited to make structures that can potentially lead to the formation of a knot.

5.1 DNA-origami six-helix-bundles

A DNA origami is generally made up of a single long 'scaffold' strand that goes through the entire structure and many 'staple' strands that specifically bind to multiple regions of the scaffold thus constraining it in a particular conformation. A single staple strand may bind with many scaffold segments. A 'helix bundle' is essentially a collection of double helices that run parallel to each other. There are points in the structure where a strand passes from one helix to another and such points are called crossovers or junctions. Generally in DNA origamis, two strands pass between helices at a junction and such crossovers are generally termed as 'double' crossovers. Origamis can be designed from a square or hexagonal lattice of helices wherein crossovers can occur between adjacent helices. The six-helix-bundles (6HB) used in this study are based on a hexagonal lattice, the cross section of which is shown in **Figure 5.3**. Each helix has three nearest neighbours in this lattice and antiparallel crossovers to each of them can occur every 7 bp. The nanotubes used here can be dissected into well defined 'slices' transverse to the contour, where the number of base pairs in each slice corresponds to the number of helices in the bundle.

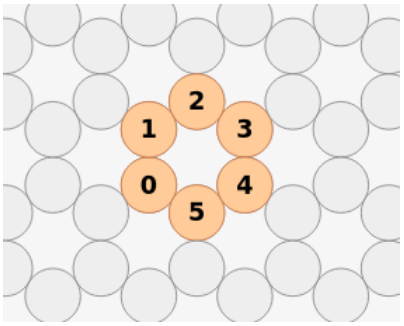


Figure 5.3: The cross section of a honeycomb-like lattice in caDNAno.

5.2 Twisted and bent nanotubes

The six-helix-bundle DNA nanotubes discussed in **Section 5.1** can be programmed to bend or twist in a particular fashion. The technique outlined by Dietz *et al.* for folding DNA into complex shapes[46] has been put to use here. DNA in a honeycomb lattice is spatially constrained and deletion or insertion of a base pair leads to local over-/under-winding along with the accumulation of tensile and compressive strains, respectively. The over-/under-winding deformations lead to global twisting whereas tensile and compressive strains are relieved by global bending. If the placement of either insertions or deletions of base pairs is symmetric, the bending deformations cancel out and twisting deformations are reinforced, resulting in purely twisted origamis. In a similar manner, purely bent origamis can be obtained using equal numbers of appropriately placed insertions and deletions in a bundle's cross section. It must be noted that structures produced using this technique will be intrinsically bent or twisted without the influence of any external force. **Figure 5.4** illustrates the method outlined above.

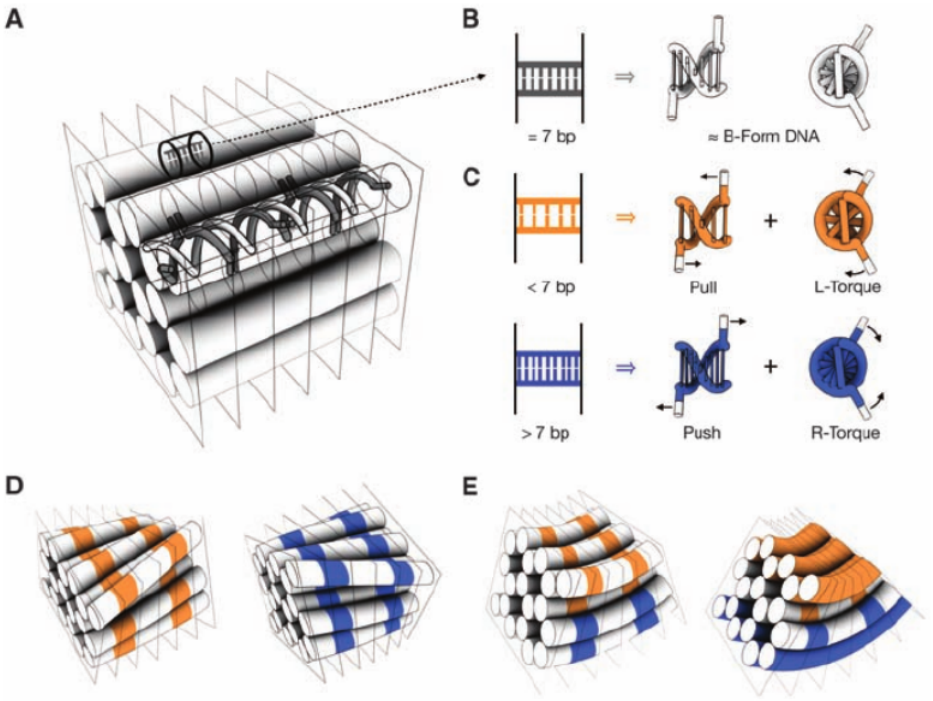


Figure 5.4: Controlling the twist and curvature in DNA origami helix bundles. (a) Helices run parallel in a honeycomb-like lattice where crossovers between neighbouring helices occur every 7 bp. The crossover planes divide the helix bundle to small sections out of which one is highlighted. (b) A section of a helix with 7 base pairs between two consecutive crossover planes, exerting no stress on its neighbours. (c) Above, a section with 5 base pairs exerting a left-handed torque on its neighbours. Below, a section with 9 base pairs exerting a right-handed torque on its neighbours. (d) Deletion/insertion of base pairs (shown in orange/blue) in appropriate positions leads to global left-/right-handed twisting, respectively. (e) A combination of appropriately placed deletions (shown in orange) and insertions (shown in blue) can be used to induce global bending. Figure adapted from Ref. [46].

5.2.1 Completely untwisted 6HB

Characterizing the twist in a DNA-origami nanotube has been made easier by Chhabra *et al.*[179], the protocol for which involves the calculation of the average value of a quantity called Θ_3. $\langle\Theta_3\rangle = 0$ implies no twist whereas a positive/negative $\langle\Theta_3\rangle$ implies a right-/left-handed twist. The current oxDNA potential has been parameterised to reproduce the results of experiments performed by Dietz *et al.*[46] according to which origamis based on the hexagonal lattice with no insertions or deletions can

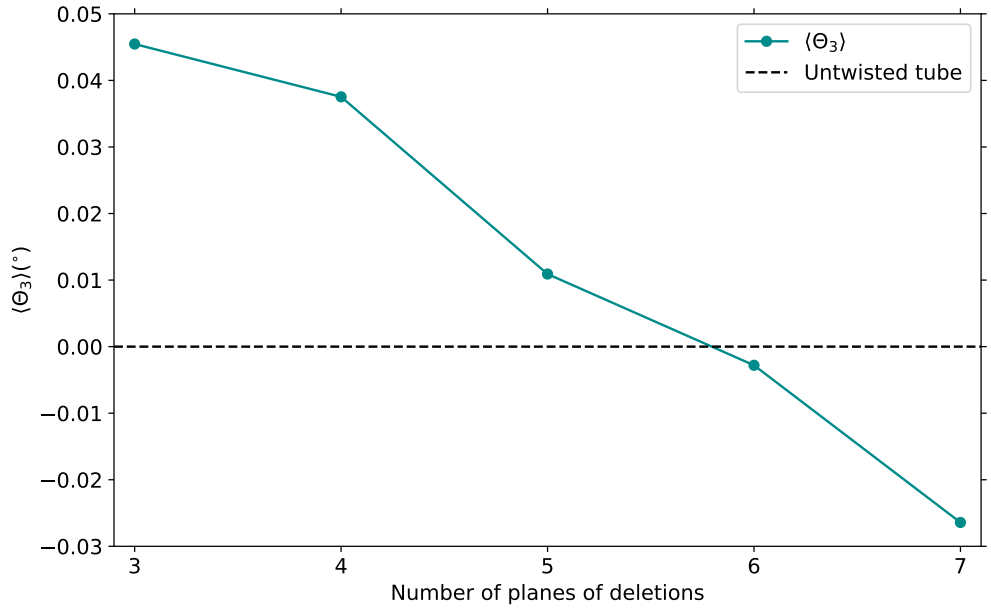


Figure 5.5: A measure of twist as a function of the number of planes of deletions in a 6HB origami.

be expected to be untwisted. However, this is not exactly true and the 6HB origami with no insertion or deletion was found to be slightly right-handed, with $\langle\Theta_3\rangle = 0.046$[179]. To get closer to a zero-twist 6HB, evenly spaced planes of deletions were used. The number of planes of deletions was varied in the range of 3-7. A plot of $\langle\Theta_3\rangle$ against the number of planes of deletions is shown in **Figure 5.5**. Assuming the number of planes of deletions to be an integer, 6 planes were found to be sufficient to give rise to an approximately untwisted rod with $\langle\Theta_3\rangle$ = -0.003. This corresponds to a left-handed twist of $\sim 3.5°$ along the length of the 6HB.

5.3 Calculation of curvature and torsion

The spatial orientation and geometry of a space curve is often denoted by its curvature and torsion. Curvature is the amount by which a curve deviates from

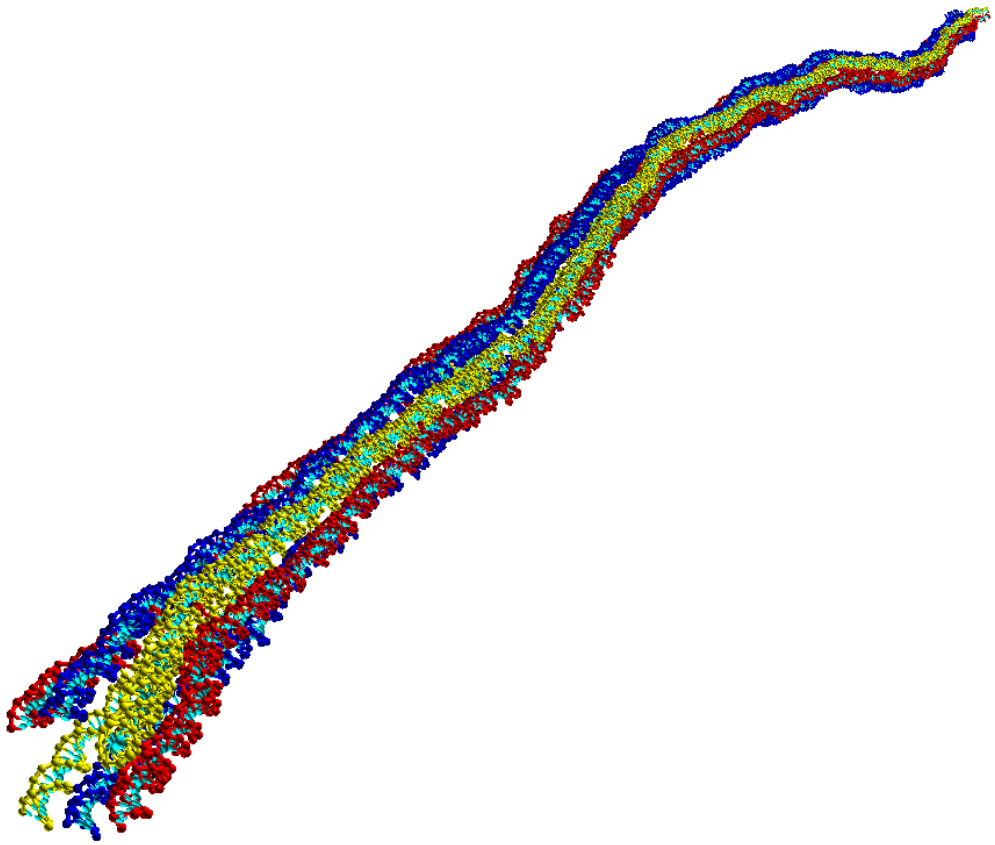


Figure 5.6: A snapshot of the 6HB origami with 6 planes of deletions. Helices diagonally opposite to each other in the hexagonal lattice are coloured the same.

being a straight line, or a surface deviates from being a plane. Curvature can also be defined as the magnitude of the acceleration of a particle moving with unit speed along the curve. Torsion measures how sharply a curve is twisting out of the plane of curvature.

For any space curve $\mathbf{C}$, with an arc length parameterisation $\gamma(s)$, the unit tangent vector $\mathbf{T}$ is defined as

$$\mathbf{T}(s) = \gamma'(s) = \frac{d\gamma(s)}{ds}. \tag{5.1}$$

The curvature κ is defined as

$$\kappa(s) = ||\mathbf{T}'(s)|| = ||\gamma''(s)||. \tag{5.2}$$

One can also define the principal normal vector $\mathbf{N}$ and the binormal vector $\mathbf{B}$ as

$$\begin{aligned} \mathbf{N}(s) &= \frac{\mathbf{T}'(s)}{\kappa}, \\ \mathbf{B}(s) &= \mathbf{T}(s) \times \mathbf{N}(s). \end{aligned} \tag{5.3}$$

The torsion τ is given by the following equation

$$\tau(s) = -\mathbf{N}(s) \cdot \mathbf{B}'(s), \tag{5.4}$$

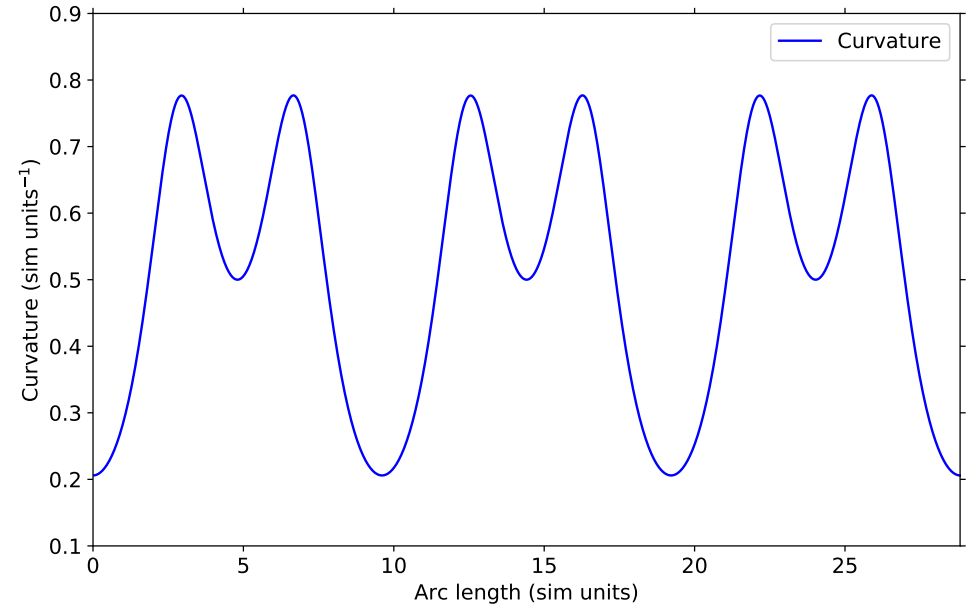
where $\mathbf{B}'(s)$ denotes the derivative of the binormal vector with respect to the arc length.

A simple parametric equation of a trefoil knot with parameter t is given by

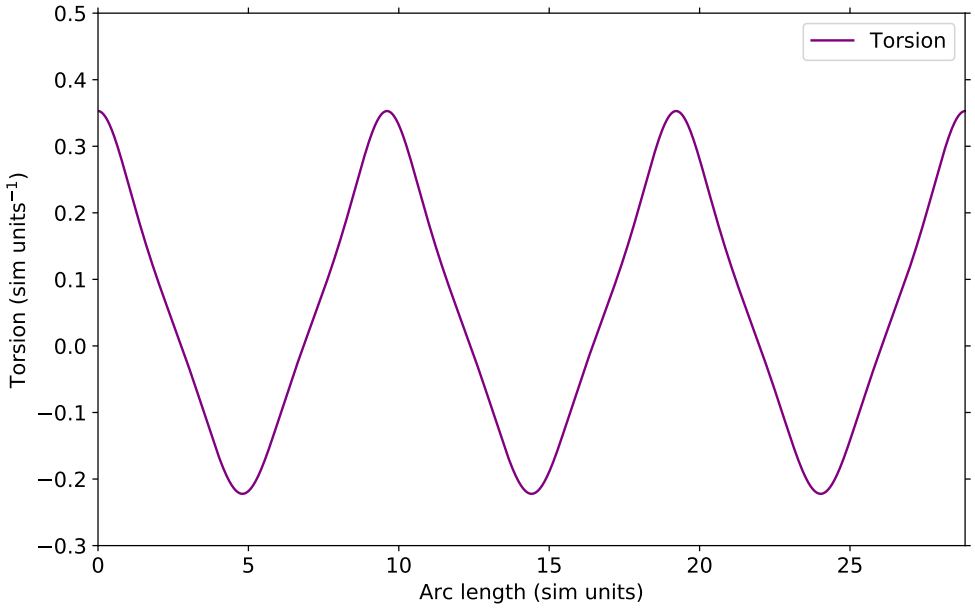
$$\begin{aligned} x(t) &= \sin(t) + 2\sin(2t), \\ y(t) &= \cos(t) - 2\cos(2t), \\ z(t) &= -\sin(3t), \end{aligned} \tag{5.5}$$

where $0 \leq t \leq 2\pi$. The curvature and torsion of the trefoil knot described by **Equation 5.5** as a function of the arc length is given in **Figure 5.7**.

To calculate, the curvature and torsion of a 6HB monomer, the centre of mass of each 'slice' perpendicular to the long axis of the origami was calculated. The centre of masses were interpolated using a cubic spline with arc length parameterisation, and the curvature and torsion were calculated using the same method described above. **Figure 5.8** shows the curvature and torsion of one of the 6HB designs as a function of the arc length. Ideally, the curvature and torsion profile of our monomer



(a)



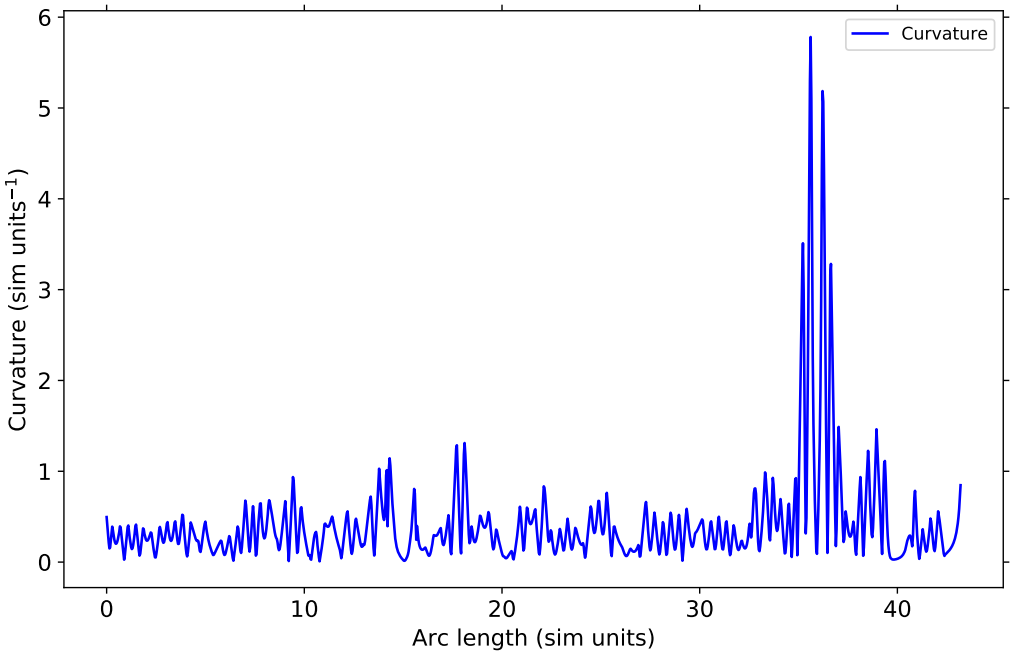
(b)

Figure 5.7: (a) Curvature and (b) torsion of a trefoil knot as a function of the arc length.

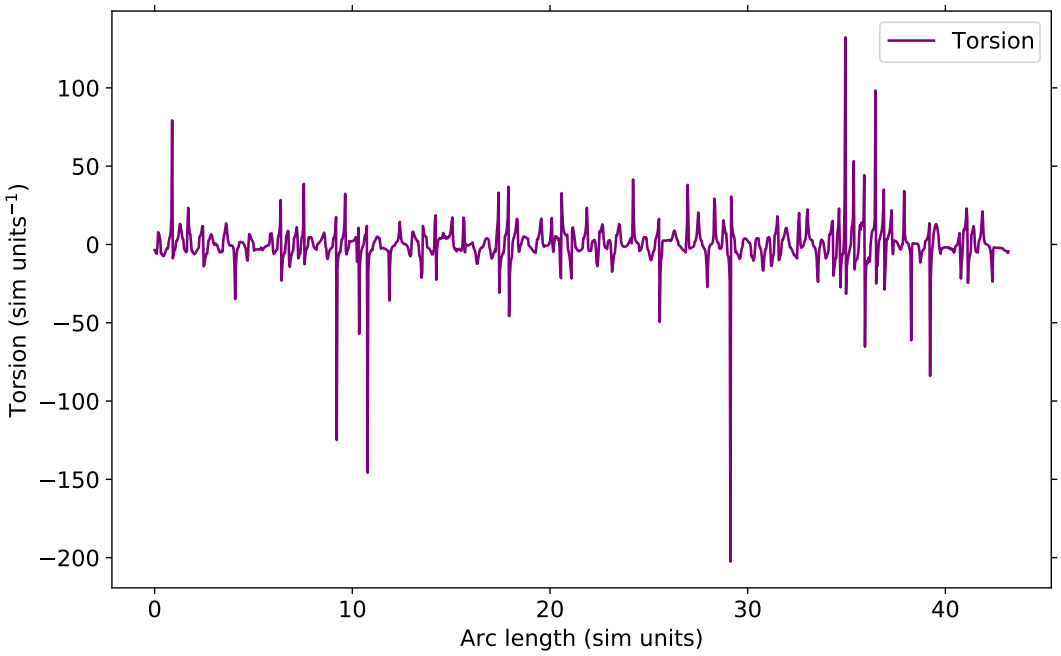
should have been similar to one-third of that of a trefoil knot but as is evident from **Figure 5.8**, the data is extremely noisy and no concrete inference could be drawn from it.

5.4 Assembling 6HBs

As mentioned in **Section 5.2**, tuning the twist and curvature of the 6HBs can be done using insertions and deletions. A slice with only insertions/deletions can give rise to twist whereas an equal number of both in a slice can lead to bending. For all our designs, '*bending*' and '*twisting*' planes were introduced in the design using caDNAno, following which the structure was minimised, relaxed and simulated till it attained a desired configuration. For most of the designs discussed henceforth, the simulation ran for 2×10^8 steps with an integration time step of 15.15×10^{-15}s. The process of monomer design turned out to be very unpredictable, in the sense that the expected configuration of a monomer differed significantly from what was actually obtained after the simulation. This was because of the non-trivial interplay between twisting and bending in the case of such large nanostructures. In order to perform the assembly, individual 6HBs with sticky ends had to be designed. These sticky ends, in oxDNA, are nothing but segments of single-stranded DNA that can bind with its complementary sequence on a separate strand. caDNAno was used to design the sticky ends and their sequences were set using a Python script[228]. After the monomers were appropriately placed, individual mutual traps between complementary bases in the ssDNA sticky ends were used to pull them together. The equilibrium distance of these traps was set to 1.2 simulation units and their



(a)



(b)

Figure 5.8: (a) Curvature and (b) torsion of a 6HB monomer (Monomer C, as described later in **Section 5.4.1**).

Monomer	Number of bending planes	Number of twisting planes
A	30	4
B	30	6
C	28	5

Table 5.1: Details of monomer design.

rates were calculated from the initial distance between them and the length of the simulation.

5.4.1 Individual monomer designs

This section deals with three monomer designs to understand the interplay between twisting and bending. It must be noted that to arrive at a design whose curvature was approximately similar to that of a trefoil knot, extensive trial and error was involved. Multiple designs were separately created and simulated, after which all but three (monomers **A**, **B** and **C**) were rejected . Monomer **A** was designed to be purely bent i.e. completely planar and the other two (**B** and **C**) were designed from **A** by incorporating a small twist in the structure. Each of the three designs: **A**, **B** and **C** were made in caDNAno (design details given in **Table 5.1**), relaxed and then simulated (shown in **Figure 5.9**). For all three monomers, the bending planes were evenly distributed throughout the structure whereas the twisting planes were evenly spaced across half the structure.

5.4.2 Kabsch algorithm for positioning 6HBs

In order to avoid entanglements and very high forces, the 6HBs (monomers) used to form the trefoil knot (trimer) must be spatially arranged so that their ends are correctly oriented. To achieve the appropriate spatial arrangement, each monomer



(a) Monomer A



(b) Monomer B



(c) Monomer C

Figure 5.9: Visualisation of each monomer from two different angles.

must be rotated so that their position roughly coincides with the points of the trefoil that it can possibly give rise to. The KABSCH algorithm[229] was used to find the optimal rotation matrix which when applied to the coordinates of a monomer, rotates the entire structure in a way that it occupies one-third of a trefoil knot of appropriate size. After the monomer had relaxed, the centre of mass of each slice was calculated and stored in a matrix **A**. It must be noted that a trefoil knot is a topological structure that can have infinitely many spatial orientations, meaning no one equation or set of equations can perfectly describe it. Nevertheless, to apply the KABSCH algorithm, a desired trefoil knot had to be defined for which the knot described in **Equation 5.5** was scaled until the size matched that of the monomer. The coordinates of this knot was stored in a matrix **B**.

From the matrices **A** and **B** defined above, the cross-covariance matrix **H** was computed as $\mathbf{H} = \mathbf{A}^{\mathbf{T}}\mathbf{B}$,where the superscript $^{\mathbf{T}}$ denotes the transpose of the matrix. Singular value decomposition of **H** gave the matrices **U**, **Σ** and **V** as $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$. A correction factor d, to ensure a right-handed coordinate system was calculated as $d = \text{sgn}(\det(\mathbf{V}\mathbf{U}^{\mathbf{T}}))$, where det() indicates the determinant and sgn() indicates the sign function. The optimal rotation matrix **R** was calculated by the following equation

$$\mathbf{R} = \mathbf{V}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{pmatrix}\mathbf{U}^{\mathbf{T}}. \tag{5.6}$$

The matrix **R** was applied onto all the position, backbone and normal vectors of each nucleotide in the configuration file of the monomer. Copies of the monomer

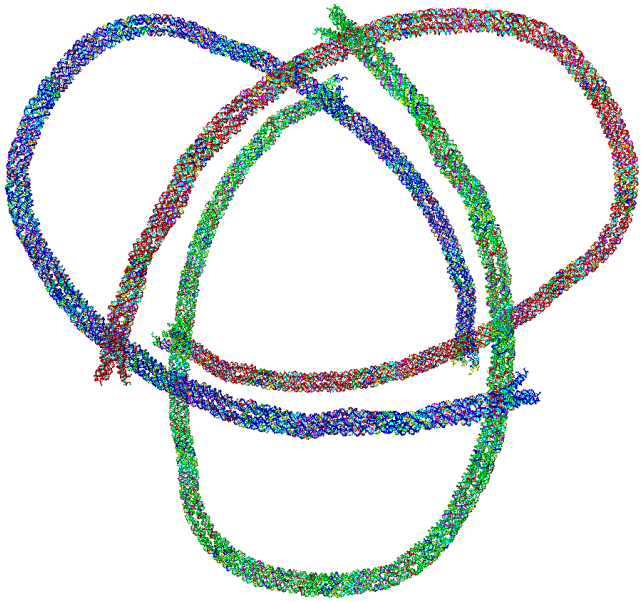


Figure 5.10: A snapshot of three monomers, each oriented by a rotation matrix calculated from the KABSCH algorithm.

were made in the system and for each copy a separate matrix **R** was calculated and applied onto it in a similar way. **Figure 5.10** shows the system generated after the above protocol is performed on monomer **C** from **Section 5.4.1**. Although it looked like the system resembled a trefoil knot, there was a significant overlap between the monomers, due to which they were separated using oxView to obtain a configuration with no overlaps (shown in **Figure 5.11**).

The system created above was simulated with mutual traps between the nucleotides in the sticky ends. A structure was obtained where two out of three junctions had formed but the third was far from coming together (shown in **Figure 5.12**). Moreover, the monomers had not threaded through and thus the structure was not knotted. Similar systems were created and simulated for monomers **A** and **B** but they did not give structures that could possibly form a knot.

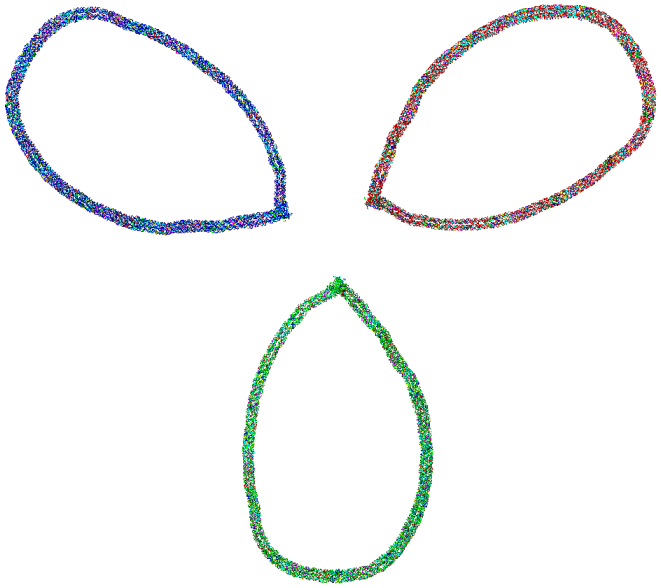


Figure 5.11: The monomers in **Figure 5.10** after re-positioning in oxView.

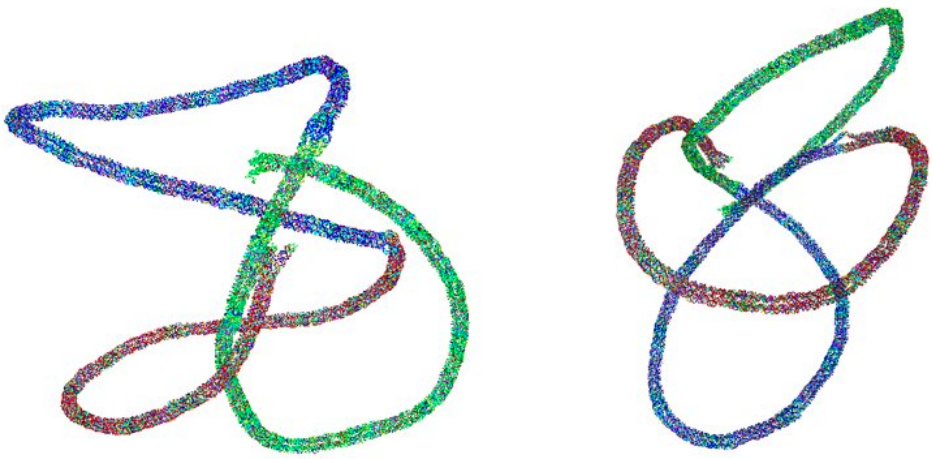


Figure 5.12: Snapshots of the final configuration of the trimer after an attempt was made to assemble it from individual 6HB monomers.

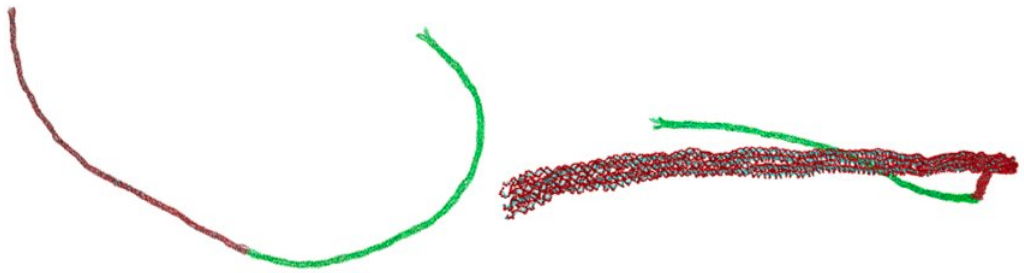


Figure 5.13: Snapshots of a monomer for a 6-particle trefoil knot composed of two origamis with different handedness of twist.

5.5 6-particle trefoil

This section outlines another approach that could possibly lead to the formation of a DNA origami trefoil knot. It was realised that the monomers designed using the previous protocol must have both right and left handed twist (see **Figure 5.7**) so that they could attain the appropriate conformation. This made these structures extremely complex to design. Instead, it might be easier if one tried to make a 6-particle knot with two relatively simpler particles. It would also potentially mean that the individual particles only possess one handedness of twist. The assembly of a 6-particle knot over a 3-particle knot can pose fewer problems because each particle would have only kind of twist and thus have less internal stress. A monomer of a 6-particle knot would be made up of two DNA-origamis: one slightly bent with strong left-handed twist (red in **Figure 5.13**) and the other with stronger bending but with a slight right-handed twist (green in **Figure 5.13**). The problem in this approach was that it was difficult to design the junction so that the curves represented by each particle came together in a way that was continuous across the junction.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In conclusion, this thesis has shown the versatility of the DNA origami approach to generate novel functional devices and as a building block for complex assemblies. It also highlights the role that simulations can play in this field, through the extra insights into the complex coupled responses of origami devices and how it can be used as a tool to optimise designs. Another important inference that can be drawn from the results presented in this thesis is the ability of simulations to guide experimental design. Our studies indicate that oxDNA can be used to explore how conformational changes affect the tension experienced by any molecule in the constant length ensemble.

The work on the origami-polymer force clamp provides a new route for the mechanical manipulation of ensembles of target molecules. One major innovation in this study was the use of a DNA origami as a programmable substrate for the organisation of target molecules. Although the origami was initially assumed to be perfectly rigid, our data shows that the force due to particle collapse can cause the origami to undergo significant bending, and the coupling between the different components leads to a complex force response that would have been hard to appreciate without our simulations. A typical OPFC manipulates $\sim 10^{11}$ target molecules which is significantly higher than the throughput level of common parallelised force spec-

troscopy tools. Another important advantage of the OPFC is its amenability for integration with various spectroscopic analysis methods.

Our work on the simulation of a nucleosome in oxDNA has enabled us to appreciate protein-DNA interactions. The ANM-oxDNA model has been used to develop models of the nucleosome in its fully and partially wrapped states. We are in the process of using these models to characterise force clamp designs that might eventually allow the structure of the partially unwrapped state to be experimentally determined.

Attempting to design DNA origamis that can self-assemble into a knot is intricate and tricky. Although the interplay between twist and bend is non-trivial, by manipulating the number of base pair insertions and deletions, a wide range of nanostructures were obtained. The origamis were programmed to have complementary sticky ends that could potentially bind with each other and result in the formation of a knot. An attempt was made to form a trimer from the monomers that had close-to-correct geometries by using mutual traps. Although knotted structures did not form, it was possible to understand the reason behind it and further modifications have been made in the design to address these issues.

6.2 Future Work

Our initial investigation of an origami-polymer force-clamp using oxDNA elucidates the potential that lies within these hybrid nanoscopic devices. Further studies of the polymer-force clamp may consider different types of DNA origami substrates with varying compliance. It would be interesting to understand how the force exerted

by the polymer sphere is distributed among the target molecules and also the DNA origami substrate and spot any patterns within them. In order to make the method more versatile, other target molecules with known mechanical responses could be examined. It would have been nice to have been able to compare the experimental time-resolved mechanical unfolding results. Although computationally very challenging due to the size and timescales involved, hopefully subsequent investigations will make that possible. Our simulations focused on two discrete thermodynamic states of the system and it would be interesting to study the transition between the two. For example, the hairpin unfolding mechanism may involve novel features due to the coupling to the mechanical response of the origami. Once time-resolved simulations can be performed, it would also be possible to study the effect of the particle collapse rate on the loading rate of the origami-polymer force clamp.

For the work on the nucleosomes, the plan going forward is to use simulations to measure the forces exerted on the nucleosome and its partially wrapped state. This will allow the force clamp design to be optimised so that it can constrain the nucleosome to be in its partially unwrapped state. The determination of the structure of this state by cryoEM would be a major breakthrough.

To bring about the assembly of DNA origami units into a higher order structure is a challenging task. Although 6HB origami blocks are, in many ways, ideal building blocks for the assembly of complex linear structures, the precise control of twist and bend is a difficult task. We hope that the ground work we have done here will pave the way for the successful design of knotted origami structures, such as a trefoil.

References

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

[2] R. Dahm, *Human Genetics* **2008**, *122*, 565–581.

[3] M. E. Jones, *The Yale journal of biology and medicine* **1953**, *26*, 80–97.

[4] P. A. Levene, *J. Bio. Chem.* **1919**, *608*, 415–424.

[5] E. Chargaff, R. Lipshitz, C. Green, *J. Bio. Chem.* **1952**, *1*, 155–160.

[6] W. T. Astbury, F. O. Bell, *Nature* **1938**, *141*, 747–748.

[7] R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740–741.

[8] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

[9] K. Hoogsteen, *Acta Crystallographica* **1963**, *16*, 907.

[10] S. Neidle, *Principles of Nucleic Acid Structure*, Elsevier, **2008**.

[11] Wikipedia, A-DNA — Wikipedia, The Free Encyclopedia, **2021**.

[12] S. C. Ha, K. Lowenhaupt, A. Rich, Y.-G. Kim, K. K. Kim, *Nature* **2005**, *437*, 1183–1186.

[13] A. Suram, J. K. S. Rao, L. K. S., V. M. A., *Neuromolecular Med.* **2002**, *2*, 289–298.

[14] S. C. Harvey, *J. Struct. Biol.* **2015**, *189*, 1–8.

[15] M. Oram, C. Sabanayagam, L. W. Black, *J. Mol. Bio.* **2008**, *381*, 61–72.

[16] K. Ray, C. R. Sabanayagam, J. R. Lakowicz, L. W. Black, *Virology* **2010**, *398*, 224–232.

[17] A. Ghosh, M. Bansal, *Acta Crystallogr D* **2003**, *59*, 620–626.

[18] D. Elliott, *Molecular biology of RNA*, Oxford University Press, Oxford, **2016**.

[19] M. Egli, N. Usman, S. G. Zhang, A. Rich, *Proc. Natl. Acad. Sci.* **1992**, *89*, 534–538.

[20] O. Romainczyk, B. Endeward, T. F. Prisner, J. W. Engels, *Mol. BioSyst.* **2011**, *7*, 1050–1052.

[21] C. Yapijakis, *In Vivo* **2009**, *23*, 507–514.

[22] G. Mendel, *Versuche über Pflanzen-Hybriden*, Brünn :Im Verlage des Vereines, **1866**.

[23] O. T. Avery, C. M. MacLeod, M. McCarty, *Exp. Med.* **1944**, *79*, 137–158.

[24] E. Schrödinger, *What is life?*, Cambridge University Press, Cambridge New York, **1992**.

[25] B. Alberts, *Molecular biology of the cell*, Garland Science, Taylor and Francis Group, New York, NY, **2015**.

[26] J. S. Mattick, I. V. Makunin, *Human Molecular Genetics* **2006**, *15*, R17–R29.

[27] F. Crick, *Nature* **1970**, *227*, 561–563.

[28] E. Pennisi, *Science* **2012**, *337*, 1159–1161.

[29] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274–293.

[30] B. McClintock, *Science* **1984**, *226*, 792–801.

[31] D. Noble, *Experimental Physiology* **2013**, *98*, 1235–1243.

[32] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 964–967.

[33] F. L. Holmes, *Trends Biochem. Sci.* **1998**, *23*, 117–120.

[34] M. Delbruck, *Proc. Natl. Acad. Sci.* **1954**, *40*, 783–788.

[35] R. Weil, J. Vinograd, *Proc. Natl. Acad. Sci.* **1963**, *50*, 730–738.

[36] R. Sinden, *DNA structure and function*, Academic Press, San Diego, **1994**.

[37] A. Bates, *DNA topology*, Oxford University Press, Oxford New York, **2005**.

[38] S. Yaron, G. L. Kolling, L. Simon, K. R. Matthews, *Appl. Environ. Microbiol.* **2000**, *66*, 4414–4420.

[39] N. C. Seeman, *J. Theor. Biol.* **1982**, *99*, 237–247.

[40] P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302.

[41] Y. Ke, L. L. Ong, W. M. Shih, P. Yin, *Science* **2012**, *338*, 1177–1183.

[42] E. Benson, A. Mohammed, J. Gardell *et al.*, *Nature* **2015**, *523*, 441–444.

[43] M. Matthies, N. P. Agarwal, T. L. Schmidt, *Nano Lett.* **2016**, *16*, 2108–2113.

[44] F. Zhang, S. Jiang, S. Wu *et al.*, *Nat. Nanotechnol.* **2015**, *10*, 779–784.

[45] S. M. Douglas, H. Dietz, T. Liedl *et al.*, *Nature* **2009**, *459*, 414–418.

[46] H. Dietz, S. M. Douglas, W. M. Shih, *Science* **2009**, *325*, 725–730.

[47] D. Han, S. Pal, J. Nangreave *et al.*, *Science* **2011**, *332*, 342–346.

[48] T. Liedl, B. Högberg, J. Tytell, D. E. Ingber, W. M. Shih, *Nat. Nanotechnol.* **2010**, *5*, 520–524.

[49] X. Fu, F. Peng, J. Lee *et al.*, *Top. Curr. Chem.* **2020**, *378*, 21.

[50] X.-c. Bai, T. G. Martin, S. H. W. Scheres, H. Dietz, *Proc. Natl. Acad. Sci.* **2012**, *109*, 20012–20017.

[51] R. Jungmann, M. S. Avendaño, J. B. Woehrstein *et al.*, *Nat. Methods* **2014**, *11*, 313–318.

[52] J.-P. J. Sobczak, T. G. Martin, T. Gerling, H. Dietz, *Science* **2012**, *338*, 1458–1461.

[53] Y. Ke, G. Bellot, N. V. Voigt, E. Fradkov, W. M. Shih, *Chem. Sci.* **2012**, *3*, 2587.

[54] K. F. Wagenbauer, C. H. Wachauf, H. Dietz, *Nat. Commun.* **2014**, *5*, 3961.

[55] X. Wei, J. Nangreave, Y. Liu, *Acc. Chem. Res.* **2014**, *47*, 1861–1870.

[56] K. E. Dunn, F. Dannenberg, T. E. Ouldridge *et al.*, *Nature* **2015**, *525*, 82–86.

[57] D. Svozil, P. Hobza, J. Šponer, *J. Phys. Chem. B* **2009**, *114*, 1191–1203.

[58] A. Mládek, M. Krepl, D. Svozil *et al.*, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7295.

[59] J. Šponer, J. E. Šponer, A. Mládek *et al.*, *Methods* **2013**, *64*, 3–11.

[60] C. A. Laughton, S. A. Harris, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 590–600.

[61] A. Pérez, F. J. Luque, M. Orozco, *Acc. Chem. Res.* **2011**, *45*, 196–205.

[62] F. Oteri, M. Falconi, G. Chillemi *et al.*, *J. Phys. Chem. C* **2011**, *115*, 16819–16827.

[63] C. Maffeo, B. Luan, A. Aksimentiev, *Nucleic Acids Res.* **2012**, *40*, 3812–3821.

[64] J. Yoo, A. Aksimentiev, *Proc. Natl. Acad. Sci.* **2013**, *110*, 20099–20104.

[65] C. Maffeo, J. Yoo, J. Comer *et al.*, *J. Condens. Matter Phys.* **2014**, *26*, 413101.

[66] O. Kratky, G. Porod, *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 1106–1122.

[67] T. Odijk, *Macromolecules* **1995**, *28*, 7016–7018.

[68] H. Yamakawa, *Macromolecules* **1977**, *10*, 692–696.

[69] J. Gore, Z. Bryant, M. Nöllmann *et al.*, *Nature* **2006**, *442*, 836–839.

[70] S. K. Nomidis, E. Skoruppa, E. Carlon, J. F. Marko, *Phys. Rev. E* **2019**, *99*, 032414.

[71] T. Sun, V. Minhas, N. Korolev *et al.*, *Front. Mol. Biosci.* **2021**, *8*, 159.

[72] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. de Pablo, *J. Chem. Phys.* **2013**, *139*, 144903.

[73] P. D. Dans, A. Zeida, M. R. Machado, S. Pantano, *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725.

[74] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, S. J. Marrink, *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945.

[75] C. Maffeo, A. Aksimentiev, *Nucleic Acids Res.* **2020**, *48*, 5135–5146.

[76] T. E. Ouldridge, I. G. Johnston, A. A. Louis, J. P. K. Doye, *J. Chem. Phys.* **2009**, *130*, 065101.

[77] J. M. Arbona, J. Elezgaray, J. P. Aimé, *EPL (Europhysics Letters)* **2012**, *100*, 28006.

[78] D.-N. Kim, F. Kilchherr, H. Dietz, M. Bathe, *Nucleic Acids Res.* **2011**, *40*, 2862–2868.

[79] K. Pan, D.-N. Kim, F. Zhang *et al.*, *Nat. Commun.* **2014**, *5*, 5578.

[80] A. Savelyev, G. A. Papoian, *Proc. Natl. Acad. Sci.* **2010**, *107*, 20340–20345.

[81] Q. Cao, C. Zuo, Y. Ma, L. Li, Z. Zhang, *Soft Matter* **2011**, *7*, 506–514.

[82] N. Korolev, D. Luo, A. Lyubartsev, L. Nordenskiöld, *Polymers* **2014**, *6*, 1655–1675.

[83] M. C. Linak, R. Tourdot, K. D. Dorfman, *J. Chem. Phys.* **2011**, *135*, 205102.

[84] J. C. Araque, A. Z. Panagiotopoulos, M. A. Robert, *J. Chem. Phys.* **2011**, *134*, 165103.

[85] A. Morriss-Andrews, J. Rottler, S. S. Plotkin, *J. Chem. Phys.* **2010**, *132*, 035105.

[86] C. Maffeo, T. T. M. Ngo, T. Ha, A. Aksimentiev, *J. Chem. Theory Comput.* **2014**, *10*, 2891–2896.

[87] T. Cragnolini, P. Derreumaux, S. Pasquali, *J. Phys. Chem. B* **2013**, *117*, 8047–8060.

[88] V. Ortiz, J. J. de Pablo, *Phys. Rev. Lett.* **2011**, *106*, 238107.

[89] D. M. Hinckley, J. P. Lequieu, J. J. de Pablo, *J. Chem. Phys.* **2014**, *141*, 035102.

[90] W. Lu, C. Bueno, N. P. Schafer *et al.*, *PLOS Comput. Biol.* **2021**, *17*, 1–21.

[91] A. Davtyan, N. P. Schafer, W. Zheng *et al.*, *J. Phys. Chem. B* **2012**, *116*, 8494–8503.

[92] A. N. Kapanidis, T. Strick, *Trends Biochem. Sci.* **2009**, *34*, 234–243.

[93] H. Rohrer, *Surf. Sci.* **1994**, *299-300*, 956–964.

[94] J. Adamcik, J.-H. Jeon, K. J. Karczewski, R. Metzler, G. Dietler, *Soft Matter* **2012**, *8*, 8651.

[95] J.-H. Jeon, J. Adamcik, G. Dietler, R. Metzler, *Phys. Rev. Lett.* **2010**, *105*, 208101.

[96] A. Y. Lushnikov, V. N. Potaman, Y. L. Lyubchenko, *Nucleic Acids Res.* **2006**, *34*, e111–e111.

[97] B. D. Sattin, M. C. Goh, *Biophys. J.* **2004**, *87*, 3430–3436.

[98] Y. L. Lyubchenko, L. S. Shlyakhtenko, T. Ando, *Methods* **2011**, *54*, 274–283.

[99] Y. Zhang, X. Hu, J. Sun *et al.*, *Microsc. Res. Tech.* **2010**, *74*, 614–626.

[100] O. Ordu, A. Lusser, N. H. Dekker, *Biophys. Rev.* **2016**, *8*, 33–49.

[101] C. Bustamante, Z. Bryant, S. B. Smith, *Nature* **2003**, *421*, 423–427.

[102] I. D. Vlaminck, C. Dekker, *Annu. Rev. Biophys.* **2012**, *41*, 453–472.

[103] Z. Bryant, F. C. Oberstrass, A. Basu, *Curr. Opin. Struct. BiologBiol.y* **2012**, *22*, 304–312.

[104] R. A. Forties, M. D. Wang, *Cell* **2014**, *157*, 4–7.

[105] M. Wang, H. Yin, R. Landick, J. Gelles, S. Block, *Biophys. J.* **1997**, *72*, 1335–1346.

[106] S. B. Smith, Y. Cui, C. Bustamante, *Science* **1996**, *271*, 795–799.

[107] A. Mossa, M. Manosas, N. Forns, J. M. Huguet, F. Ritort, *J. Stat. Mech. Theory Exp.* **2009**, *2009*, P02060.

[108] M. Manosas, D. Collin, F. Ritort, *Phys. Rev. Lett.* **2006**, *96*, 218301.

[109] J.-D. Wen, M. Manosas, P. T. Li *et al.*, *Biophys. J.* **2007**, *92*, 2996–3009.

[110] S. Forth, C. Deufel, M. Y. Sheinin *et al.*, *Phys. Rev. Lett.* **2008**, *100*, 148301.

[111] A. L. Porta, M. D. Wang, *Phys. Rev. Lett.* **2004**, *92*, 190801.

[112] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, *Science* **1996**, *271*, 1835–1837.

[113] T. Strick, J.-F. Allemand, D. Bensimon, V. Croquette, *Biophys. J.* **1998**, *74*, 2016–2028.

[114] T. R. Strick, V. Croquette, D. Bensimon, *Nature* **2000**, *404*, 901–904.

[115] F. Mosconi, J. F. Allemand, D. Bensimon, V. Croquette, *Phys. Rev. Lett.* **2009**, *102*, 078301.

[116] H. Brutzer, N. Luzzietti, D. Klaue, R. Seidel, *Biophys. J.* **2010**, *98*, 1267–1276.

[117] A. Tempestini, V. Cassina, D. Brogioli *et al.*, *Nucleic Acids Res.* **2012**, *41*, 2009–2019.

[118] D. Salerno, A. Tempestini, I. Mai *et al.*, *Phys. Rev. Lett.* **2012**, *109*, 118303.

[119] H. Meng, J. Bosman, T. van der Heijden, J. van Noort, *Biophys. J.* **2014**, *106*, 1174–1181.

[120] T. R. Strick, M.-N. Dessinges, G. Charvin *et al.*, *Rep. Prog. Phys.* **2002**, *66*, 1–45.

[121] T. Ramreddy, R. Sachidanandam, T. R. Strick, *Nucleic Acids Res.* **2011**, *39*, 4275–4283.

[122] F. C. Oberstrass, L. E. Fernandes, Z. Bryant, *Proc. Natl. Acad. Sci.* **2012**, *109*, 6106–6111.

[123] J. Lipfert, G. M. Skinner, J. M. Keegstra *et al.*, *Proc. Natl. Acad. Sci.* **2014**, *111*, 15408–15413.

[124] K. Howan, A. J. Smith, L. F. Westblade *et al.*, *Nature* **2012**, *490*, 431–434.

[125] M. Manosas, M. M. Spiering, F. Ding, V. Croquette, S. J. Benkovic, *Nucleic Acids Res.* **2012**, *40*, 6187–6198.

[126] D. A. Koster, A. Crut, S. Shuman, M.-A. Bjornsti, N. H. Dekker, *Cell* **2010**, *142*, 519–530.

[127] M. Nöllmann, M. D. Stone, Z. Bryant *et al.*, *Nat. Struct. Mol. Biol.* **2007**, *14*, 264–271.

[128] P. Gross, N. Laurens, L. B. Oddershede *et al.*, *Nat. Phys.* **2011**, *7*, 731–736.

[129] G. A. King, P. Gross, U. Bockelmann *et al.*, *Proc. Natl. Acad. Sci.* **2013**, *110*, 3859–3864.

[130] A. L. Forget, S. C. Kowalczykowski, *Nature* **2012**, *482*, 423–427.

[131] S. M. Hamdan, J. J. Loparo, M. Takahashi, C. C. Richardson, A. M. van Oijen, *Nature* **2008**, *457*, 336–339.

[132] A. M. van Oijen, J. J. Loparo, *Annu. Rev. Biophys.* **2010**, *39*, 429–448.

[133] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, T. Ha, *Annu. Rev. Biochem.* **2008**, *77*, 51–76.

[134] H. Schneckenburger, *Methods Appl. Fluoresc.* **2019**, *8*, 013001.

[135] R. Vafabakhsh, T. Ha, *Science* **2012**, *337*, 1097–1101.

[136] J. Y. Lee, B. Okumus, D. S. Kim, T. Ha, *Proc. Natl. Acad. Sci.* **2005**, *102*, 18938–18943.

[137] S. Hohng, R. Zhou, M. K. Nahas *et al.*, *Science* **2007**, *318*, 279–283.

[138] T. Cordes, Y. Santoso, A. I. Tomescu *et al.*, *Biochemistry* **2010**, *49*, 9171–9180.

[139] M. Lee, S. H. Kim, S.-C. Hong, *Proc. Natl. Acad. Sci.* **2010**, *107*, 4985–4990.

[140] S. Hohng, S. Lee, J. Lee, M. H. Jo, *Chem. Soc. Rev.* **2014**, *43*, 1007–1013.

[141] X. S. Xie, P. J. Choi, G.-W. Li, N. K. Lee, G. Lia, *Annu. Rev. Biophys.* **2008**, *37*, 417–444.

[142] H. C. Andersen, *J. Chem. Phys.* **1980**, *72*, 2384–2393.

[143] J. Russo, P. Tartaglia, F. Sciortino, *J. Chem. Phys.* **2009**, *131*, 014504.

[144] L. Rovigatti, P. Šulc, I. Z. Reguly, F. Romano, *J. Comput. Chem.* **2014**, *36*, 1–8.

[145] J. Yoo, A. Aksimentiev, *J. Phys. Chem. Lett.* **2011**, *3*, 45–50.

[146] S. Plimpton, **1995**, *117*, 1–19.

[147] J. P. K. Doye, H. Fowler, D. Prešern *et al.*, The oxDNA coarse-grained model as a tool to simulate DNA origami, arXiv:2004.05052, **2020**.

[148] A. Sengar, T. E. Ouldridge, O. Henrich, L. Rovigatti, P. Šulc, *Front. Mol. Biosci.* **2021**, *8*, 551.

[149] E. Skoruppa, M. Laleman, S. K. Nomidis, E. Carlon, *J. Chem. Phys.* **2017**, *146*, 214902.

[150] B. E. K. Snodin, F. Randisi, M. Mosayebi *et al.*, *J. Chem. Phys.* **2015**, *142*, 234901.

[151] A. Suma, E. Poppleton, M. Matthies *et al.*, *J. Comput. Chem.* **2019**, *40*, 2586–2595.

[152] A. Buchberger, C. R. Simmons, N. E. Fahmi, R. Freeman, N. Stephanopoulos, *J. Am. Chem. Soc.* **2019**, *142*, 1406–1416.

[153] Y. Xu, S. Jiang, C. R. Simmons *et al.*, *ACS Nano* **2019**, *13*, 3545–3554.

[154] S. Li, Q. Jiang, S. Liu *et al.*, *Nat. Biotechnol.* **2018**, *36*, 258–264.

[155] J. Jin, E. G. Baker, C. W. Wood *et al.*, *ACS Nano* **2019**, *13*, 9927–9935.

[156] A. K. Sieradzan, A. Giełdoń, Y. Yin *et al.*, *J. Comput. Chem.* **2018**, *39*, 2360–2370.

[157] G. Mishra, Y. Levy, *Proc. Natl. Acad. Sci.* **2015**, *112*, 5033–5038.

[158] C. Tan, T. Terakawa, S. Takada, *J. Am. Chem. Soc.* **2016**, *138*, 8512–8522.

[159] C. Tan, S. Takada, *J. Chem. Theory Comput.* **2018**, *14*, 3877–3889.

[160] B. Zhang, W. Zheng, G. A. Papoian, P. G. Wolynes, *J. Am. Chem. Soc.* **2016**, *138*, 8126–8133.

[161] R. V. Honorato, J. Roel-Touris, A. M. J. J. Bonvin, *Front. Mol. Biosci.* **2019**, *6*.

[162] J. Procyk, E. Poppleton, P. Šulc, *Soft Matter* **2021**, *17*, 3586–3593.

[163] V. K. Misra, J. L. Hecht, A.-S. Yang, B. Honig, *Biophys. J.* **1998**, *75*, 2262–2273.

[164] D. T. Edwards, J. K. Faulk, M.-A. LeBlanc, T. T. Perkins, *Biophys. J.* **2017**, *113*, 2595–2600.

[165] M. Rief, *Science* **1997**, *276*, 1109–1112.

[166] F. M. Fazal, S. M. Block, *Nat. Photonics* **2011**, *5*, 318–321.

[167] L. Ju, Y. Chen, K. Li *et al.*, *Scientific Reports* **2017**, *7*.

[168] B. Liu, W. Chen, B. D. Evavold, C. Zhu, *Cell* **2014**, *157*, 357–368.

[169] Y. F. Dufrêne, E. Evans, A. Engel *et al.*, *Nat. Methods* **2011**, *8*, 123–127.

[170] M. Soltani, J. Lin, R. A. Forties *et al.*, *Nat. Nanotechnol.* **2014**, *9*, 448–452.

[171] G. Sitters, D. Kamsma, G. Thalhammer *et al.*, *Nat. Methods* **2014**, *12*, 47–50.

[172] D. Yang, A. Ward, K. Halvorsen, W. P. Wong, *Nat. Commun.* **2016**, *7*.

[173] B. Park, S. Lee, J. Kwon *et al.*, *Scientific Reports* **2021**, *11*.

[174] H. Su, Z. Liu, Y. Liu *et al.*, *Nano Lett.* **2018**, *18*, 2630–2636.

[175] Y. Zhang, Y. Qiu, A. T. Blanchard *et al.*, *Proc. Natl. Acad. Sci.* **2017**, *115*, 325–330.

[176] C. E. Castro, H.-J. Su, A. E. Marras, L. Zhou, J. Johnson, *Nanoscale* **2015**, *7*, 5913–5921.

[177] C. K. Wong, C. Tang, J. S. Schreck, J. P. K. Doye, Characterizing the free-energy landscapes of DNA origamis, arXiv:2108.06517, **2021**.

[178] E. Poppleton, J. Bohlin, M. Matthies *et al.*, *Nucleic Acids Res.* **2020**, *48*, e72–e72.

[179] H. Chhabra, G. Mishra, Y. Cao *et al.*, *J. Chem. Theory Comput.* **2020**, *16*, 7748–7763.

[180] M. Kube, F. Kohler, E. Feigl *et al.*, *Nat. Commun.* **2020**, *11*.

[181] C. Lee, K. S. Kim, Y.-J. Kim, J. Y. Lee, D.-N. Kim, *ACS Nano* **2019**, *13*, PMID: 31291091, 8329–8336.

[182] J. Widom, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 285–327.

[183] R. D. Kornberg, Y. Lorch, *Cell* **1999**, *98*, 285–294.

[184] J. L. Workman, R. E. Kingston, *Annu. Rev. Biochem.* **1998**, *67*, 545–579.

[185] R. D. Kornberg, *Annu. Rev. Biochem.* **1977**, *46*, 931–954.

[186] G. Felsenfeld, *Nature* **1978**, *271*, 115–122.

[187] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *J. Mol. Bio.* **2002**, *319*, 1097–1113.

[188] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251–260.

[189] K. Luger, *Curr. Opin. Genet. Dev.* **2003**, *13*, 127–135.

[190] R. D. Kornberg, *Science* **1974**, *184*, 868–871.

[191] E. Y. Chua, V. K. Vogirala, O. Inian *et al.*, *Nucleic Acids Res.* **2016**, *44*, 8013–8019.

[192] W. H. Taylor, P. J. Hagerman, *J. Mol. Bio.* **1990**, *212*, 363–376.

[193] D. Shore, J. Langowski, R. L. Baldwin, *Proc. Natl. Acad. Sci.* **1981**, *78*, 4833–4837.

[194] D. M. Crothers, J. Drak, J. D. Kahn, S. D. Levene in *Methods in Enzymology*, Elsevier, **1992**, pp. 3–29.

[195] M. Vologodskaia, A. Vologodskii, *J. Mol. Bio.* **2002**, *317*, 205–213.

[196] K. Luger, T. J. Richmond, *Curr. Opin. Struct. Biol.* **1998**, *8*, 33–40.

[197] T. J. Richmond, C. A. Davey, *Nature* **2003**, *423*, 145–150.

[198] E. Segal, J. Widom, *Trends Genet.* **2009**, *25*, 335–343.

[199] J. D. McGhee, G. Felsenfeld, *Annu. Rev. Biochem.* **1980**, *49*, 1115–1156.

[200] G. E. Zentner, S. Henikoff, *Nat. Struct. Mol. Biol.* **2013**, *20*, 259–266.

[201] D. Duzdevich, S. Redding, E. C. Greene, *Chem. Rev.* **2014**, *114*, 3072–3086.

[202] I. Heller, T. P. Hoekstra, G. A. King, E. J. G. Peterman, G. J. L. Wuite, *Chem. Rev.* **2014**, *114*, 3087–3119.

[203] C. Monico, M. Capitanio, G. Belcastro, F. Vanzi, F. Pavone, *Int. J. Mol. Sci.* **2013**, *14*, 3961–3992.

[204] R. Buning, J. van Noort, *Biochimie* **2010**, *92*, 1729–1740.

[205] J. S. Choy, T.-H. Lee, *Trends Biochem. Sci.* **2012**, *37*, 425–435.

[206] J. L. Killian, M. Li, M. Y. Sheinin, M. D. Wang, *Curr. Opin. Struct. Biol.* **2012**, *22*, 80–87.

[207] C. Lavelle, J.-M. Victor, J. Zlatanova, *Int. J. Mol. Sci.* **2010**, *11*, 1557–1579.

[208] S. J. Petesch, J. T. Lis, *Trends Genet.* **2012**, *28*, 285–294.

[209] J. Zlatanova, T. C. Bishop, J.-M. Victor, V. Jackson, K. van Holde, *Structure* **2009**, *17*, 160–171.

[210] G. Li, M. Levitus, C. Bustamante, J. Widom, *Nat. Struct. Mol. Biol.* **2004**, *12*, 46–53.

[211] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, C. Bustamante, *Science* **2009**, *325*, 626–628.

[212] L. Bintu, T. Ishibashi, M. Dangkulwanich *et al.*, *Cell* **2012**, *151*, 738–749.

[213] S. Mihardja, A. J. Spakowitz, Y. Zhang, C. Bustamante, *Proc. Natl. Acad. Sci.* **2006**, *103*, 15871–15876.

[214] P. C. Nickels, B. Wünsch, P. Holzmeister *et al.*, *Science* **2016**, *354*, 305–307.

[215] M. C. Engel, F. Romano, A. A. Louis, J. P. K. Doye, *J. Chem. Theory Comput.* **2020**, *16*, 7764–7775.

[216] D. W. Sumners, *Prog. Theor. Phys.* **2011**, *191*, 1–19.

[217] C. B. Anfinsen, *Biochem. J.* **1972**, *128*, 737–749.

[218] K. E. Drexler, *Proc. Natl. Acad. Sci.* **1981**, *78*, 5275–5278.

[219] R. M. Dirks, *Nucleic Acids Res.* **2004**, *32*, 1392–1403.

[220] R. S. Forgan, J.-P. Sauvage, J. F. Stoddart, *Chem. Rev.* **2011**, *111*, 5434–5464.

[221] J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic, A. Stasiak, *Proc. Natl. Acad. Sci.* **2012**, *109*, E1715–E1723.

[222] J. W. Schmidberger, J. A. Wilce, A. J. Weightman, J. C. Whisstock, M. C. Wilce, *J. Mol. Bio.* **2008**, *378*, 284–294.

[223] D. Bölinger, J. I. Sułkowska, H.-P. Hsu *et al.*, *PLoS Comput. Biol.* **2010**, *6*, e1000731.

[224] T. C. Sayre, T. M. Lee, N. P. King, T. O. Yeates, *Protein Eng. Des. Sel.* **2011**, *24*, 627–630.

[225] J. I. Sulkowska, P. Sulkowski, P. Szymczak, M. Cieplak, *Proc. Natl. Acad. Sci.* **2008**, *105*, 19714–19719.

[226] N. P. King, A. W. Jacobitz, M. R. Sawaya, L. Goldschmidt, T. O. Yeates, *Proc. Natl. Acad. Sci.* **2010**, *107*, 20732–20737.

[227] J. Arsuaga, M. Vazquez, P. McGuirk *et al.*, *Proc. Natl. Acad. Sci.* **2005**, *102*, 9165–9169.

[228] M. Wong, oxDNA UTILS documentation, **2019**.

[229] W. Kabsch, *Acta Crystallogr. A* **1976**, *32*, 922–923.