PhD Thesis

University of Sheffield

Investigating the role of supercoiling in DNA damage detection using all-atom implicit solvent molecular dynamics simulations

A Thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy in the School of Mathematical and Physical Sciences by

Victoria E. Hill

Department of Chemistry

July 2025

Abstract

DNA damage is not uniformly distributed across the genome, and crystal structures show that repair proteins interact with flipped-out bases or bent DNA. When supercoiled, DNA often forms plectonemes: plait-like structures of interwound superhelices with sharply bent apices. To accommodate the bend, flipped bases, kinks, and bubbles form. Because plectonemes protrude from the dense genome, it has been hypothesised that damage may localise to the plectoneme tips to improve accessibility to repair proteins and provide bent or flipped structures for binding. Here, we use DNA minicircles to study the interplay between DNA sequence and structure in positioning plectonemes under both positive and negative supercoiling. We present all-atom, implicit-solvent molecular dynamics simulations of minicircles of two sizes and multiple sequences. The simulations' ability to capture sequence-dependent effects was validated experimentally using a novel smFRET assay and supported by data from the literature. Our results confirm implicit solvent molecular dynamics to be an accurate and accessible method to study the effects driving plectoneme pinning. We introduced all twelve mismatches into sequences with different plectoneme density landscapes, enabling us to examine both damage-driven pinning and the competition between sequence and damage for the tip. We found that the mismatches had different propensities for pinning the plectoneme, which further depended on the surrounding sequence context and direction of supercoiling. Multiple consecutive mismatches were introduced to study the relative influences of sequence versus bubble formation and elucidate a novel sequence-dependent threshold below which the bubble sequence dominates over bubble size in positioning the plectoneme. Overall, the conformational response of damaged DNA under torsional strain may explain why some regions of the genome are repaired more efficiently than others and provide one solution to the protein-DNA search problem.

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Work presented in this thesis has formed part of the following publication, as indicated at the beginning of the corresponding chapter:

- **Victoria E. Hill**, Sophie E. Fountain, Quentin M. Smith, David S. Rueda, Agnes Noy, and Timothy D. Craggs. "Implicit Solvent Simulations Capture Sequence-Dependent Plectoneme Pinning". 2025. *Manuscript submitted for publication.*

Date 21st July 2025 .....

Victoria E. Hill

To Jade, without whom this would not have been possible...

All joking aside, first and foremost, my thanks must go to JADE2, funded by EPSRC, and the allocations I received via HecBioSim and the University of Sheffield. I've estimated that, throughout my PhD, I have used approximately 50 years of GPU time on JADE2. Had it not been for their generous allocations and the suspension of billing, this work would not have been possible, and the project certainly wouldn't have grown into what it ultimately became. My thanks also go to the BBSRC White Rose DTP for funding both myself and this research, as well as to STFC and the DiRAC Innovation scheme for funding my placement which, ultimately, led me to becoming a better researcher.

I would also like to thank the many people who have contributed—whether through their scientific skills and knowledge, or through their friendship and emotional support—throughout my PhD, but especially Agnes Noy, Sophie Fountain, Mahmoud Abdelhamid, Anyah Settle, and Simon Fairbanks. I must also thank the numerous academics and collaborators who supported my work and engaged with me in countless discussions about the wonders of DNA sequence and topology. Although there are too many to name, they played a significant role in shaping this research.

Finally, my greatest thanks go to my family, whose unwavering support and encouragement throughout this PhD, and the many years leading up to it, have allowed me to pursue a project that I am truly passionate about. I promise you will never have to hear the word "minicircles" again. And to Luke—his support and patience during what was probably the most stressful period of my life has been invaluable. Writing this thesis would have been infinitely more difficult without his comfort and understanding, and the company of my favourite writing buddy, Aine.

Contents

List of Important Abbreviations

AV Accessible Volume

BER Base Excision Repair

bp base pair

BVA Burst Variance Analysis

DNA DeoxyriboNucleic Acid

DNAP DNA Polymerase

dsDNA double-stranded DNA

GB Generalised Born

ISD Intercalation-induced Supercoiling of DNA

Lk Linking number

MD Molecular Dynamics

MMR [DNA] MisMatch Repair

mRNA messenger RNA

NER Nucleotide Excision Repair

nt nucleotides

RER Riboncleotide Excision Repair

RNA RiboNucleic Acid

RNAP RNA Polymerase

SIDD Stress-Induced Duplex Destabilization

smFRET single-molecule Förster resonance energy transfer

SSB Single-Strand Breaks

ssDNA single-stranded DNA

tRNA transfer RNA

WLC Worm-Like Chain

List of Figures

List of Tables

This thesis describes a series of simulations that investigate the mechanical properties and sequence-dependent conformational dynamics of DNA under superhelical tension. **Chapter 1** provides the biological background for this work. It begins with the broader structure and biological functions of DNA before focusing specifically on DNA supercoiling. The mechanisms of DNA damage recognition and repair are introduced to frame the protein–DNA search problem. We then collate the existing evidence suggesting that DNA is not merely the passive partner in biological processes that it was once thought to be, but actively modulates its own interactions via a sequence–structure–function relationship. Finally, we review existing approaches for studying sequence-dependent DNA properties and DNA supercoiling, highlighting the need for multiscale investigations that combine both theoretical and experimental methods. **Chapter 2** describes, in detail, the methodology and underpinning theoretical background used in this work. This includes an overview of the AMBER force field, the principles of molecular dynamics simulations, and descriptions of both explicit and implicit solvent models. Key simulation parameters—such as the use of thermostats, barostats, and restraints—are outlined. This is followed by a description of single-molecule FRET, as well as the two predictive algorithms central to this thesis: the SIDD algorithm for bubble location prediction, and the Dekker algorithm for predicting plectoneme location.

Chapter 3 presents a large dataset of implicit solvent simulations of supercoiled DNA minicircles, which is used to investigate several properties relevant to DNA supercoiling. Four sequences are compared against results from the literature, while three sequences were specifically designed to exhibit distinct plectoneme and bubble conformational landscapes. These were used to test the ability of molecular dynamics simulations to reproduce both experimental observations and theoretical predictions. We further explore the role of bubble formation in driving plectoneme pinning and its interplay with other structural factors such as intrinsic curvature. To do so, twist changes are induced by modifying simulation conditions. The results are validated against a novel smFRET assay.

Chapter 4 extends this work by investigating supercoiled DNA minicircles containing mismatches. The plectoneme pinning effects of 12 different mismatches are explored in the two distinct sequence contexts from the previous chapter. We also examine the influence of multiple mismatches and the modulation of bubble properties through changes in twist. The competition between DNA sequence and damage for the plectoneme tip is studied, and the role of bubble formation as a driving force of plectoneme positioning is further evaluated. Together, this forms the most comprehensive investigation of mismatch-driven plectoneme pinning to date.

Finally, **Chapter 5** synthesizes the results of the previous chapters. It summarizes the key findings, evaluates the overall success of the project, and discusses the broader biological implications. We discuss how our results provide novel insights into bubble-driven plectoneme pinning and propose future research directions to extend and refine this work.

1

Introduction

The double-helical structure of deoxyribonucleic acid (DNA) was famously published in 1953 by Watson and Crick [1] thanks to the X-ray diffraction work of Rosalind Franklin and Raymond Gosling [2]. Whilst this original model thought the structure to be rigid, we now know that this is not the case – DNA is highly dynamic. The large structure allows for many degrees of freedom that can cause motions on timescales ranging from nanoseconds to milliseconds. This range of dynamics and the resulting conformations are crucial and specific to each of its biological functions. Historically, there has been a protein-centric view of cellular processes such as transcription and replication; however, the field now leans towards the idea that DNA is not a passive partner in these interactions. Studying the conformational dynamics of DNA and understanding the principles and mechanisms underlying the conformational shifts tells us not just about DNA itself, but also about the broader molecular processes that govern cellular life. DNA's physical properties and functional versatility are intricately linked, revealing the complex relationship between its sequence structure and its roles in cellular processes.

1.1 DNA

DNA is a polymer comprised of four base units: cytosine (C), guanine (G), adenine (A), and thymine (T). The nitrogenous bases are further categorised into the purines (A and G) with a double ring structure and the pyrimidines (C and T) with a single six-membered ring (fig. 1.1).

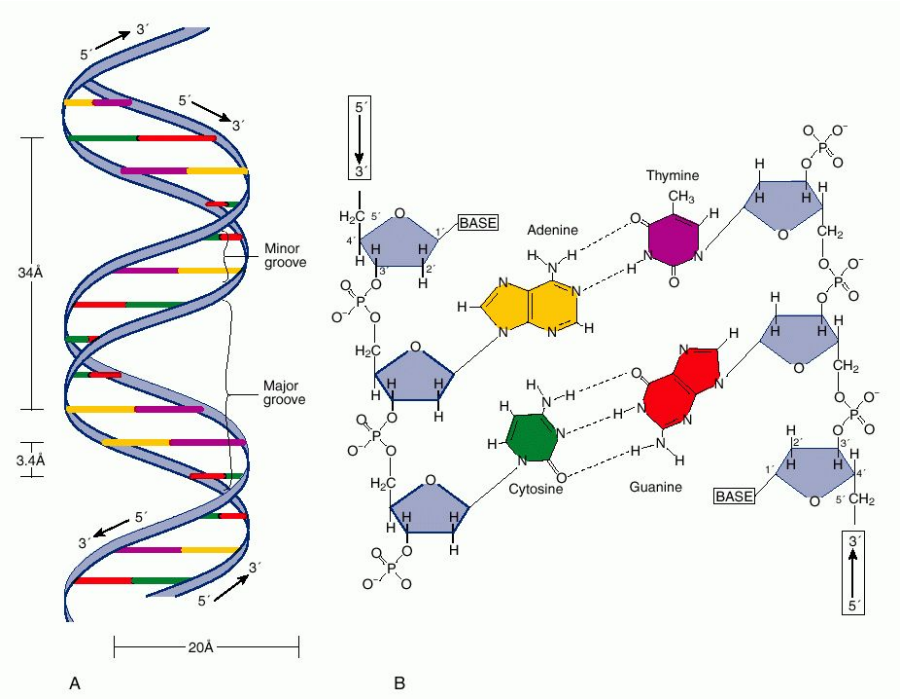

Figure 1.1: The structure of B-DNA. Figure taken from the 3DNA software package [4].

The bases are joined to one another along the strand via a sugar-phosphate backbone that links the sugar of one nucleotide to the phosphate of the next nucleotide in a fashion known as phosphodiester linkage. The bases, phosphate group, and deoxyribose sugar together are known as a nucleotide. They form two antiparallel polynucleotide strands that coil around one another to form the double helix. The double-stranded DNA forms such that the bases between the two strands hydrogen bond with one another to create a base pairing. Here, A forms a pair with T and C forms a pair with G. The total GC content of the human genome is 40.9% [3]. Due to the asymmetric directionality of the DNA strands, the strands run 3' (pronounced three prime) to 5' (pronounced five prime) on one strand, with the 3' a terminal hydroxyl group and the 5' a terminal phosphate group. The other strand runs in the 5' to 3' direction with the same corresponding termini. In addition to the hydrogen bonding afforded by the base pairing, the hydrophobic interactions as a result of base stacking between the adjacent bases on the same strand further stabilises the double helix (fig. 1.1).

Of the three main DNA structures found in nature, A-DNA, B-DNA, and Z-DNA, the B form predominates. In this structure, there are around 10.5 base pairs (bp) per turn in the helix which equates to a right-handed rotation of 34.3° per base pair [5–7]. The regularity of the double helical structure creates repeating structural features known as the major and minor grooves. The major and minor grooves are 22 Å and 12 Å in width, respectively. The difference arises due to the strands not being directly opposite each other. The offset pairing of the two strands is described by the opening between the base pairs (fig. 1.2). However, in reality, the overall conformation of DNA may deviate from this canonical form as it is strongly linked to its sequence and role. For example, repeating A · T units at least 4 base pairs in length, have a

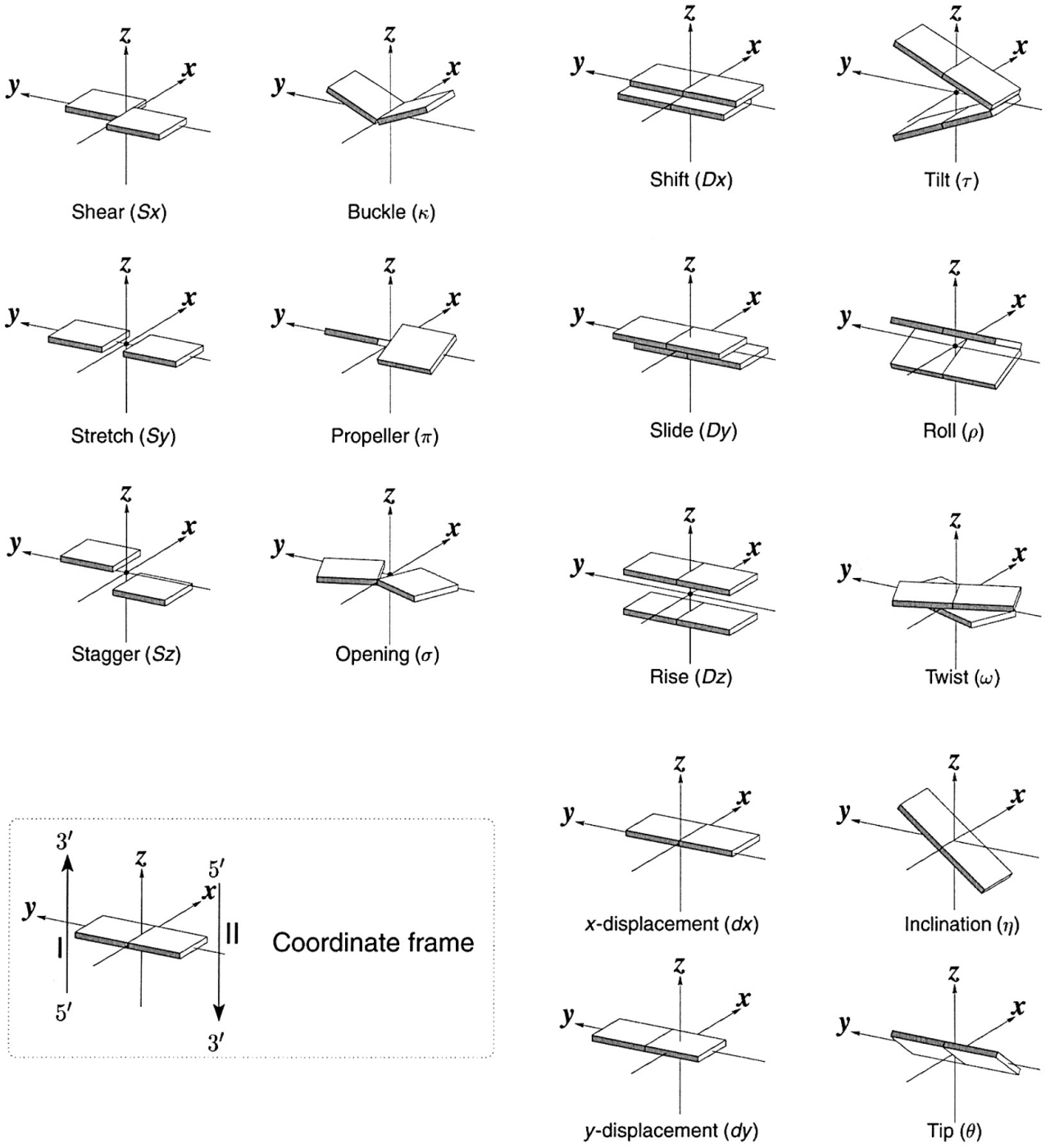

Figure 1.2: Base pair and base step parameters. The shaded edge facing the viewer is the minor-groove side. The columns from left to right correspond to the translational base pair parameters, rotational base pair parameters, translational base step parameters, and rotational base step parameters. Figure taken the 3DNA software package [9].

narrower minor groove due to exhibiting intrinsic curvature [8].

Base-pair and Base-step parameters

Describing the geometry of a base pair, i.e. the two bases opposite one another across the two strands; and a base step, i.e. the relationship to the following base pair, is crucial to describing the full higher structure. In 1988, a workshop was held to standardise the definition

and nomenclature of nucleic acid structure components. The resulting Cambridge University Engineering Department Helix Computation scheme (CEHS) (fig. 1.2) brought nomenclature in line with the recommended practice from the International Union of Pure and Applied Chemistry (IUPAC) [10, 11]—the world authority on chemical nomenclature, terminology, and measurement standards. The result was six base pair parameters that are split into three translational parameters: shear, stretch, and stagger; and three rotational parameters: buckle, propeller twist, and opening. There are a further six base step parameters, again with three translations: shift, slide, and rise; and three rotations: tilt, roll, and twist. The axes with which these are calculated are positioned such so that the same value will be obtained when calculating in both the 3' to 5' and 5' to 3' directions, with only a possible change of sign. The guidelines, such as the direction and positioning of the x, y, and z axes are known as the Cambridge accord.

Following this, there arose a need for standard base pair co-ordinates with which comparisons could be made in order to investigate deviations away from normality. The Tsukuba convention [12] defines the standardised base pairs based on high-resolution X-ray crystal structures following optimisation of the hydrogen-bond donor-acceptor distances. This takes advantage of the planarity of the DNA and so, in the optimal models, the co-ordinates are defined such that the base pair and base step parameters have reference values of zero. Parameter measurements, therefore, are relative deviations from the ideal structure as in reality, DNA does not exist as described in the ideal model (table 1.3).

The base pair and base step parameters differ between the pairs themselves, and also depend on the directionality of the pair, i.e. GC differs from CG (table 1.4) [14, 15]. The twist, roll, and slide parameters show the greatest variability across all base steps, with twist spread over 10° and roll spread over 14° compared to tilt which has a range of only 3°. Of the translation parameters, the shift values stay within an 0.8 Å span compared to the 2.1 Å range of the slide values [15]. Note that the twist between base pairs must be present due to the helical structure of DNA. The differences in twist across base steps can be explained by considering steric clashes in the major and minor grooves [15]. Where there are large exocyclic groups in the grooves, and, therefore, the largest steric clashes, there are low twist values to accommodate the clash. This is the case for CG with NH_2—NH_2 clashes in both the major and minor groove; and for AG that has an NH_2—NH_2 clash in the major groove and an NH_2—O clash in the minor groove.

Table 1.3: *The base pair parameters from X-ray crystal structures as used in the Structure and Conformation of Helical Nucleic Acids: Rebuilding Program (SCHNArP) [13]. The angle parameters are given in degrees, and the slide parameter in* Å.

Base Pair	Propeller	Opening	Buckle	Stretch	Stagger	Shear
Flat base pair	0.00	0.00	0.00	5.45	0.00	0.00
A-T	-15.52	7.88	-0.79	5.39	-0.055	0.11
G-C	-9.18	2.82	-2.08	5.48	0.02	0.00

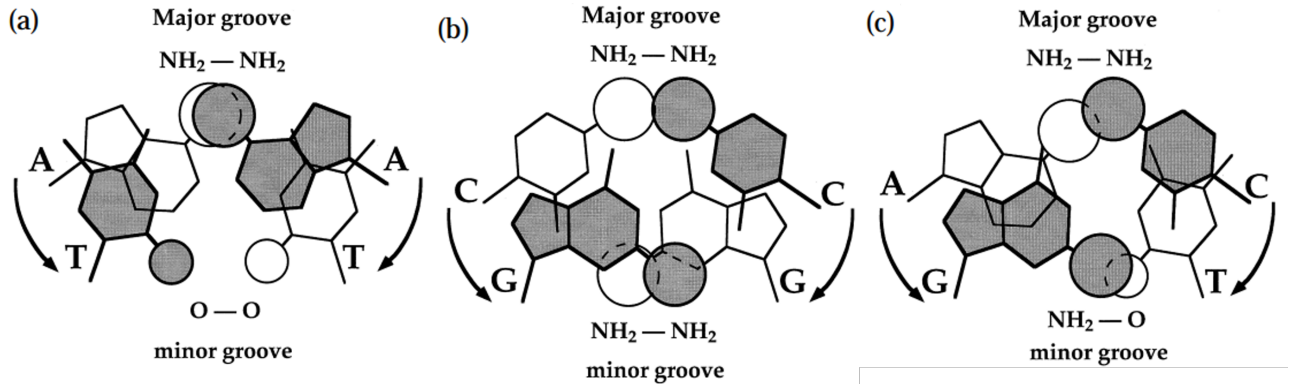

Figure 1.5: A projection of the (a) AT; (b) CG; (c) AG base steps showing the exocyclic groups involved in the steric clashes of the major and minor grooves. The shaded base pairs are closest to the viewer. Image taken from Gorin *et al.* [15].

This contrasts, for example, TA which has only an O—O clash in the major groove and a H—H clash in the minor groove and a large twist value (fig. 1.5). In protein-DNA complexes, the deformations follow the same trends; however, the dispersion of the parameters is larger [16].

Whilst the absolute values of the base pair and step parameters well-describe the conformation of the DNA, the intrinsic correlations between certain parameters further explain the conformational dynamics of the DNA. For example, the twist of a base step is strongly correlated to the slide, roll, cup, and rise parameters [15]. Cup is defined as the difference in buckle between the two base pairs within the step. This can be explained by the rigidity of the sugar-phosphate backbone. The C—C distance along the chain varies by 0.3 Å, which is only 6% of the 4.9

Table 1.4: *The average of each base step parameter over N base steps in 38 crystal structures of B-DNA without bound proteins or small molecules, or chemical modifications. The angle parameters are given in degrees, and the slide parameter in* Å*. The signs are as defined by the Cambridge accord [10]. The standard deviation is quoted in parenthesis. Table reproduced from Gorin et al. [15].*

Step	N	Twist	Tilt	Roll	Slide	Propeller
GC	27	$38.3_{(3.8)}$	$0.0_{(4.6)}$	$-7.0_{(6.1)}$	$0.29_{(0.37)}$	$-11.7_{(4.2)}$
AC	10	$35.8_{(4.4)}$	$-0.9_{(3.6)}$	$0.4_{(6.7)}$	$-0.13_{(0.54)}$	$-6.7_{(8.3)}$
AT	34	$33.4_{(3.5)}$	$0.0_{(2.1)}$	$-0.6_{(3.5)}$	$-0.37_{(0.23)}$	$-16.6_{(5.8)}$
TA	9	$40.0_{(3.8)}$	$0.0_{(3.7)}$	$2.6_{(6.4)}$	$0.74_{(0.87)}$	$-11.1_{(3.1)}$
CA	23	$36.9_{(9.6)}$	$0.6_{(3.1)}$	$1.1_{(6.3)}$	$1.46_{(1.04)}$	$-8.6_{(4.3)}$
CG	10	$31.1_{(4.7)}$	$0.0_{(4.3)}$	$6.6_{(3.2)}$	$0.63_{(0.29)}$	$-11.2_{(6.1)}$
GA	25	$39.3_{(3.3)}$	$-0.4_{(2.7)}$	$-0.1_{(3.0)}$	$-0.07_{(0.34)}$	$-15.1_{(4.7)}$
AA	50	$35.8_{(3.1)}$	$-0.4_{(4.2)}$	$0.5_{(3.2)}$	$-0.03_{(0.29)}$	$-17.3_{(7.0)}$
GG	3	$33.4_{(3.3)}$	$-1.1_{(3.9)}$	$6.5_{(2.9)}$	$0.60_{(0.05)}$	$-12.8_{(4.3)}$
AG	4	$30.5_{(4.9)}$	$-2.6_{(1.3)}$	$2.9_{(6.9)}$	$0.47_{(0.68)}$	$-14.3_{(10.4)}$
Total	195	$36.1_{(5.2)}$	$0.0_{(3.6)}$	$-0.2_{(5.7)}$	$0.21_{(0.74)}$	$-13.9_{(6.7)}$

Å equilibrium distance. The twist parameter directly modulates this value [17]. In order to maintain this distance, an increase in twist is often accompanied by a compensatory decrease in cup and roll, and increase in slide and rise [18, 19]. When rise increases, there is a decrease in buckle, and when the tilt increases, there is a compensatory decrease in stagger [12]. The most pronounced structural perturbations occur when a base pair parameter is increased at one nucleotide, coupled with the same parameter decreasing at the successive nucleotide. For example, a negative buckle at base i accompanied by a positive buckle at base $i+1$. These induce large differences, i.e. large values of $\Delta Buckle$, that create extreme dimer steps. In general, there is a nearly instantaneous coupling between the movement of two neighbouring base pairs [20]. However, how base pair parameters are coupled to one another and coupled along successive base steps describes the overall flexibility of the DNA [21].

The dependence of base pair and base step parameters on sequence was the fundamental argument that founded the theory that DNA sequence has an effect on structure. Since DNA's role as a carrier of genetic information is intrinsically tied to its sequence, this supports the theory that both sequence and structure work synergistically to enable DNA to carry out its various functions.

The Functions of DNA

DNA exists as chromosomes that contain all of our genetic material. Its primary role is to store the genetic information that is used to produce proteins in a process known as the central dogma of molecular biology [22]. The central dogma was first described by Francis Crick in September 1957 in a lecture that even now frames how we understand the central processes surrounding life [23]. He stated that "once information has got into a protein it can't get out again" (fig. 1.6). Here, he recognises six valid transfers of information: DNA → DNA (DNA replication); DNA → RNA and RNA → protein which together are the two steps of protein synthesis; RNA → RNA (RNA replication in viruses); a potential for RNA → DNA (hence the dotted line) which we now know to be reverse transcription; and a potential for DNA → protein omitting the role of RNA altogether. Crucially, the information cannot be transferred from protein back to nucleic acid, meaning that once the genetic information is held in the protein, the protein cannot alter DNA sequence.

In the human body, 3 billion base pairs are arranged into 46 chromosomes that make up the human genome [25]. Chromosomes contain genes: segments of DNA that code for a specific protein. There are approximately 20,000 protein-coding genes which account for 40% of the genome; however, a gene is made up of both coding regions called exons and non-coding introns [26]. This results in only 1.15% of the total genome being made of coding DNA. Upstream from the coding region is the promoter region where RNA polymerase (RNAP) and regulatory proteins called transcription factors bind to initiate transcription (fig. 1.7). Once bound, the DNA is unwound to create the template and coding strands and form the transcription bubble.

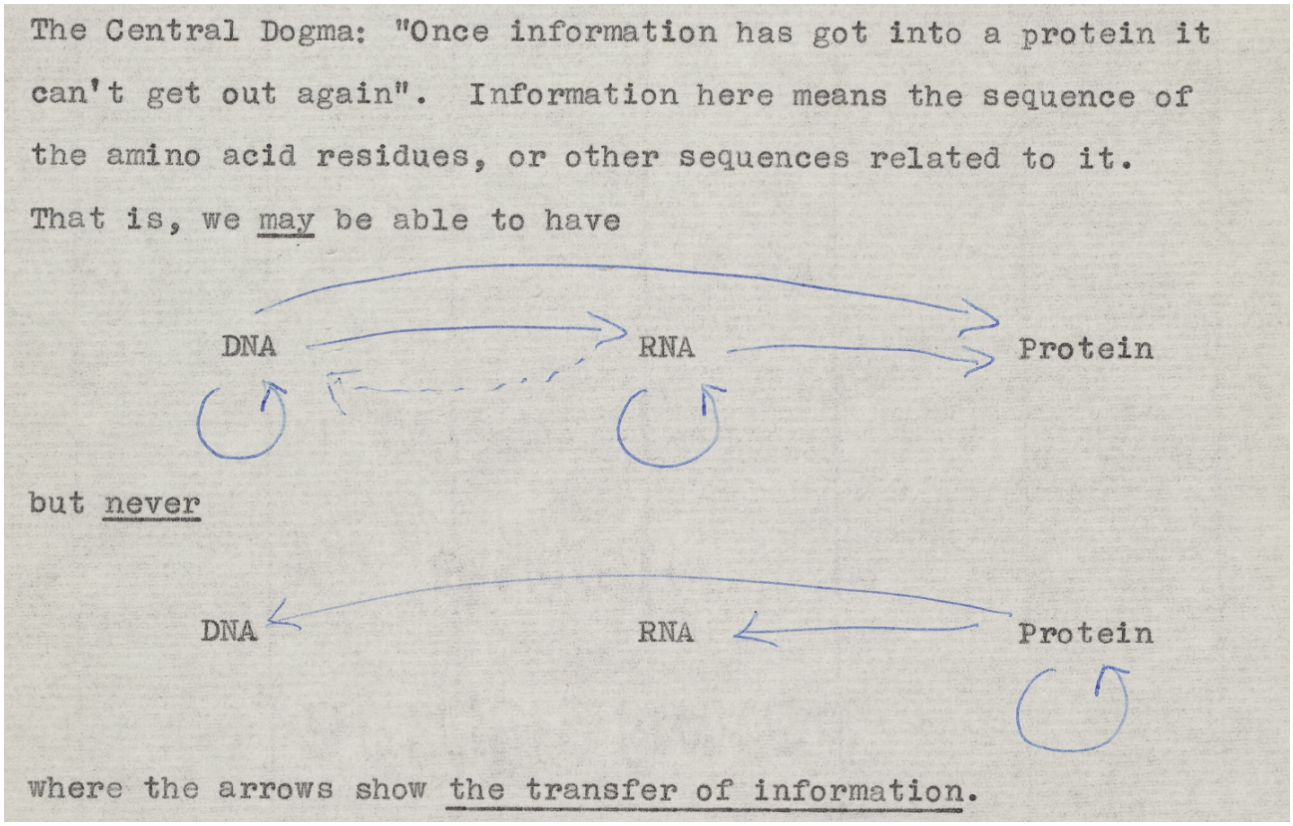

Figure 1.6: A section of the unpublished note by Francis Crick entitled "Ideas on Protein Synthesis (Oct. 1956)". The note is the first hypothesis of the central dogma that was later published in 1958 [22]. It reads "The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it." Image taken from the Wellcome Collection [24].

It creates a strand of RNA complementary to the template strand known as messenger RNA, or mRNA. The mRNA is the RNA equivalent of the coding strand, with a ribose sugar instead of a deoxyribose, and uracil (U) in place of thymine. RNAP transcribes the coding region until it reaches a termination sequence. This is the DNA $\rightarrow$ RNA transfer of genetic information.

Subsequently, the RNA $\rightarrow$ protein transfer occurs in a process known as translation (fig. 1.7). Here, the mRNA is decoded in the ribosome by reading three consecutive bases known as codons. This usually begins with an AUG start codon or initiator methionine codon. Each codon codes for a specific amino acid. Transfer RNA (tRNA) carrying the appropriate amino acid chain match with each codon using the complementary anti-codon to create the peptide chain. Translation ends upon reaching a UAA, UAG, or UGA stop codon.

The central dogma also encompasses DNA replication, a key process in cell division during cell growth, repair, and biological inheritance that produces an identical copy of the DNA molecule. As with transcription, the strands of DNA are separated; however, during replication, both strands now act as template strands. The point at which the two strands become unwound is known as the replication fork and replication occurs outwards in both directions from the origin to create the replication bubble (fig. 1.8). DNA helicase is the enzyme responsible for the breaking of the hydrogen bonding and removing any bound proteins [29–32]. The unwinding

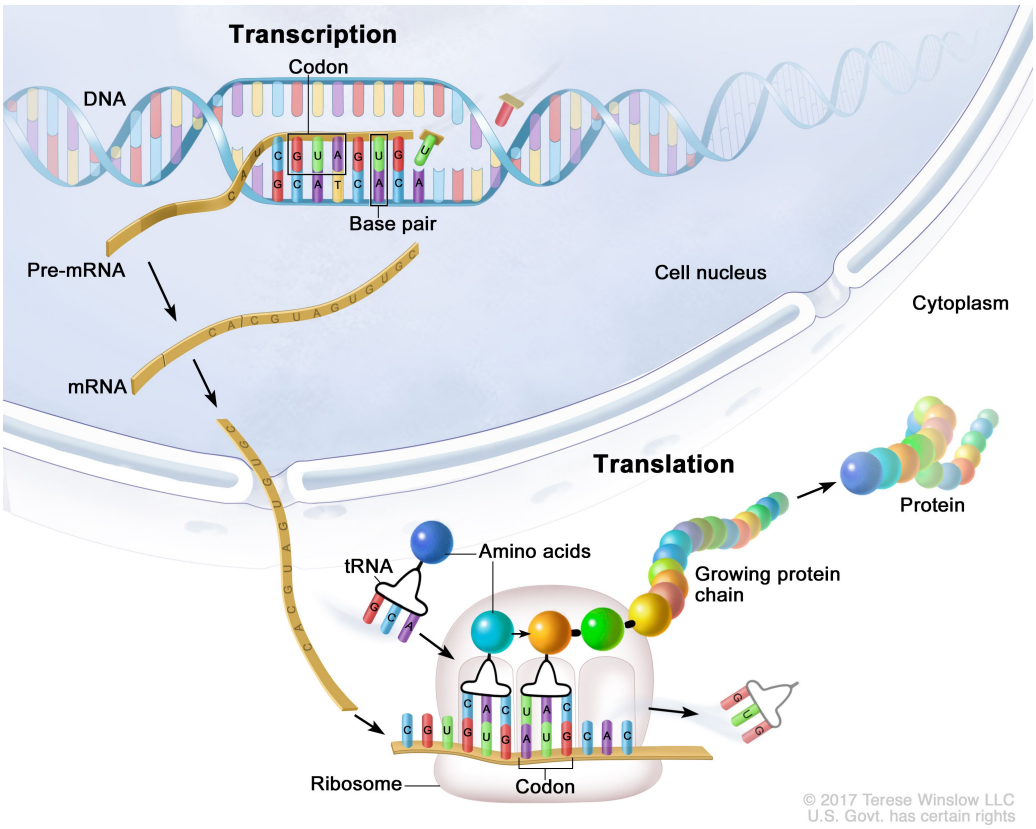

Figure 1.7: Transcription is the process of forming mRNA from DNA in the first step of protein synthesis. mRNA is then translated into a protein in the cytoplasm by tRNA. Image taken from the National Cancer Institute [27].

kinetics are highly sequence-dependent, depending both upon the sequence passing through the helicase and on the resulting single-stranded(ss)DNA due to their inchworm-like motion [33]. Replication begins at AT-rich sequences that are known as the origin of replication and are recognised by initiator proteins. The combination of DNA, initiator proteins, and DNA polymerase is known as the preinitiation complex. The role of DNA polymerase is to read the DNA in the 3' to 5' direction and then synthesise the new DNA strand in the 5' to 3' direction. This occurs for both strands simultaneously in the replisome, the complex molecular machine that encompasses all biomolecules involved in replication. Due to replication occurring in the 3' to 5' direction on both strands, and the antiparallel nature of the two strands, this leads to the two strands being replicated in opposite directions relative to the replication fork. The strand being continuously synthesised in the direction of the growing replication fork is the leading strand. In the other direction, the lagging strand requires a more complex process. Here, DNA primase initiates the synthesis of a short RNA primer complementary to the lagging strand that is extended by DNA polymerase. The resulting Okazaki fragments are joined together by DNA ligase. Replication occurs until it reaches a termination site. However, due to the linear nature of eukaryotic DNA, the replication machinery is unable to reach the ends of strands and so DNA is lost from the end of the chromosome with each replication cycle. Telomeres are regions of non-coding repetitive DNA at the ends that exist to prevent the loss of genes as a result of this. Telomere shortening refers to the gradual loss of these protective ends and is a natural

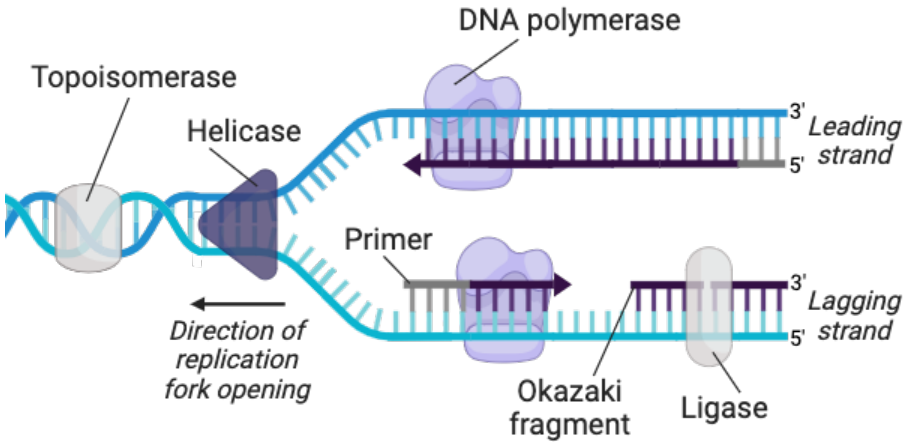

Figure 1.8: DNA replication is the process of synthesising an exact copy of the exiting DNA molecule. Image taken from Khan Academy [28].

hallmark of aging.

1.2 DNA Supercoiling

DNA topology refers to the three-dimensional shape of DNA and describes how the two strands are intertwined with respect to one another. It is described by the linking number, Lk, and encompasses DNA supercoiling, knotting, and catenation. Supercoiling is the under- or over-winding of the DNA double helix, compared to B-DNA, that creates torsional stress. Positive supercoiling occurs when DNA is over-wound, whilst negative supercoiling results from under-twisting. These create torsional strain that causes the DNA to writhe in order to relieve that strain. Writhe is the degree of coiling of the helix—literally a sum of the crossing points in the resulting supercoiled structure. The linking number is the sum of twist (number of turns in the helix) and writhe and the global conformation is governed by the twist/writhe partitioning (fig. 1.9) [34]. In order for there to be a build-up of torsional stress, the DNA must be constrained; therefore, this phenomenon is exclusive to DNA with closed or fixed ends; for example, the DNA loops found in nature.

For a given linking number, the twist and writhe are interconvertible and can fluctuate to afford a wide range of conformations (fig. 1.10); however, the sum of the two must remain constant. In order to change the level of supercoiling, and therefore the linking number, nicks or breaks in the DNA are required to break the phosphodiester backbone and release the torsional strain. To normalize the linking number to DNA of any length, the superhelical density, σ, is defined by $\Delta Lk/Lk_0$ where Lk_0 is the linking number, or number of helical turns, of the relaxed conformation. There are 10.5 base pairs per turn of the helix, and so Lk_0 can be calculated by dividing the length of the DNA in base pairs by 10.5. The result of the torsional stress is either toroidal coiling, which is ubiquitous in DNA condensation [35], or the more prevalent formation of the lower energy plectonemic structures [36]. In prokaryotes, DNA is predominantly maintained in a negatively supercoiled state with a superhelical density of -0.06 [37]. Conversely,

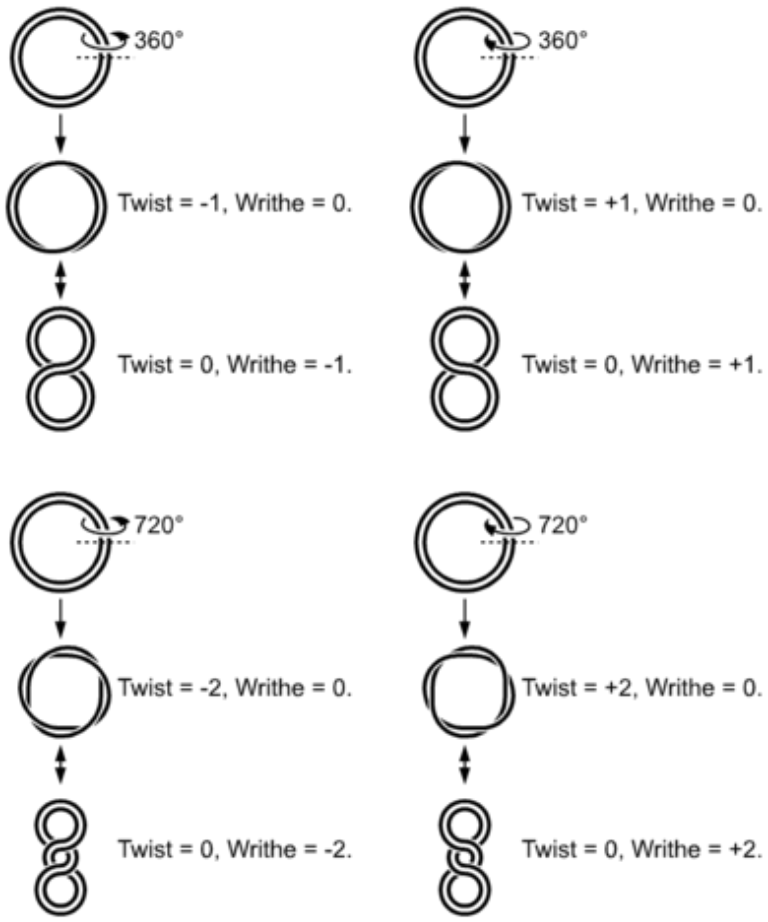

Figure 1.9: The twist and writhe of the helix can be either positive or negative depending on its direction. The twist is the twisting of the duplex itself, whilst writhe signifies a coiling of the global conformation. Figure taken from Wheeler [34].

in thermophilic bacteria, it is maintained in a positive supercoil in order to prevent thermally induced denaturation in their high temperature environment [38, 39]. Supercoiling affects a number of biological functions involving DNA transcription, recombination, and replication.

Plectonemes

A plectoneme is the crossing of DNA strands over one-another to form a plait-like structure of interwound superhelices (fig. 1.10). They are twisted in such a way that they cannot be separated without breaking them. They form as a result of writhe—the rotation of the duplex around the superhelical axis. The most simple conformation resulting from this is a figure eight. The two lobes of the figure eight are rotated with respect to one another, clockwise in the case of negative supercoiling, and anticlockwise under positive supercoiling. For each increase in writhe, the lobes will undergo an additional rotation about their axis to sample different conformations. The point at which the distant sections of the DNA meet is known as the crossing point. Upon writhing, the DNA is highly dynamic and may adopt a wide range of conformations [40] dependent on the positioning of the crossing point, the twist/writhe

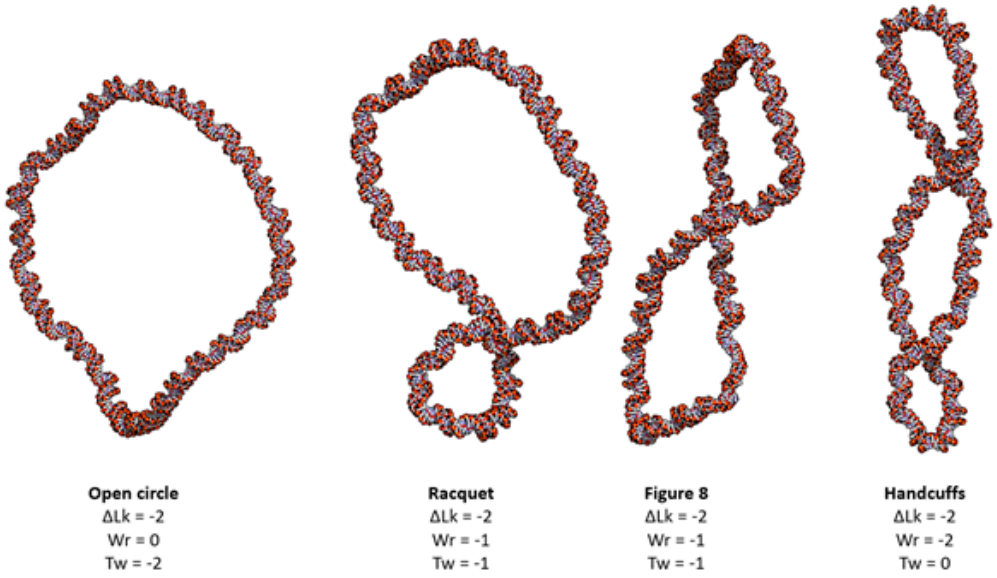

Figure 1.10: Depending on the writhe/twist partitioning, supercoiled DNA can adopt different conformations. This is shown on a model circular system. When the torsional stress is stored completely as twist, the DNA adopts an open conformation. Upon writhing, the resulting plectonemes are highly dynamic and can adopt a wide variety of structures. Figure contains snapshots from molecular dynamics simulations of a 339 bp minicircle (see chapter 3).

partitioning, and bend angle at the plectoneme tip. The tip of a plectoneme refers to the apex of the structure; in the case of a minicircle, two tips exist at the distal ends of the DNA. These tips are sharply bent with a bend angle of up to 240° seen across the plectoneme tip arc [41]. Pinning the plectoneme is the phenomenon of specific regions of DNA consistently localising at the plectoneme tip, which can influence its stability and dynamic properties. This can also be achieved through protein binding [42].

Biological Role

The packing of dozens of centimetres of DNA into the micrometre-sized nucleus is a vital process known as DNA condensation (fig. 1.11). In eukaryotic DNA, the double helix is wrapped around an octamer of four highly basic proteins called histones that together form a nucleosome [44]. The wrapping results in toroidal supercoiling which aids packing of the DNA into an organised structure [45]. The wrapping has been shown to be ambidextrous as left-handed nucleosomes form on negatively supercoiled DNA, whereas right-handed nucleosomes form on positively supercoiled DNA [46]. Linker DNA between neighbouring nucleosomes creates the 10 nm "bead on a string" model; however, when linker histones bind in this region, the DNA is further compacted into a 30 nm fibre. This can be compacted further still to form a chromosome. Chromatin refers to the complex of DNA and histones in any of the three levels of compactions. Humans have 23 pairs of chromosomes containing between approximately 45 and 250 million base pairs, compacted into structures on the order of micro-meters. However, the level of compaction at any given point depends on the current phase of the cell cycle. During interphase, when the cell is actively growing and replicating, the chromatin remains structurally "loose" to

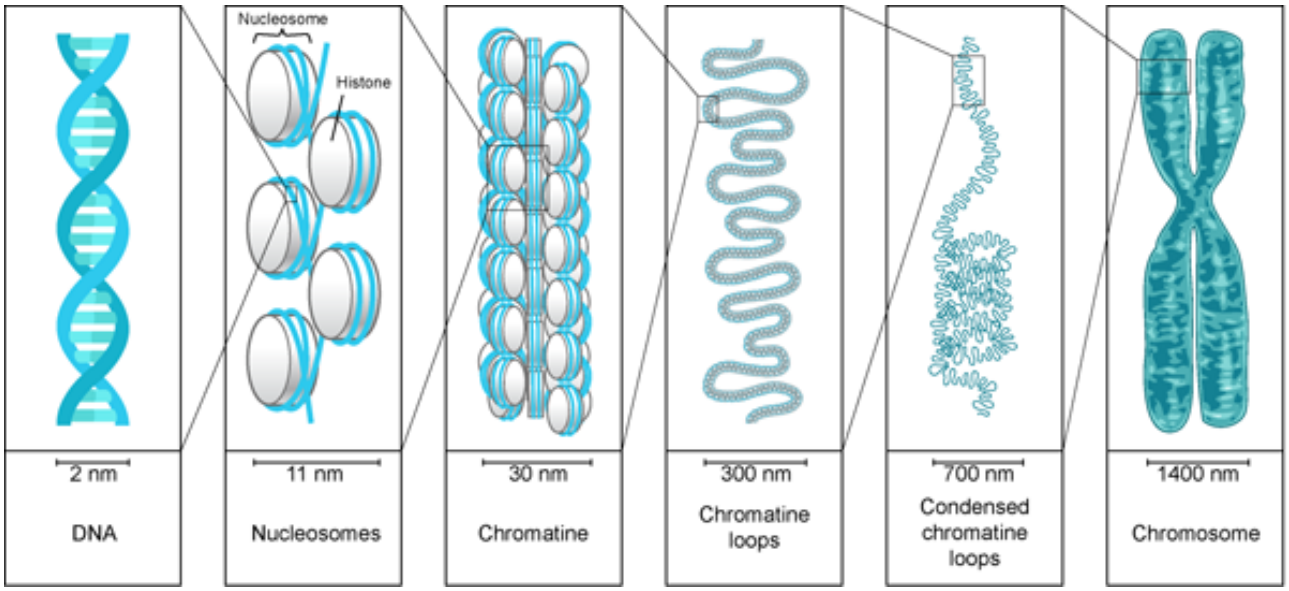

Figure 1.11: The major structures of DNA compaction. Image taken from Shmoop Biology [43].

allow access to proteins for transcription and replication. The genes being actively transcribed are found in the euchromatin, or bead on a string state. This is the case for 92% of the human genome [47]. The inactive state is associated with the tightly packed heterochromatin state of the 30 nm fibre. In prokaryotes, plectonemic supercoiling is the dominant form due to the circular nature of the chromosome [45]. Conversely, in eukaryotes, plectonemic supercoiling typically only arises when the DNA is not tightly constrained by the nucleosome. It plays a key role in managing torsional stress during vital processes like transcription and replication. Hereinafter, I focus the discussion solely on plectonemic supercoiling.

Dynamic changes in supercoiling act as a regulatory mechanism, influencing gene activation or repression in response to environmental cues or cellular needs. In the case of transcription, RNA polymerase is required to rotate relative to the DNA due to its helical nature. As the size of the growing RNA chain increases, this becomes difficult, thus making it more energetically feasible to rotate the DNA itself. As a result, positive supercoils are generated ahead of the advancing RNAP, whilst negative supercoils form behind it (fig. 1.12). This is known as the twin-domain model of transcription [49]. Negative supercoiling, or unwinding, inherently makes it easier to open the helix for starting transcription, thus contributing to basal transcriptional regulation [50]. The sharp bend of the plectoneme tips aid RNA polymerase as DNA bending is required to open the polymerase complex at the start of transcription [51]. Topoisomerases are the class of enzymes responsible for releasing the supercoils generated from RNAP movement. In eukaryotes, DNA topoisomerase of type I changes the linking number by +/- 1 via a transient single-stranded break, whilst topoisomerase type II utilises a double-stranded break. They function by nicking either one or both strands of the DNA whilst simultaneously binding to the cleaved DNA via a phosphotyrosyl bond. The other strand is passed through the break to change the linking number and the nicked backbone is restored. The accumulation of the positive supercoils impairs transcription elongation and so, along with topoisomerases, serve

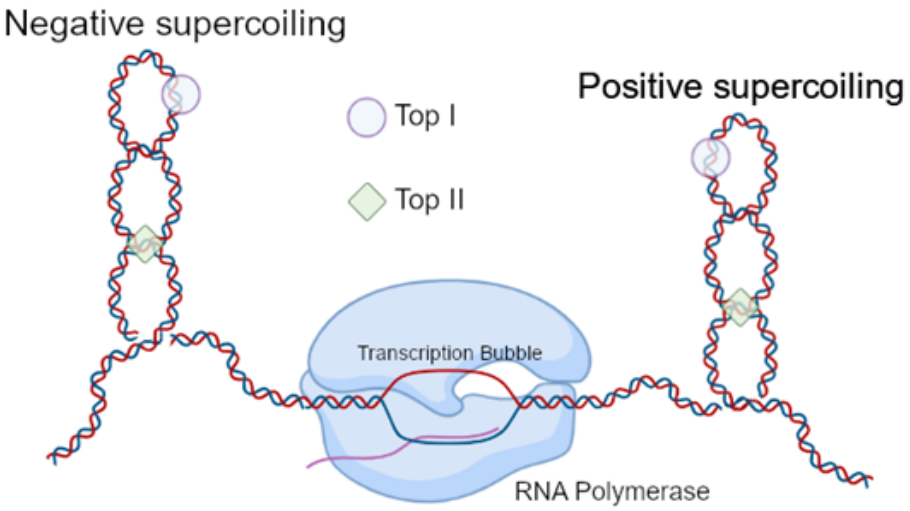

Figure 1.12: In the twin-domain model of transcription, DNA becomes positively supercoiled ahead of RNA polymerase and negatively supercoiled behind. The supercoiled nature is controlled by topoisomerases (top) I and II. In between, the transcription bubble is a completely unwound segment where the coding strand (red) and the template strand (blue) separate, and a transcript RNA strand (pink) is formed complementary to the latter. The direction of transcription is running left to right. Figure adapted from Ma *et al.* [48].

to regulate the process via feedback mediation [48, 52, 53]. Additionally, without the removal of supercoils, the topology of the DNA can be greatly affected resulting in non-B-DNA forms such as R-loops [54], cruciforms [55, 56], Z-DNA [57, 58], and P-DNA [59–61]. These all act as either transcription factor binding targets or transcription obstacles in order to regulate the process. For example, R-loops are an RNA-DNA hybrid of double-stranded DNA joined to a displaced single strand of RNA. R-loop formation inhibits transcription [62, 63]. Additionally, they are important motifs for preventing the methylation of some promoters [64], but are also targets for DNA damage themselves due to the exposure of the strands [65, 66] and thus must be well-regulated [67]. It has been shown that the formation of R-loops is facilitated by negative supercoiling by a minimum of one order of magnitude [68].

The direction of supercoiling further affects gene expression. Unwinding facilitates the strand opening of promoter regions to enable transcription initiation and the recruitment of regulatory proteins such as the TATA binding proteins [69]. Topological promoter coupling further enables long-distance effects as transcription at one promoter region is able to affect a distant promoter region due to propagation of the supercoiling wave [70, 71]. Conversely, positive supercoiling elongates RNA polymerases [72] and aids transcription by removing DNA-bound proteins that would otherwise block the progress [73]. However, a build-up of supercoiling represses transcription [74] and has been shown to diminish the synthesis of mRNA [75].

The transcription process for highly expressed genes occurs in bursts - this is the case for both prokaryotes [71, 76–78], and eukaryotes [79–81]. It is the phenomenon of ceasing production of transcripts over a sufficient period of time to inhibit transcription and elongation. The mechanism by which this occurs has been shown to be linked to the supercoiled nature of the chromosomal

DNA [82] as genes that are highly expressed are more strongly linked to topoisomerase II and negative supercoils and less so towards positive supercoils [83]. Transcriptional bursting increases the diversity of a population and enhances survival [84–86]. This bursting results from supercoiling dynamics as dissociation and rebinding of topoisomerases causes stochastic variations in superhelical density which drives the phenomenon [82]. Overall, supercoiling facilitates gene activation and may provide an explanation for the evolutionary clustering of genes together within the chromosome [87] as plectonemes bring areas of the DNA together that otherwise would not be, to regulate gene expression [88]. This especially applies to enhancer and promoter regions of the DNA. Plectonemes are important for nucleosome positioning [41] and their positioning is reliant on signals from near the promoter regions [50]. Supercoiling is a means of preventing repositioning within the nucleosome at coding regions and the prevention of supercoiling diffusion at gene boundaries has a regulatory role [83].

Similarly, there is a compensatory overwinding ahead of the replication fork due to the necessary unwinding for strand separation during DNA replication [89, 90]. As the size of the strand-separated region increases as replication progresses, further turns in the helix are removed. This would reach a critical point of compensatory positive writhe where the increase in energy is too large and further replication of the DNA is blocked. Topoisomerases alleviate this stress by progressively reducing the linking number to ensure smooth progression of the replication machinery and to avoid replication stalling [91].

Finally, an increasing writhe is also linked to an increased compaction of the DNA [40] which aids in cellular packing. However, at the superhelical density of genomic DNA, there is an unexpected reduction in compaction due to the onset of kinks and denaturation bubble defects [40]. The onset of defects is able to better dissipate the torsional stress compared to the high-energy increase in writhe. The advantage of this is two-fold: firstly, the defects can act as flexible hinges which makes a greater range of DNA conformations accessible; secondly, they allow for a global relaxation of the DNA which permits a greater range of conformations, thus increasing the overall accessibility for DNA-protein binding.

When DNA supercoils due to any of the aforementioned actions, if the ends of the molecule are allowed to rotate, the superhelical strain will dissipate. Therefore, for supercoiling to have a structural or functional role, it must exist within a constrained environment. In the eukaryotic genome, large (1 Mbp) topologically associated domains (TADs) exist as loops and provide a structural foundation to the genome [92, 93]. Recent evidence has shown these regions to be supercoiled [94]. The boundaries contain an abundance of binding sites for the CTCF transcription factor suggesting that they may act as supercoiling boundary elements [95]. Mapping the distribution of DNA supercoiling has unveiled the presence of 100 kb dynamic supercoiled domains that are both over- and under-wound [87]. The patterns of the domains were dependent on both topoisomerases and transcription and so they form due to the balance between the two. Only 10% of the domains corresponded to TAD boundaries. These large

supercoiling domains are joined by smaller foci of approximately 1.5 kb at transcription start sites [96] that, together with the broader supercoiling domains, are organised into larger structural loops [95]. It has been suggested that larger domains form via diffusion of the smaller foci at promoters and that the drag from the interactions with chromatin may hinder rotation and be a sufficient barrier to supercoiling diffusion at boundaries. However, the relationship between the smaller foci and the broader domains, and the mechanisms by which they are topologically constrained remains unclear.

Sequence Specificity

The positioning of plectonemes is strongly influenced by DNA sequence [41]. This is not surprising given their affinity for promoter sequences and crucial sequence-mediated biological roles. As plectoneme tips are inherently sharply bent, there is an energy barrier imposed to their formation. However, if the sequence in that region is able to kink or locally melt, or if it is already intrinsically curved or more flexible than the surrounding DNA, the barrier height decreases [41]. These properties are often correlated with the AT distribution within the sequence and, biologically, plectonemes locate at AT-rich regions as they are made more accessible to RNAP, transcription factors, and topoisomerases. However, intrinsically curved GC-rich sequences have also been shown to favour plectoneme positioning in positively supercoiled DNA [41], showing that it is indeed curvature, and not necessarily sequence, that drives the pinning.

The ability of DNA to form defects, such as kinks, or to open, which aids plectoneme positioning [41], is also linked to sequence as well as the bendability of the local region [97]. Defects refers to any local disruption to base pairing and stacking. AT sequences have a lower thermodynamic stability than GC base pairs, and alternating purine-pyrimidine sequences, such as AT, have weakened base stacking compared to homopolymeric sequences, such as AA [97]. As a result, base pair combinations are able to absorb torsional stress to varying degrees [98]. This allows AT-rich sequences to act as sinks for superhelical stress [97]. Therefore, AT-rich sequences are often correlated with defects, increased writhe, and plectoneme positioning, whereas an increased GC content, due to its higher rigidity, tends to melt [99]. Furthermore, co-operative kink formation on opposite sides of a topologically constrained circular DNA system is optimal for relieving both bending and torsional stress [97] which suggests an additional influence on the relative positioning of these regions within the system and not just the local sequence. The formation of kinks localises counterions to that region in order to stabilise the build-up of negative charge as a result of the kink bringing regions of DNA in close proximity to one another [100]. This built upon an initial model that compared base-flipping and denaturation as a consequence of supercoiling [101]. Where there is a smooth bend, the stress is evenly distributed. However, as the bending angle progresses to a writhe-mediated sharp bend, as found in plectoneme tips, base flipping ensues in order to act as a flexible hinge and relieve some of the strain. Denaturation bubbles may also act as flexible hinges in negatively supercoiled DNA in order to allow kink formation and favour a more open global DNA conformation due to

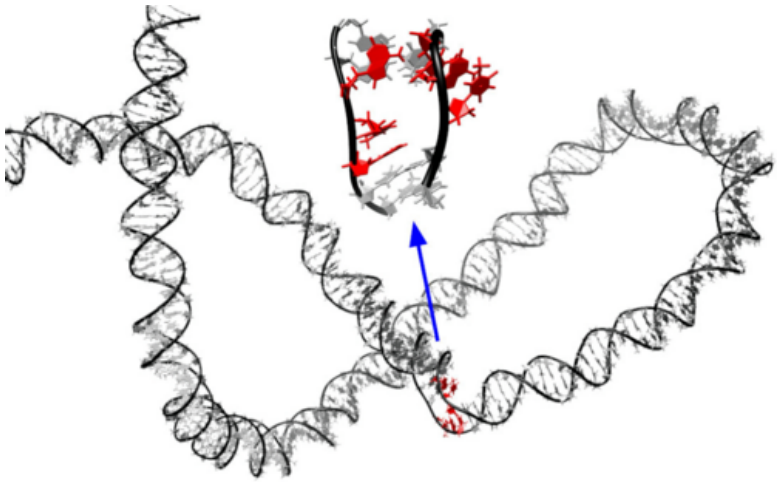

Figure 1.13: The conformation of a 3 nt bubble (red) within a plectoneme. Figure taken from Burman and Noy [106].

the dissipation of torsional stress.

Plectonemes are dynamic structures [102] which is important for long-range communication and gene regulation [103]. On the 20 ms timescale of magnetic tweezer experiments, both a diffusive and a hopping motion have been observed [102]. The former requires the DNA to "slither" within the supercoil, in addition to the overall movement of the plectoneme itself. This is slower with respect to the hopping motion, which sees one plectoneme disappear with the simultaneous nucleation of a separate plectoneme at a distant site along the DNA. The plectonemes have been observed to hop over distances of around 5 μm. The energy for such motion is governed by the nucleation and writhe transfer, which requires the intervening DNA to rotate, and it is independent of the ruggedness of the energy landscape.

A crucial stage of plectoneme nucleation, and therefore dynamics, is DNA buckling [36]. This refers to the transition from linear DNA to a plectonemic structure under torsional strain, as opposed to storing the stress in the form of twist [104]. It occurs abruptly [105] and the transition state between the two is a highly twisted single loop which allows the process via the disruption of the harmonic elasticity of the DNA [36]. The local defects that aid plectoneme positioning may also increase the rate of nucleation by stabilising this transition state [36].

It is important to note that intrinsic curvature has been shown to be the main driver of plectoneme positioning only in positively supercoiled DNA [41]. The same cannot be said for negatively supercoiled DNA where it is more difficult to disentangle the effects of intrinsic curvature and bubble formation. The roles of positive vs negative supercoiling are similar and yet distinct; thus, explaining the biological significance of the fact that positive and negative plectonemes have similar but distinct sequence-driven positioning landscapes. A melted region over at least three nucleotides (nt), called a denaturation bubble (fig. 1.13), describes the breaking of the hydrogen bonding in a short region of double-stranded(ds)DNA to form single-stranded(ss)DNA bubbles [107, 108]. Bubble formation is highly sequence-specific and absorbs

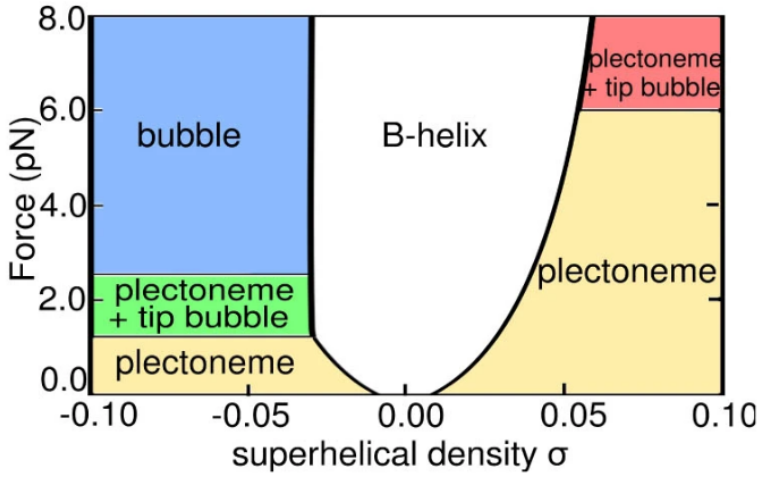

Figure 1.14: The supercoiling response to torsion is often desscribed as a "hat curve". Under negative supercoiling, plectonemes exist at low force and tip bubbles form as force increases. Beyond approximately 2 pN, the bubble regime dominates. Conversely, under positive supercoiling, the plectoneme regime dominates up to approximately 6 pN force, at which point tip bubbles form. Figure taken from Matek *et al.* [112].

the torsional stress caused by under-twist by breaking bonds [109] and predominates at high force regimes [110]. In physiologically relevant conditions, the response is plectoneme-dominant with tip bubbles that pin the plectoneme in place and limit diffusion [111, 112]. The existence of a bubble further imparts flexibility onto the local DNA which, flexibility itself, is known to aid plectoneme positioning [41]. The highly sequence-specific, as well as plectoneme-positioning, nature of the bubbles means that Dekker model of intrinsic curvature and flexibility cannot be considered alone in the case of negative supercoiling. The supercoiling response to torsion and tension is known as the “hat curve” (fig. 1.14) [110] and describes the relationship between twist, and plectoneme and bubble formation.

The location of bubble formation can be predicted using the Stress-Induced Duplex Destabilization (SIDD) algorithm [113]. Typically, bubbles occur in AT-rich regions due to the relative thermodynamic weakness of the AT base pair compared to GC [114, 115]. The initiation steps of transcription and replication both require local separation of the DNA strands and so it stands to reason that these form away from coding sequences and are especially prevalent amongst promoter regions and other major regulatory sites [116, 117]. The SIDD model is limited in that it assumes all torsional stress is stored as twist; thus, over-predicting bubble formation in plectonemic systems [118]. To date, there is no cohesive model that combines the effects of intrinsic curvature and bubble formation to predict plectoneme localisation in negatively supercoiled DNA.

The Persistence Length

Somewhat contradictorily, DNA is referred to as both rigid and flexible depending on the context, and understanding the mechanical intricacies of DNA is crucial as they undoubtedly influence

Table 1.15: *The nearest neighbour (NN) parameters, ΔG°_{37}, of each base step. The model assumes that the stability of the base pair is dependent on the identity and orientation of the neighbouring base. These are the parameters incorporated into the SIDD model. Data from the SantaLucia Nearest Neighbour parameter set [119].*

Base Step	ΔG°_{37}/kcal mol^{-1}
A A	-1.0
A T	-0.88
T A	-0.58
C A	-1.45
G T	-1.44
C T	-1.28
G A	1.30
C G	-2.17
G C	-2.24
G G	-1.84

biological function. The persistence length is a measure borrowed from polymer physics that is heavily discussed in the DNA field. The persistence length is a measure of stiffness, with a higher length meaning a stiffer polymer. It quantifies the minimum length at which thermal fluctuations can spontaneously cause the polymer to bend. The persistence length of DNA is approximately 50 nm, or 150 bp [120]. It is an important property for considering the probability of cyclisation and loop formation, and DNA extension, to name a few.

DNA is one of nature's stiffest polymers [121]. Yet consider the primary role of DNA: to store genetic information. The human nuclear genome is 207 cm in length [3] and so packing the long double helix into the very limited space of the nucleus depends largely on DNA stiffness [122]. However, DNA does not simply store the data. To extract and use the information, processes such as transcription and replication rely largely on DNA unwinding, looping, and supercoiling. Thus, DNA function requires the complex balance between compression for packaging and accessibility for transcription and replication. As a result, it is a locally rigid structure at distances below its persistence length, but a globally flexible structure [123].

DNA stiffness can be attributed both to the negatively charged nature of the phosphate backbone and the stabilising base pair stacking interactions. The intramolecular repulsion from the backbone resists local bending and expands the DNA, whilst the stacking interactions are compressive and make conformational change unfavourable. At physiological conditions, the contribution of these two forces to DNA's rigidity are approximately equal [123]. As base stacking is a major contributor to DNA stiffness, and the free energies of stacking are dependent on sequence [124], it was initially proposed that the sequence-dependence arose from these

differences [125–127]. However, this has since been disproven and it has been shown that the relationship relies on a much more complex interplay of factors [128].

The sequence-dependent differences in persistence length arise largely due to differences in the intrinsic shape of sequences with only a small contribution from the stiffness of their component base pairs [129]. It is for this reason that there is often a distinction between the static, l_s, and dynamic, l_d, persistence lengths where the traditional persistence length, l_p, is equal to $1/l_s + 1/l_d$ [130]. Here, the static persistence length describes the intrinsic curvature of the DNA due to its specific sequence, whilst the dynamic persistence length describes the flexibility of the DNA due to thermal fluctuations, or in other words, its bending rigidity. There is only a weak sequence-dependent contribution to the dynamic persistence length.

By measuring the persistence length of various, approximately 200 bp, quasi-periodic sequences, the persistence length of small fragments was calculated [131]. It was estimated that these sequences deviate from the average persistence length value by up to 15%. This is significant as the same study found that a variation in the persistence length of 2.1 nm is sufficient to change the bending energy in the nucleosome by 10-fold [131]. Such an amount may explain the selectivity of nucleosome binding as well as protein binding affinities in such cases where the DNA is strongly bent.

It is without doubt that there is a sequence-dependence to the persistence length of DNA and, by extension, its mechanical properties; however, a survey of current literature shows that there is not one consensus able to fully describe the intricacies. Whilst some studies describe an almost perfect linear trend between increasing GC content and increasing persistence length [132], others show a negative quadratic relationship peaking around 50% GC [131]. However, some sequences are exceptional in their cases and are generally well-agreed upon in the field. For example, poly(A) sequences are exceptionally straight and stiff, whilst repeating, phased poly(A) sequences called A-tracts are overall exceptionally bent and stiff [129, 133]. In general, YR (pyrimidine-purine) steps are the most flexible due to their hinge-like behaviour whilst RY (purine-pyrimidine) steps are the most rigid. RR (purine-purine) steps fall somewhere in the middle [16, 128, 134]. However, this is oversimplified, and it is often the case that A/T rich sequences are more extreme in their mechanical properties, including being both rigid (AAAA, AATT) and flexible (TATA, CACA) compared to G/C rich sequences that sit around the "average" [133, 135]. The flexibility of the various steps has been attributed to the flexibility in all three of the angular parameters (twist, roll, and tilt) as well as in rise, with the stiffness in each parameter following the overall trend in stiffness: YR < RR < RY [128]. YR steps are the most flexible in the cases of both roll and tilt; whilst RR and RY have similar tilt force constants and differ only in that the roll of RY is stiffer than RR. In the case of twist and the YR steps, TA is stiffer than CG or CA. The stiffness in tilt and roll shows little-to-no variation in the values dependent on surrounding sequence context, whilst rise shows some and twist has a significant variation in some steps. Shift and slide show no easily discernible dependence

on sequence. By investigating the correlation coefficients of the base step parameters between neighbouring base pairs, it was found that the majority of the base pairs were only correlated to the nearest neighbour with slide being positively correlated, and twist, shift, title, and rise being negatively correlated. Roll is only very weakly negatively correlated. Rise, especially, and slide are further correlated to the next-nearest neighbour.

However, it is unfeasible to account for differences in persistence length on a global scale by only considering base pair steps. There is limited understanding on how these bends on a base step level combine to exert a global bend [136]. This is further compounded by the influence of periodicity, i.e. the placement of the sequence in relation to the phase of the helical pitch, on the "bendability" of DNA. This is well-studied in the case of A-tracts and the phasing of these explains the high intrinsic curvature induced by these motifs despite they themselves being stiff. Overall, this topic remains largely unresolved.

1.3 DNA Damage and Repair

DNA is susceptible to damage at various stages of the cell cycle. During transcription and replication, a major role of RNAP and DNAP is to proof-read the DNA and check for aberrant bases. RNAP is known to pause and back-track to avoid transcribing damaged DNA [137], whilst DNAP is able to correct mismatched bases [138]. This is vital as errors in these processes can lead to mutations and epigenetic changes [139]. Chemical modifications such as oxidation, alkylation, and deamination can change the hydrogen bonding pattern and result in incorrect base-pairing within the DNA. If the cell retains the DNA damage, transcription can be blocked thus preventing the translation into a protein; or transcription can occur resulting in a mutant transcript and, later, a mutated protein. These proteins can be important in the development of tumours [140]. A rate of up to 1,000,000 molecular lesions per cell per day has been suggested [141] and the steady-state levels of damage represent the balance between damage accumulation and repair. Once a cell reaches a critical accumulation of damage, or can no longer effectively repair its damage, one of three things may happen: senescence, which is a permanent state of cell dormancy; apoptosis, which is the programmed cell death; or unchecked proliferation of the cell, a major hallmark of cancer. There are a number of DNA repair pathways in the body that exist to combat this.

DNA Mismatch Repair

A point mutation is where a single nucleotide is inserted, deleted, or changed within the DNA sequence. They occur rarely and spontaneously at a frequency of 1 occurrence per 10^9-10^{10} base pairs per cell division [143]. The result is a mismatch, where the bases in the base pair are non-complementary. Mismatches mostly arise due to errors during replication [144], but can also be caused by ionizing radiation, mutagenic chemicals, and via spontaneous deamination [145]. Strand slippage and the formation of higher secondary DNA structures, especially within

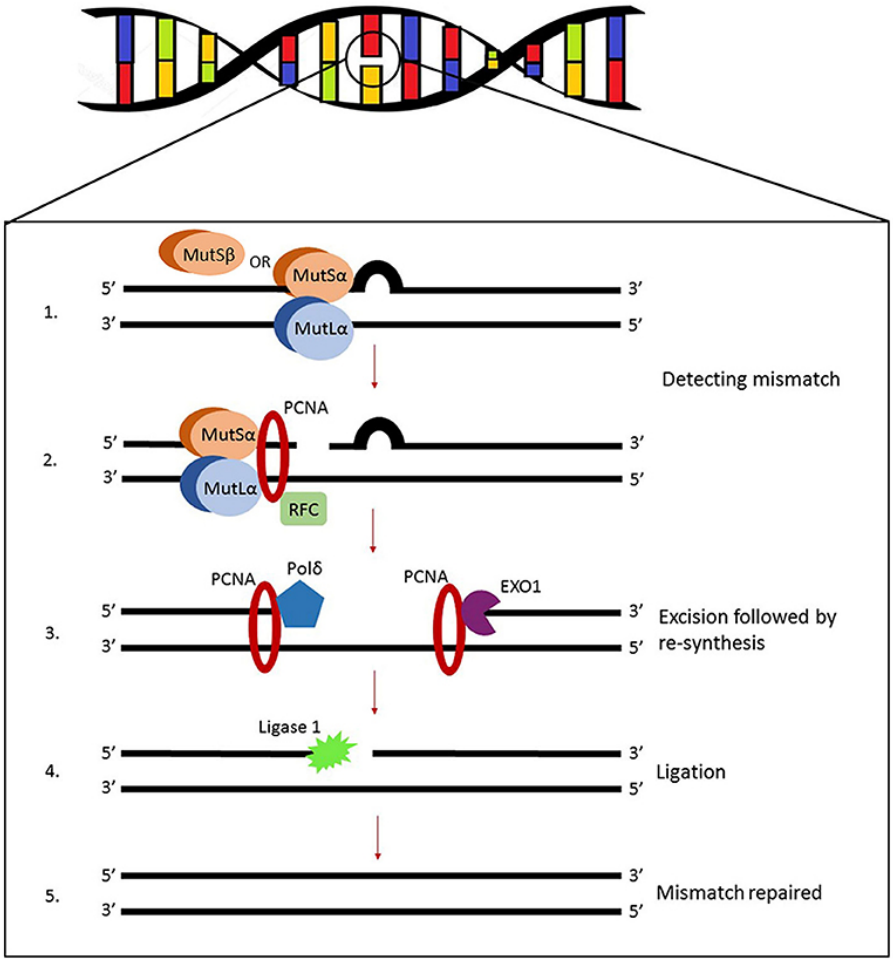

Figure 1.16: The eukaryotic MMR system. Figure taken Pećina-Šlaus *et al.* [142].

repetitive sequences, is also responsible for DNA mismatches when the base pairs are incorrectly processed during repair, replication, or recombination [143]. Consequently, if left uncorrected, they can cause genetic mutations due to their impact on protein sequence.

DNA mismatch repair (MMR) is the system responsible for recognising and repairing mismatched bases (fig. 1.16). It is the second line of defence for mismatches arising from replication. DNAP has an error rate of approximately 10^{-4}-10^{-5} per nucleotide during DNA synthesis, which can be reduced to 10^{-7} via proof-reading [146]. DNAP is able to recognise and reject mismatched bases because of their poor base stacking compared to canonical base pairs [147]. MMR reduces the error rate further to approximately 10^{-9}-10^{-10} per nucleotide [148].

In eukaryotes, the major components of the repair system are the MutS Homolog (MSH) and MutL Homolog (MLH) proteins, named due to their homology with the *E. coli* MMR system involving MutS, MutL, and MutH [149]. The pathway consists of five steps and is highly conserved amongst bacteria and eukaryotes [150]. These are recognition, incision, removal, resynthesis, and ligation. In mammalian cells, MutSα binds to small mismatches up to 3 bp in length, whilst MutSβ recognises segments of up to 13 bp [148]. ADP is then exchanged for ATP to form a clamp-like structure that slides along the DNA and recruits the MutL homologs [151].

The newly synthesized DNA strand is nicked to provide the entry point for Exo1 to excise the mismatch containing strand. DNAP then resynthesises the excised strand.

A major result of DNA mismatches is microsatellite instability (MSI). Microsatellites are repeating 1-5 bp DNA segments up to 60 bp in length. The human genome contains between 50,000 and 100,000 dinucleotide microsatellites, i.e. being of two repeating base pairs, with higher microsatellites being less common [152]. They are often located in non-coding regions; but can also be found in coding and regulatory regions. Due to their repeating nature, a mutation in a microsatellite will lead to the gain or loss of an entire repeat unit. MSI is the variation in microsatellite repeat lengths. Mismatch repair deficiency is associated with some tumour types, mainly colorectal; whilst MSI have been associated with a wide range of cancers including breast, cervical, lung, and prostate [153]. It has been estimated that mismatches are associated with up to 30% of spontaneous cancers [154].

Base Excision Repair

The base excision repair (BER) mechanism (fig. 1.17) repairs small base lesions due to oxidations, alkylations, deaminations, and uracil incorporation. DNA glycosylases are responsible for the recognition and initiation of this pathway [155]. They remove the base whilst keeping the sugar-phosphate backbone intact, creating a structural motif known as an abasic, or AP, site. They do this by first flipping the DNA lesion out of the double bond, then cleaving the N-glycosidic bond. The abasic site is cleaved by AP endonuclease which creates a nick in the DNA backbone to create a single-stranded break. DNA polymerases synthesise either one new nucleotide in a process called short-patch BER, or they can synthesise up to 10 new nucleotides in long-patch BER. The choice between the two is not currently understood [156].

In addition to being BER intermediates, abasic sites can also form spontaneously or as a result of exogenous damage. It is much more common in purines, with approximately 10,000 apurinic sites generated daily, compared to just 500 apyrimidinic sites [157, 158]. Left unrepaired, they can cause mutations and stalling at the replication fork. The oxidation of guanine to produce 8-Oxoguanine (8-oxoG) receives a significant amount of attention in the field. Oxidative damage is a major reaction to reactive oxygen species (ROS) that can cause base and sugar alterations, crosslinking, and sugar-base cyclisation [159]. They arise from exogenous agents such as air pollution, UV light, ionising radiation, exposure to pesticides and metals, as well as some lifestyle and diet choices. 8-Oxoguanine is one of the most common DNA lesions resulting from this [160] with a steady-state level of approximately 2,400 per cell [161]. However, the levels increase with age [162, 163] and are a measure of oxidative stress, which in itself is an important marker for the development of cancer [164], Parkinson's disease [165], Alzheimer's disease [166], heart failure [167], and autism [168] to name a few. 8-oxoG can form a Hoogsteen base pair with adenine resulting in a GC to TA substitution [169].

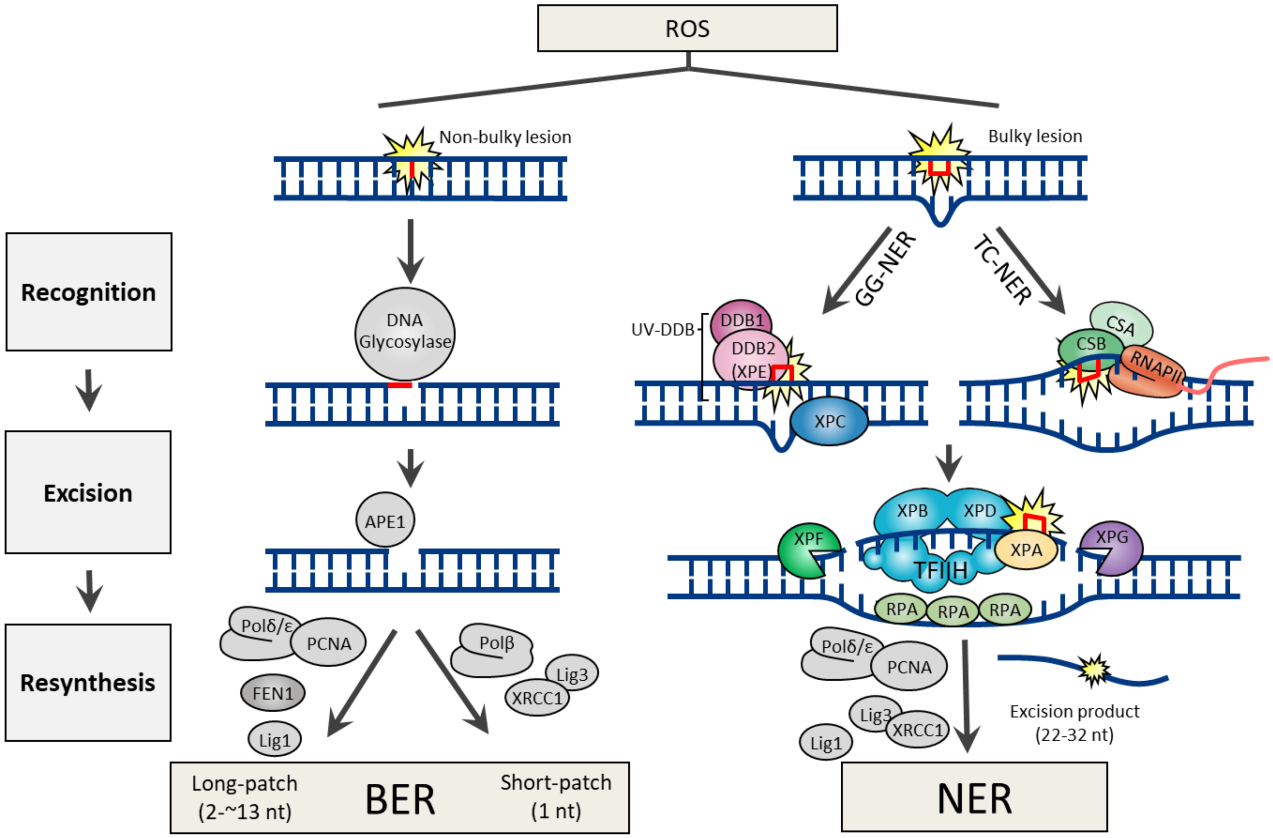

Figure 1.17: Both the BER and NER pathways proceed via the same three general steps: recognition, excision, and resynthesis. Figure taken Lee and Kang [170].

Nucleotide Excision Repair

The nucleotide excision repair (NER) pathway (fig. 1.17) repairs bulky DNA adducts induced by UV light. These lesions are usually in the form of thymine dimers—the product of [2+2] cycloaddition across the C=C double bonds resulting in covalent bond formation between adjacent thymine bases [171]. Repair can proceed via global genomic NER (GG-NER) where the damage is recognised by the helix distortion, or transcription-coupled NER (TC-NER) where the pathway is initiated upon RNAP stalling at a lesion. Upon recognition, both pathways utilise endonucleases to remove a 12-24 nt segment of ssDNA containing the lesion [172] leaving the undamaged strand to act as a template strand for DNAP to synthesise the complementary sequence. DNA ligase then re-forms the double strand. A deficiency in the NER pathway greatly increases the risk of cancer development. Due to the role of UV radiation in the creation of damage, skin cancer is by far the greatest risk with a greater than 5,000-fold increase [173]. The accuracy of the repair pathway is crucial, as up to 100 UV-mediated reactions can occur per second upon exposure of a skin cell to sunlight. Unrepaired lesions can create errors in transcription and replication, completely stall polymerase function, and contribute to the production of melanin and sunburn.

Ribonucleotide Excision Repair

By far the most abundant form of DNA lesion is the incorporation of a ribonucleotide into DNA; in other words, the oxidation of the 2' carbon of the sugar ring. As a result, the ribose site is approximately 100,000 times more likely to undergo spontaneous hydrolysis [174] which makes the DNA strand more susceptible to breakage and mutagenesis [175–177]. The number of ribonucleotide lesion events exceeds the total events of all other forms of DNA damage combined by an order of magnitude [178]. They arise due to errors in DNA synthesis [175, 179] at a rate of approximately one every few thousand base pairs [175], as well as during transcription [180, 181], and telomere elongation [182]. The active site of DNAP is able to distinguish between the normal deoxyribose and ribose lesion [183]; however, the high damage rate challenges its fidelity and large accumulation of ribonucleotides can lead to replication fork stalling [184]. It has been estimated that the proof-reading abilities of DNAP in terms of ribonucleotides is only one-third as effective as when noticing incorrect DNA base pairing [185].

The RNase H2 endonuclease is responsible for recognising a single lesion and initiating the ribonucleotide excision repair (RER) pathway [186–189], whilst RNase H1 recognises a series of four consecutive ribonucleotides [190] such as is the case with R-loops. The enzymes cleave the 5'-phosphodiester bond to the ribonucleotide [191] that, when left intact, has been linked to the genetic neurological disorder Aicardi–Goutières syndrome [192, 193]. The enzyme introduces a single-strand break 5' of the ribonucleotide creating a 3'-hydroxyl and a 5'-phosphate group. Polymerases perform strand displacement synthesis on the 3'OH DNA creating a flap that is removed by an exonuclease. DNA ligase then seals the nick [188] in a pathway analogous to short-patch BER.

Direct Reversal

O6-methylguanine is the alkylation of guanine's carbonyl oxygen. Like 8-oxoguanine, it causes a mutation in the DNA due to incorrect base pairing. Here, it pairs with T causing a GC to AT transition in the DNA. Most often, the methylation is due to N-nitroso compounds found in tobacco smoke and some foods, and form in the gastrointestinal tract, especially after consuming red meat [194]. It has long been established as the most mutagenic and carcinogenic alkylated base [195], with estimates varying between one mutation per eight unrepaired O6-methylguanine lesions [196], to one in three [197]. The mutagenicity of the base is largely sequence-dependent, with the more sterically accessible bases being more easily repaired [198]. Repair is primarily carried out by the MGMT methyltransferase. It de-methylates the guanine in an S_N2 mechanism and is not regenerated, thus being referred to as a suicide enzyme. Deficiency in MGMT is associated with 61% of cervical cancers [199], 40-90% of colorectal cancers [200–203], 87% of thyroid cancers [204], and 32-88% of stomach cancers [205, 206], amongst others. Other examples of direct reversal include the AlkB protein in *E. coli* that removes the alkylation damage resulting from S_N2 chemical agents [207].

Single- and Double- Stranded Breaks

During the aforementioned damage repair pathways, single-stranded breaks (SSBs), also called nicks, are formed inherently as a result of lesion excision. Escape of these intermediates requires careful detection and repair. SSBs are also formed during normal lagging-strand DNA synthesis in the form of Okazaki fragments, and as a result of damage in the form of stalled replication forks or directly due to oxidation of the deoxyribose. They are the most prevalent form of DNA damage and arise at a rate of one every 1-10 seconds per cell, requiring efficient removal to prevent RNAP and DNAP stalling and collapse, R-loop formation, and genomic instability [208]. They have aberrant 3' and/or 5' termini as a result of their mechanism of formation. This can range from a 3'-phospate arising due to oxidative stress [209], bonds to topoisomerases as a result of abortive activities and normal DNA repair [210, 211], and adenylated 5' ends due to abortive DNA ligase action [212, 213]. The poly(ADP-ribose) (PARP) enzymes detect SSBs [214] and initiates a pathway that is dependent on the SSB terminus damage. DNA end processing can thus occur via a wide variety of enzymes [215]. Gap filling can then occur via short- or long-patch in a mechanism analogous to BER, which is then followed by ligation.

Double-strand breaks are particularly hazardous due to their role in genome rearrangements, as well as mutations and heterozygosity [216]. Left unrepaired, they lead to cell death and cancer. They can arise naturally during DNA replication, stalling upon encounter of a SSB, and during the BER of clustered lesions; but also as a result of ionising radiation and mutagenic chemicals breaking the phosphodiester backbone. They can be repaired via homology-directed repair (HDR) using a sister template chromatid or, in non-replicating cells, non-homologous end joining (NHEJ) that proceeds via end processing and ligation. The latter is typically guided by single-stranded overhangs, called microhomologies, that must be directly compatible in order for the repair to proceed accurately [217, 218]. Hallmarks of tumour cells such as telomere fusion and translocation can arise due to NHEJ inaccuracy [219].

1.4 Beyond Passive DNA

The recognition of DNA sequence by proteins was initially thought to be driven by base readout: a complementary pattern of hydrogen bonding between the amino acids of the protein and the bases of the DNA [220]. Each base pair has a specific signature in the major groove in the form of their specific pattern of hydrogen bonding [221]. However, the phenomenon of shape readout is now emerging [222]. This suggests that DNA is not the passive partner in protein-DNA interactions that it was once thought to be and a balance between DNA sequence and topology is necessary for DNA to perform its biological role.

In reality, DNA binding proteins must combine both shape and sequence mechanisms of readout to achieve binding specificity. In general, base, or direct, readout can allow DNA to bind to a family of proteins; whilst shape, or indirect, readout offers the higher resolution

distinction between members of the same DNA-binding protein family [223]. Conformations and deformations of DNA provide a mechanism of protein recognition that acts synergistically with sequence. It is well-known that AT-rich sequences narrow the minor groove and have an important biological role in nucleosome organisation [224, 225], transcription regulation [226–228], recombination [229, 230], replication [231], and gene silencing [232] due to their intrinsic curvature and unique mechanical properties [223, 224, 229, 230]. Arginine makes up 60% of protein-DNA contacts within narrower minor grooves, compared to only 22% in minor grooves of the normal width. It is also significantly more abundant than the second-highest contact, lysine, at less than 20% [222]. Within all narrow minor grooves where arginine contacts, there is a 78% probability that the closest contacting base pair is an AT showing a synergy between DNA sequence and conformation.

DNA is known to adopt different conformations to aid in protein binding. For example, numerous proteins have been shown to induce large bends [233, 234] and open the minor groove [235, 236] in order to optimise the DNA-protein interface. Probably the most notable example is that of the nucleosome where bending and wrapping of the DNA is crucial [237–239] and can lead to the formation of complex 3D structural motifs such as hairpins and loops [240]. But the conformational changes are not limited to binding sites. Spacer regions that separate the binding sites of the individual protein subunits may also be affected in order to bring the sites into the optimal position [241, 242]. The repair efficiency of the BER pathway varies significantly depending on the position and orientation of the damaged base within the nucleosome [243–245]. Bases facing towards the histone octamer or closer to the nucleosome dyad were less well repaired than an outward-facing base or one near the nucleosome edge [245–248]. However, DNA plays an active role in the pathway as partial unwrapping from the histone octamer enabled the access of the repair proteins to the sterically occluded bases.

However, the conformational change of the DNA does not need to span numerous base pairs or be a significant deviation in order to exert a pronounced effect. In prokaryotes, DNA curves upstream of transcription start sites [249–251] as it has been shown that curved DNA facilitates RNA polymerase binding [252–255] and codes the signal for plectoneme tip formation [41]. DNA bending and kinking is a common occurrence in DNA-protein binding [256]. The resulting bend can be as little as 12° as is the case for the bacteriophage 434 repressor [257] or in excess of 160° as seen with the integration host factor (IHF) [258]. Recently, it has been shown that DNA with some intrinsic curvature improves IHF binding that is further enhanced by supercoiled DNA [42]. Here, the DNA has an active role in driving the protein interaction and may also modulate the protein's function. The DNA-IHF configuration changes across replication [259], phage Mu transcription [260], and Tn transposition [261], which is further influenced by the level and direction of DNA supercoiling [42].

DNA supercoiling is an important mechanism of achieving conformational recognition. Overall, supercoiling releases DNA from the free energy minimum that linear canonical DNA is trapped

in, and the wider variety of conformations creates more potential protein recognition sites. Some proteins show a higher affinity for a specific direction of supercoiling. For example, topoisomerase IV was able to recognise and relax positively supercoiled DNA approximately 3-fold faster than negatively supercoiled substrates [262] and with an approximately 5-fold higher affinity than for relaxed DNA [263]. Secondly, in the first step of transcription, the TFIID transcription factor binds to the TATA promoter sequence and recruits the other components to the promoter region to form the preinitiation complex [264, 265]. The TFIID transcription factor has a much higher affinity for negatively supercoiled TATA elements [69] and the TATA box is able to absorb the majority of the torsional stress by melting in order to achieve an overall twist close to that of B-DNA [266]. Finally, triplex-forming oligonucleotides (TFOs) are major groove binders that bind to locally underwound 16-mers [220]. They form a triplex DNA of the single-stranded TFO bound to the target double-stranded DNA [267]. They have roles as transcription factors, cross-linking agents, and as a means to control gene expression [268]. The formation of the triplex structure is specific to negative supercoiling as positively supercoiled DNA cannot form the Hoogsteen hydrogen bonds necessary to counteract the electrostatic penalties of triplex formation.

Supercoiling has recently been shown to be involved in DNA allostery. Allostery in proteins is well-documented; however, cases of DNA modulating protein binding at two distant sites are only just emerging [269–275]. It has been shown that the relative positioning of two DNA-bound proteins influences the torsional stress of the DNA and that the lowest energy conformations are when separated by full helical turns [276]. Interestingly, the conformation of the proteins bound to supercoiled DNA is also lower in energy than the equivalent relaxed DNA complex, suggesting an allosteric role for the torsional stress and twisting stiffness. It has been suggested that elastic deformation allows for mechanical signalling to improve the cooperative binding of multiple proteins. This contrasts the idea that DNA supercoiling mediates allostery structurally by bringing distant sites together via looping [277, 278]. But it has not been the first time that tension has been the modulator of DNA allostery. It has also been shown that tension is propagated along a spacer region between two A-tracts to improve protein binding at these sites in a non-supercoiled system [269].

The protein-DNA search problem

This brings about the question - how do the proteins know where their target sequence is in the first place? It seems unfeasible that they can be continuously searching the length of the DNA when there are 10^6-10^9 decoy sites to contend with [279]. This is often described as the *protein-DNA search problem* and has been a major area of research since 1970 when Riggs *et al.* showed that the *lac* repressor binds to its target DNA with a rate up to two orders of magnitude faster than that of 3D diffusion [280]. There are, currently, two main diffusion models: 1D diffusion along the DNA backbone via non-specific protein binding during the sliding search; and 3D bulk diffusion in the cytoplasm. Combining the two in a manner known as facilitated

diffusion can significantly speed up the diffusion compared to the two models in isolation [279, 281]. It has been suggested that the proteins conduct a 1D search over some maximum length of DNA, and then 3D diffuse to a separate region and conduct a 1D search there. The inclusion of sliding to the surrounding bases increases the number of nucleotides the protein searches per 3D hop, allowing the protein to reach the target due to the association of the surrounding bases in a mechanism known as the antenna effect [282]. The model that achieves the fastest delivery of DNA-binding proteins further takes into account DNA conformation [279]. Somewhat unsurprisingly, if the approximate length of searchable DNA by 1D diffusion is left unoccupied by DNA-binding proteins and nucleosomes in the region of the target DNA, and the rest of the DNA is packed and made inaccessible by regulatory proteins, then the search can be up to 100 times faster than in naked DNA [279]. It has been shown that DNA looping strongly influences the search process and that proteins are significantly more likely to make contact with the outermost point of the loop than the bulk of the DNA [283]. Furthermore, the 3D jump distance is highly correlated with the spacing between successive loops. We suggest that the location of DNA damage to the plectoneme tip is one answer to the protein-DNA search problem it fulfils the aforementioned criteria.

DNA Damage and Supercoiling

The idea that damage pins the plectoneme tips in order to reduce the search space for damage-recognising proteins is supported by a number of factors. The common hypothesis is that DNA lesions reduce base stacking and, as a result, shorten the persistence length. The flexible hinge that results is a common structural feature found to be recognised by all repair proteins from a survey of crystal structures [147]. Crystal structures of MutS bound to mismatches show sharp kinks at the mismatch towards the major groove [284, 285] and the loss of base stacking [285]. Since the bending stress at the plectoneme tip often causes these structural defects, this would favour protein binding at the tip. Furthermore, MMR preferentially protects actively transcribed genes [286] where supercoiling and plectoneme formation is prevalent. Experiments using magnetic tweezers [287], ISD [111], and coarse-grained modelling [288] have shown that under positive superhelical stress, mismatches locate to the tip of the plectoneme. The centre of the plectoneme coincides with the mismatch placement as they are able to effectively stabilise the sharp bend at the apex [288]. Indeed, the larger the mismatched sequence, the more deterministically it pins the plectoneme [288] and the larger the plectoneme due to the inherent flexibility of the damaged sequence [111]. Kinks and flipped bases in the apex of the plectoneme lower the energy of the end-loop in order to stabilise it upon protein-lesion binding. Furthermore, protein binding will be stronger due to the energetic contributions from the binding that elevate bending stress at the plectoneme tip [287]. MutS binding to a mismatch is known to be stabilised by DNA supercoiling, resulting in a longer protein dwell time [289]. Supercoiling had no significant effect on the conformational change of MutS to the sliding clamp form, which is coupled with DNA unbending, suggesting that the role of supercoiling is in the initial recognition

step.

Thymine dimers have been shown, using coarse-grained modelling, to reduce the superhelical density required for plectoneme and bubble formation [290]. At physiologically relevant conditions, the dimers locate to the plectoneme tip and increase the likelihood of tip-bubble formation thus creating highly bent and denatured states for protein-binding. The existence of tip-bubbles in plectonemes is well-established [112, 118]. Recently, it has been shown that whilst some glycosylases merely tolerate bubbles [291, 292], one human enzyme was found to have increased repair efficiency for 8-oxoguanine [291] and a cytosine oxidation product [293] at bubbles compared to ssDNA and dsDNA. Finally, the bacterial enzyme RNase T is responsible for processing RNA substrates in DNA damage repair pathways in *E. coli*. It is able to recognise lesion sites in bubbles and bulges, as well as duplex DNA [294].

Whilst BER initiators recognise very specific forms of damage in terms of their chemical make-ups, both NER and MMR initiators recognise a wide range of DNA lesions [147]. They do not recognise substrates with specific chemical motifs and so it has been hypothesised that they use conformational readout instead [295]. This is further supported by the increased flexibility at these damage sites that may create a disturbance in the energy landscape to allow a wider range of accessible conformational space. In the BER pathway, recognition of a lesion by glycosylases is based on their hydrogen bonding potential, distribution of electric charge, and their shape [147]. In general, the base excision repair pathway is initiated with DNA glycosylases recognising a damaged base, flipping it, and excising it [296]. Upon base flipping, the local DNA is kinked to form the stable protein-lesion complex [297]. Base-flipping is especially prevalent in DNA damage repair [298] as a means to break the base pairing and stacking interactions and expose the base to the repair proteins. Base flipping and kinking are also prevalent in DNA supercoiling both as a means to dissipate torsional strain in place of increasing the size of the plectoneme, and to accommodate the sharp bending angles of the plectoneme tip [40, 112, 299].

1.5 Sequence-Dependent Supercoiling Methods

As with anything, the level of resolution required of your investigation depends on the question that you are answering; this applies to both experiments and theoretical models. Supercoiling is particularly difficult as local changes generated at the base pair level are able to propagate to large global effects, up to the order of Mega base pairs [94]. Theoretical modelling of DNA supercoiling ranges from atomic-level molecular dynamics simulations that are limited to a maximum of a couple hundred base pairs, to continuum models that describe supercoiling as a field and can look at the effect of polymerases on the order of kilo base pairs [300]. Experimentally, to observe supercoiling in cells and take into account the action of e.g. nucleosomes and chromatin, only the twist, and not the writhe, is measurable [87, 96]. In free DNA, the change in twist and writhe is summed to obtain the change in linking number; however, the twist-writhe balance for a given linking number is more ambiguous when protein-bound [301] and this, coupled to

the fact that supercoils are highly dynamic and rapidly propagating [102], makes investigating supercoiling especially challenging.

Experimental Methods

The diverse conformational landscape of supercoiled DNA is challenging for traditional atomistic resolution structural methods such as nuclear magnetic resonance (NMR) and X-ray crystallography [40]. They are limited to static images of only short (approximately 30 bp) duplexes. Achieving atomic-level resolution of supercoiled DNA minicircles has been experimentally accomplished using atomic force microscopy (AFM) [40]; however, there is no way in which to know the position of the DNA, in terms of sequence, within the image and so no sequence-dependent properties can be studied. Conversely, the single-molecule experiments that allow us to probe sequence-dependent conformational dynamics lose the granularity necessary to understand why these dynamics arise. In fact, in order to adequately investigate the sequence-dependence of plectoneme positioning experimentally, the field has been limited to two scales thus far.

Methods that map supercoiling on a genomic level function by the preferential binding of biochemical probes to supercoiled DNA. These probes tend to be selective to only one direction of supercoiling [301], but are able to accommodate the influences of chromatin and nucleosomes and other cellular components relevant to supercoiling. However, establishing an accurate zero-torsion baseline upon which the resulting torsion map depends is non-trivial [302]. On a smaller scale, it has become necessary to develop novel methods to specifically investigate DNA supercoiling. Whilst these are limited in their study of free DNA, they offer a highly increased resolution and easy manipulation of sequence. Intercalation-induced supercoiling of DNA (ISD) is a high-throughput single-molecule assay that allows both the control of superhelical density, unlike AFM and electron microscopy, and easy visualisation using a fluorescence microscope [111]. The method employs intercalating dyes to induce controlled supercoiling in a surface-bound, linear DNA fragment. It is bound at each DNA terminus in such a way that it is torsionally constrained and allows for easy visualisation of the plectonemes without the need for mechanical manipulation. However, the resolution is limited by the optical diffraction limit and is unable to distinguish plectonemes separated by less than 1 kbp. Furthermore, the supercoiling is induced via the stochastic binding of the dye and so lacks the direct control of supercoiling and linking number. A novel combination of optical and magnetic tweezers, and fluorescence microscopy named COMBI-Tweez has been developed to combat these limitations; although it functions in less high-throughput [118]. It has improved upon the stochastic nature of ISD and traditional optical tweezers and allows the visualisation of bubbles unlike traditional magnetic tweezer experiments; however, it cannot differentiate between plectoneme and bubble in its visualisation. Furthermore, both methods rely on a highly custom set-up that has yet to be replicated. Overall, this necessitates a multidisciplinary approach to understanding the sequence-dependent conformational dynamics of supercoiled DNA.

Theoretical Modelling

As DNA is a polymer, it can be understood using both biological models and polymer physics. To understand the mechanical stress response of DNA, various physical models have been employed. Amongst the most widely used is the worm-like chain (WLC) model. Often, the persistence length, l_p, is measured by using the WLC model in tandem with experimental results [122]. However, in order for the model to retain its validity, the assumptions of the experiment must be thoroughly considered and met [303]. In the WLC model, the polymer, DNA, is modelled as a continuous, isotropic elastic rod. This is often used to describe semi-flexible polymers. The bending energy of DNA can be explained using the WLC model by considering a chain of segments (or base pairs) of uniform-length, d, connected at nodes. The bending angle at node i, θ_i, is governed by the bending energy described in equation 1.1 [304].

$$E_i^{WLC} = \frac{1}{2} k_B T \frac{l_p}{d} (1 - cos\theta_i) \tag{1.1}$$

Unexpectedly, it was discovered that in short (approximately 100 bp) stretches of DNA, the probability of cyclization was as much as five orders of magnitude higher than that predicted by the WLC model [305]. This sparked a debate where some follow-up work supported the WLC model for strongly bent DNA [303, 306, 307], whilst others showed it to be more flexible than the WLC model describes [308–311]. Understanding this is crucial as DNA on a tens of bp length scale is bent hundreds of degrees in processes such as prokaryotic transcription initiation [312], eukaryotic and viral DNA packing [313, 314], and when bent by architectural proteins, e.g. IHF [234].

A number of models have been proposed to account for the high flexibility of strongly bent DNA. Amongst the most popular of these are the meltable WLC (MWLC) [315] and kinkable WLC (KWLC) models [316]. The MWLC model theorizes that local opening of base pairs creates bubbles that act as flexible hinges and lower the energetic cost of bending, thus increasing the looping probability of the short DNA [315, 317]. In this model, additional variables are added that consider the two distinct states: a melted and a hybridized state (equation 1.2) [304]. This introduces the thermodynamic penalty for melting, $\Delta\mu$.

$$E^{MWLC}(k_B T)^{-1} = \delta_H \left[\frac{l_P^H}{d} (1 - cos\theta) \right] + \delta_M \left[\frac{\Delta\mu}{k_B} + \frac{l_P^M}{d} (1 - cos\theta) \right] \tag{1.2}$$

The KWLC model differs in that kinking is defined as local base pair unstacking. This allows large local bend angles but, unlike in the MWLC model, preserves base pairing. Although kinking and bubbling can occur in tandem [122]. In the KWLC model, there is a completely flexible molten state with a persistence length, l_p^M, equal to 0. In this model, a preferred bend angle, γ, is introduced to the molten state (equation 1.3) [304].

$$E^{KWLC}(k_BT)^{-1} = \delta_H\left[\frac{l_p^H}{d}(1-cos\theta)\right] + \delta M\left[\frac{\Delta\mu}{k_B} + \frac{l_P^M}{d}(cos\theta-\gamma)^2\right] \tag{1.3}$$

However, all of these models are highly generalised and lack any description of sequence-dependence. The discrete WLC model was developed to combat this [318]. Here, this takes into account the individual dinucleotide stiffness values, K, and the torsion values, ω, are summed over all dinucleotide steps of type n to give W (equation 1.4).

$$E^{circle}(\phi) = \sum_{1}^{n} K_n W_{1,n}^{circle}(\phi) \tag{1.4}$$

The twistable WLC (tWLC) model was developed as a means to incorporate supercoiling and its effects into the WLC model [319]. This expanded upon an extensible WLC model [320, 321] that introduced the stretch modulus, S, and relative extension of the DNA, d/L_c, as a result of a stretching force, F. The tWLC [319] introduced an energy cross term between the twist and stretch degrees of freedom, $g(F)$, and the twist rigidity of DNA, C, to overall model the twist-stretch response of DNA (equation 1.5).

$$\frac{d}{L_c} = 1 - \frac{1}{2}\left(\frac{k_BT}{Fl_p}\right)^{\frac{1}{2}} + \frac{C}{-g(F)^2 + SC}F \tag{1.5}$$

To date, there is no method of modelling the combination of sequence-dependence and supercoiling on a large scale. And so, we now focus on the increased granularity down to the atomic level. Below this, studying DNA supercoiling at a quantum level is in its infancy; however, it has been suggested that the superhelical stress lowers the energy barrier for quantum tunnelling, thus facilitating tautomerisation and DNA mutation [322].

Coarse-graining has been an attractive option for studying increased timescales and large systems. This has allowed DNA minicircles of over 1,000 bp in length to be studied [288, 323] and timescales in the order of μs to be achieved [97, 288]. Coarse-graining simplifies the representation of the system from the atomistic level to a reduced number of interaction sites, or beads, where each site can describe a varying number of atoms grouped together. Grouping more atoms increases the level of coarse-graining and decreases the computational cost for a simulation of a given size and length but also reduces the accuracy due to the reduced degrees of freedom within the model. The beads interact with one another according to a potential energy function, and models are often parameterised for similar systems, such as only DNA and RNA, or only proteins in order to effectively reproduce all-atom MD and experimental data. Much in the same way as the parameterisation of empirical force fields.

Coarse-grained simulations are not yet sequence-specific enough to study the intricacies of, for example, mutating just one base within a short stretch of DNA as presented later in this thesis

(see chapter 3). Currently, the CGeNArate model is sequence-specific on the tetramer level and can back-map back to atomic-level resolution via a machine-learned reconstruction algorithm [324]. However, the nature of the method means that kinks and defects are not captured in the simulations and may not be suited to simulating highly stressed DNA. Furthermore, the sequence-specificity discussed in the paper is in terms of helical parameters, tested on the 56-mer scale, and it is unclear whether the model would capture the sequence-specific plectoneme pinning as this was not the focus of the work nor the work from which the original all-atom minicircle simulations were taken from. A second coarse-grained elastic rigid base pair model is sequence-dependent to the nucleotide level [325] when incorporating only the rotational degrees of freedom and can model on the level of a helical repeat where the stretch modulus is important and considers both rotational and translational degrees of freedom. Even still, the model is not able to capture higher-order effects such as kinks. The popular oxDNA model [326, 327] has been used successfully to study DNA supercoiling dynamics [97, 288] as it well-reproduces DNA behaviour under torsional stress [112] and has shown both kink [328] and bubble [112] formation. However, despite its well-parameterised sequence-dependent thermodynamic properties, it is only able to reproduce structural and mechanical properties on the general level [329]. Therefore, all-atom resolution is still required.

Atomistic molecular dynamics has been instrumental in early studies in the field by showing the existence of kinks, bubbles and other structural irregularities [299, 330], investigating the effects of supercoiling on helical deformations [331, 332], bending stress [318], and even charge density [333]. However, explicit solvent is prohibitive for simulating DNA on the scale of more than a couple hundred base pairs [325]. Currently, implicit solvent all-atom molecular dynamics is the only method capable of bridging that gap.

The various WLC models are able to well-describe phenomenon on a large length scale or in bulk; however, DNA is shown to exert sequence-dependent behaviour on a scale of thousands of base pairs [41]. The experimental field is rapidly advancing; however, they are still limited in their ability to reach the resolutions required for sequence-specific dynamics. Even with the advent of experimental methods developed specifically for the study of DNA supercoiling, atomistic molecular dynamics are still necessary in tandem to properly understand the results [118]. Molecular simulations uniquely provide answers to the global sequence-dependence whilst still offering the atomic-level resolution necessary to understand the underlying mechanisms. However, in reality, they are but a model and a multi-scale approach combining both theoretical and experimental investigation at all length scales between atomic and genomic resolution is required to truly and completely understand the function and conformational dynamics of supercoiled DNA.

1.6 Aims

The aim of this work is to, firstly, validate the use of all-atom implicit solvent simulations for investigating the sequence-dependent effects of DNA supercoiling. No one cohesive model currently exists that is capable of capturing the combined influences of intrinsic curvature, local flexibility, and bubble formation on plectoneme positioning. We seek to demonstrate that implicit solvent simulations offer an accessible and informative approach to further investigate these trends at atomic-level resolution. To this end, we designed DNA sequences with specific, predicted plectoneme conformational landscapes and tested whether these features are reliably reproduced in our simulations. Through this, we aim to elucidate the role of bubble formation in shaping the conformational dynamics of negatively supercoiled DNA—an area that remains currently overlooked in the literature.

The second, and central, aim of this thesis is to investigate whether DNA supercoiling plays a role within the DNA mismatch recognition pathway and, therefore, can help to explain the variable repair rates of mismatches across the genome. We hypothesise that DNA mismatches exhibit varying propensities for localizing to the plectoneme tip, and that this is further influenced by their surrounding sequence context. If true, this localization could represent a mechanism for facilitating protein recognition during the DNA damage search process. Thus, we aim to uncover whether DNA supercoiling offers a solution to the protein-DNA search problem. As no existing model can accurately predict bubble formation in the presence of DNA mismatches, we aim to uncover the mechanical properties inherent to DNA mismatches that are driving their location to the plectoneme tip.

2

Research Methods

In this chapter, I explain the theory and principles behind molecular dynamics simulations and single-molecule FRET, as well as the two main predictive algorithms central to this work. This does not serve to function as an experimental section, but to give insight into how the methods work and why the various parameters were chosen. The detailed methods specific to the experiments can be found in the corresponding results chapters.

DNA can be studied complementarily via both experimental and theoretical approaches. They work synergistically to give a more complete description of the conformational dynamics of a biological system, as well as the principles and mechanisms underpinning said dynamics. Theoretical methods can reveal atomic-scale details that are difficult to access experimentally; whilst, in this work, experimental techniques serve to validate and guide simulations. Together, they provide a more robust and multifaceted understanding than either method could alone.

2.1 Molecular Dynamics

Molecular dynamics (MD) is a physics-based computer simulation method used to study the movement of atoms within a biological system. Here, Newton's equations of motion are solved to build up a trajectory of a biomolecule. Simply, the acceleration of a body is the second derivative of its position with respect to time; therefore, by knowing the mass of each atom within our system, we can employ $F = ma$ to investigate how a system moves over time (eq.

2.1).

$$\vec{\boldsymbol{F}} = m\vec{\boldsymbol{a}} = m\frac{d^2\vec{\boldsymbol{x}}(t)}{dt^2} \tag{2.1}$$

The results of a simulation, therefore, are the coordinates of each atom as a function of time, which represent a trajectory of the system. These simple equations of motion are the foundation of molecular dynamics simulations; however, understanding the intricacies of the simulations becomes much more complex. Here, the simulation methodology is largely discussed in terms of its application to the study of DNA with a specific focus on the various models and parameters used in this thesis.

Force fields

Due to the foundation of the technique relying on $F = ma$, perhaps the most obvious question surrounds what the force acting on the atoms is. Here, the potential energy is described as the combination of intramolecular and intermolecular interactions. This can generally be written as:

$$E_{total} = E_{bonded} + E_{non-bonded} \tag{2.2}$$

A force field is a combination of parameters that define these potentials for a given set of biomolecules. For example, it is common to have separate force fields to describe proteins, nucleic acids, and small molecules. The generic, functional form of the AMBER force field—often applied to simulations of DNA—combines bond length, angle, and torsional potentials as the bonded terms, with electrostatic and van der Waals non-bonded potentials (eq. 2.3).

$$\begin{aligned} V(r^N) = & \sum_{bonds} k_{bi}(l_i - l_i^0)^2 + \sum_{angles} k_{ai}(\theta_i - \theta_i^0)^2 \\ & + \sum_{torsions} \sum_{n} \frac{1}{2} V_i^n [1 + cos(n\omega_i - \gamma_i)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \end{aligned} \tag{2.3}$$

Other force fields, such as the CHARMM force field, include improper potentials to account for out-of-plane-bending, and cross-terms that can couple the different variables, such as angles and bond length. Below, we discuss each term in detail as applicable to the AMBER force fields.

Bonded terms

The simplest bonded term is that of the bond length interaction. This describes the vibrations of a pair of covalently bonded atoms and is Hookean in nature: the bonds are modelled as beads on a spring, and the force is proportional to the length of the spring. The bond potential (eq. 2.4)

depends on the spring constant, k_{ij}, as the bond length, l, oscillates about its equilibrium value, l_0. Notably, the term is binary in that the atoms are either bonded or not. Thus, conventional force fields are not able to model the breaking or forming of chemical bonds.

$$V_{bond} = k_{bond}(l - l_0)^2 \tag{2.4}$$

This potential is inherently one-dimensional. To define a two-dimensional potential, a bonded triplet may be considered. Here, the relative orientations of the electron orbitals in a triplet define an equilibrium bond angle, θ_0, between the three atoms. Thus, the potential for the oscillation of the angle, θ, can be described in the same way as the bond length (eq. 2.5).

$$V_{angle} = k_{angle}(\theta - \theta_0)^2 \tag{2.5}$$

Finally, the three-dimensional potential is dependent upon the dihedral angle. This is the angle between the two planes created by a set of four linearly bonded atoms that share a common central bond. This term is expressed as a Fourier series, as a single bond may have more than one term. The energy depends on the multiplicity, n, which is defined by the bond order, the phase angle, γ, the barrier height to torsional rotation, V_n, and the dihedral angle, ω (eq. 2.6).

$$V_{dihedral} = V_n[1 + cos(n\omega - \gamma)] \tag{2.6}$$

Non-bonded Terms

The nonbonded terms are the most computationally expensive part of the calculation. They describe the van der Waals and electrostatic interactions between atom pairs. The Lennard-Jones potential is arguably the most well-studied and widely used pair potential in the field of molecular dynamics [334] and describes the weak interactions between partial charges (eq. 2.7). Here, the depth of the potential well, ϵ, determines the strength of the interaction. The potential is dependent on the distance between the interacting particles, r, in relation to the value at which the potential is zero, σ.

$$V_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.7}$$

The idealised substances that interact exclusively through this potential are referred to as "Lennard-Jonesium". The potential describes the phenomenon of two interacting particles repelling one another at close distance—the positive term that approximates the strong Pauli exclusion principle—but attracting each other at a moderate distance—the negative term that arises from weaker van der Waals interactions. At a large distance, the particles stop interacting. The 12^{th} power is chosen to match experimental data, whilst the 6^{th} power matches London

dispersion forces [335]. Specifically, this is the 12,6-Lennard-Jones potential. However, the repulsive term is, in practice, too steep and often causes an overestimation of the pressure of the system.

To obtain the correct parameters when describing the interaction between two different types of atoms, combining rules are used. This has the benefit of avoiding huge parameter sets containing the parameters for each combination of atom type. The most common, and those used by the AMBER force field, are the Lorentz-Berthelot rules that combine the arithmetic average to calculate σ_{ij} (eq. 2.8) [336] and a geometric average to calculate ϵ_{ij} (eq. 2.9) [337]. The Lorentz rule assumes the atoms to be two hard spheres, and, therefore, σ is the effective radius between the two particles.

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{2.8}$$

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{\frac{1}{2}} \tag{2.9}$$

Finally, the Coulomb potential describes the electrostatic interactions between two atoms by assuming point charges at each position (eq. 2.10). It utilises the atomic charges, q, to describe the chemical bonding—ranging from covalent and polar covalent to ionic bonding—between two atoms i and j. It is distance, r, dependent and takes into account the permittivity of free space, ϵ_0. Together, the Coulomb and Lennard-Jones interactions can accurately describe hydrogen bonding [338] despite not explicitly including terms for this interaction.

$$V_{Coulomb} = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \tag{2.10}$$

The calculation of the non-bonded interactions is the most computationally demanding aspect of a simulation. The number of non-bonded interactions scales as the square of the number of atoms in the system; as such, it is impractical to calculate every non-bonded interaction. Therefore, a cutoff is employed whereby beyond this distance, the Lennard-Jones interactions are not calculated. This reduces computation time significantly whilst maintaining accuracy. The same cannot be done for Coulomb interactions, where the strength of the interaction falls off as $1/r$, and a hard cut-off will lead to significant inaccuracies [339]. Therefore, the speed-up of electrostatic calculations is usually afforded by the Particle Mesh Ewald procedure [340]. Here, the long-range interactions are estimated by a continuum model, whilst the sum of all pairwise Coulomb interactions is only used to calculate interactions at short range, within a cut-off radius. The long-range interactions are calculated using 3D grids in reciprocal Fourier space. Specifically, the charges are assigned to the grid using interpolation and the potential at the grid point is calculated by the inverse transformation. The speed of the algorithm scales as $N \cdot log(N)$ [340], where N is the number of particles, compared to N^2 when directly computing

for all particles in the system.

1-4 interactions are a special case of interaction scaling. Where connected atoms are separated by up to two bonds—this is the case for all 1-2 and 1-3 atom pairs—the bonded interactions are sufficient to describe these interactions. However, the 1-4 atom pairs are separated by three bonds, and here, both bonded—in the form of the dihedral angle—and non-bonded terms are required. At this length scale, the discussed non-bonded interactions are too strong and must be scaled down. The magnitude of the scaling factor is force field dependent. In the AMBER force field, by default, the van der Waals interactions are scaled down by a factor of 2, whilst the electrostatic interactions are scaled down by a factor of 1.2.

Restraints

Restraints, sometimes referred to as NMR restraints due to their historical role in the use of NMR structure refinement, can be applied to any of the geometries that make up the bonded terms of the AMBER potential. These restraints restrict the parameter of interest to oscillate only around pre-defined values. To do so, an energy function is used that takes the shape of a flat-welled parabola (fig. 2.1). Here, four values corresponding to the bond length, angle, or torsional angle are input in increasing magnitude. The target values are those within the range of r2 and r3, where all values are equally acceptable. The use of r1 and r4 values, rather than restricting to only oscillating around the desired values, is to avoid a sharp transition between zero and non-zero potentials and provide a continuous and differentiable potential, required for the numerical calculations. Therefore, the energy potentials in the r1-r2 and r3-r4 regions are quadratic in nature with respect to the difference between the geometric value and the target values, whilst below r1 and above r4, the energy penalty increases linearly as the geometric parameter moves further away from the target value.

We have used distance restraints in the initial relaxation of our minicircles to maintain the hydrogen bonding within the base pairs and avoid any premature melting.

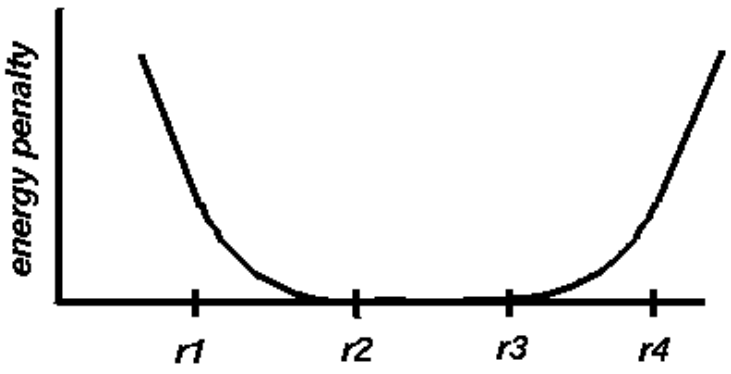

Figure 2.1: The energy function that describes the enforcement of NMR restraints. Values between r2 and r3 are equally acceptable. Figure taken from Amber MD [341].

Parameterisation and the bsc1 force field

The various constants—such as the equilibrium angles and lengths, and the potential well depths—are specific to the atoms involved in the interactions. Calculating these parameters, known as parameterisation, for a specific system is time-consuming and has traditionally relied on quantum mechanical calculations. More recently, machine learning techniques have emerged to speed up the process [342]. Nonetheless, it would be impractical to develop a generic force field capable of accurately modelling any given system. Thus, a force field is generally specific to a given type of biomolecule, eg. DNA, protein.

Force field development is crucial as technological advancements increase the typical length scales of simulations, and small errors in the force field can propagate to large deviations in the conformational dynamics at increased timescales. For example, in the late 1990s, the parm94 and parm99 force fields were developed to simulate DNA at a time when the 1 ns timescale was considered to be "state-of-the-art" [343]. As the timescales accessible to simulations increased, severe distortions of the DNA emerged that required reparameterization of the torsional term to create *parmbsc0*. This force field accurately modelled the structural and dynamic properties of DNA on the 10 ns timescale typically used at the time.

This thesis uses its successor: the *parmbsc1* force field [344], which improves upon the underestimation of twist, deviations in sugar puckering, and biases in the torsions seen in the previous *parmbsc0*. Thus, parameterisation efforts were focused on the backbone degrees of freedom; the non-bonded interactions remained unchanged due to well-reproducing hydration free energies and hydrogen bond stabilities. The parameterisation was done from high-level quantum mechanical calculations and refined from initial guesses of the gas phase fitted parameters.

The *parmbsc1* force field is capable of modelling sequence-dependent structural features—such as the unique properties of A-tracts—as well as the conformation of Hoogsteen-DNA and *syn* nucleotides [344]. It accurately captures the structural features of many canonical and non-canonical structures and, crucially, it also reproduces their associated mechanical properties. Its ability to reproduce the dielectric constant, stiffness parameters, persistence length, torsional, and stretching modules of DNA has specifically been validated. Finally, and most importantly for the work presented in this thesis, it was able to reproduce structural features in both relaxed and supercoiled minicircles. Here, in the relaxed topoisomers, there were no denatured regions and only one kink. In contrast, distortions were seen under negative supercoiling that agreed with experimental results.

Integration

Due to the second derivative relationship between acceleration and position, integration of the equations of motion is required to study the evolution of the system. The biomolecules typically being simulated using molecular dynamics contain a large number of atoms, on the scale of

thousands. For reference, the 339 bp DNA minicircles presented later in this thesis contain over 21,500 atoms. Thus, it is impossible to solve the equations of motion for these systems analytically; numerical methods are required.

The simplest integration method is the Euler algorithm. Here, starting from an initial structure of the system of interest where each particle has a known position, r, and velocity, v, at time, t, we can calculate the new velocity (eq. 2.11) and, from that, position (eq. 2.12), after some time, Δt, has passed. The acceleration is calculated via $F = ma$, and calculating this for each atom in the system over many timesteps builds up our trajectory.

$$v(t + \Delta t) = v(t) + a(t)\Delta t \tag{2.11}$$

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{2.12}$$

Whilst simple, this method is inappropriate for biomolecular simulation for multiple reasons. Firstly, it incorrectly assumes that the acceleration remains constant. This leads to the breaking of two key principles required of a molecular dynamics integration algorithm in that it does not conserve energy, nor is it reversible in time. Furthermore, whilst the method is accurate for a single time step, the accumulation of errors over time leads to large drifts in total energy. To overcome this, the original Verlet algorithm was developed (eq. 2.13) that uses the positions of the atoms over two successive time steps in order to account for changes in acceleration. However, requiring the position at two timesteps is inconvenient in the case of the initial starting of the simulation, where only the current positions are available. Furthermore, the velocities, which are useful for calculating observables such as kinetic energy, are only accessible once the next positions are calculated (eq. 2.14).

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \tag{2.13}$$

$$v(t + \Delta t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} \tag{2.14}$$

One solution to this, and the integration method employed by Amber, is the Velocity Verlet algorithm. It is mathematically equivalent to the previous Verlet algorithm, whilst solving the problem of the first timestep. It relies on the fact that acceleration is a function of atomic coordinates due to its derivation from the interaction potential. Here, the co-ordinates, r, velocities, v, and acceleration, a, at time t are used to calculate the updated co-ordinates at time $t + \Delta t$ (eq. 2.15). Secondly, the acceleration at the subsequent timestep is calculated using the interaction potential at the new positions, $r(t + \Delta t)$, to then calculate the updated velocities for use at the next time step (eq. 2.16). Explicitly including the acceleration at the two successive timesteps allows the algorithm to be both time-reversible and energy-conserving.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{2.15}$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}(a(t) + a(t + \Delta t))\Delta t \tag{2.16}$$

The Timestep

The choice of timestep is a trade-off between computational cost—as shorter timesteps require a greater number of calculations for the same total simulation length—and accuracy. The error accumulation in the Verlet methods scales as Δt^2; thus, the usual practice is to choose the largest value with which the system remains stable. This is dependent upon the biomolecule under investigation, as the timestep should be determined by the fastest motions within the system. DNA is an organic molecule; therefore, the fastest motion is the bond stretching involving a hydrogen atom. In practice, a timestep of 1 fs is recommended in order to reliably describe these motions; however, constraining hydrogen bonds, such as via the SHAKE algorithm, allows the timestep to be doubled without a significant compromise in energy conservation.

SHAKE

The SHAKE algorithm [345] is an iterative algorithm that sequentially constrains all of the required bonds to a fixed length and/or angle, rather than allowing them to oscillate, until a desired tolerance is reached. Common practice is to constrain all bonds involving hydrogen as the removal of these vibrations has a negligible effect on system dynamics, whilst significantly reducing the computational cost of the simulation. Whilst SHAKE was the first developed algorithm, and the algorithm employed by Amber, other constraint procedures are available, such as SETTLE [346] and LINCS [347].

Solvent Models

Biomolecules exist, *in vivo*, in an aqueous environment. Thus, to accurately simulate these biomolecules, their environment must also be accurately modelled. Whilst this thesis relies heavily on the use of implicit solvent models, its use is still limited by its accuracy and, currently, explicit solvent is most often the model of choice during molecular dynamics simulations.

Explicit Solvent Simulations

Explicit solvent models, as the name implies, use discrete atoms to explicitly include the individual water (or other solvent choice) and ions in the system. Here, water is used to fill the simulation box and accounts for over 80% of the particles within the simulation. Calculating these water-water interactions is the most computationally intensive part of a simulation, and, therefore, the chosen water model must be both accurate and efficient. To be considered accurate,

a good water model aims to reproduce a number of bulk properties: the static dielectric constant, the heat of vaporisation, the isobaric heat capacity, the thermal expansion coefficient, the isothermal compressibility, and the self-diffusion coefficient. In general, an explicit water model can be categorised by whether the model is rigid or flexible, the number of interaction sites, and whether the model includes polarisation effects.

Rigid models are the simplest model available as they only explicitly model the non-bonded interactions. The bonding interactions are implicitly treated using holonomic constraints: relationships between the atoms that are described only in terms of their coordinates, but neglect any of their derivatives, such as velocity or acceleration. Analogous to the force field-based modelling of the biomolecule, the electrostatic interactions are calculated using Coulomb's law, whilst the Lennard-Jones potential models the dispersion and repulsion forces (eq. 2.17) [348]. There is a distinction between the distance between the two charged sites, r_{ij}, used to calculate the Coulomb interaction, and the distance between the two oxygen atoms, r_{O-O}, used to calculate the Lennard-Jones interaction. The charged sites can be on either an atom or a dummy site, such as a lone pair, depending on the model chosen. However, most often, the key assumption is that only the oxygen atom contributes to the Lennard-Jones potential. Rigid models further include, explicitly, the electrostatic constant, k_C. A flexible model, conversely, would consider the anharmonic O-H stretching behaviour and would well describe the dynamical behaviour of the solvent [349].

$$E_{ab} = \sum_{i}^{ona} \sum_{j}^{onb} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{O-O}^{12}} - \frac{B}{r_{O-O}^{6}} \tag{2.17}$$

The number of charged sites within the water model can vary between two and six; however, the three-charge three-point models are most commonly used due to their computational efficiency and sufficient accuracy [350]. Here, the three interaction points correspond to the three atoms of the water molecule, and each site has a point charge. As detailed above, the oxygen carries the Lennard-Jones parameters. A number of different three-site models exist, with the distinction between them being the geometric parameters: namely, the O-H distance and H-O-H angle. Increasing the number of sites places dummy atoms near the oxygen to carry negative charge, simulating the lone pairs and improving the electrostatic distributions. A polarization correction can also improve solvent density and the diffusion constant by inducing small, massless charges that, when attached via harmonic springs, can create an induced dipole moment [351, 352]; or by directly calculating the induced dipole depending on the surrounding electrostatic field [353, 354]. These, therefore, no longer model the water by fixed-charge representations. The choice of water model largely depends on the properties under investigation.

Boundary Conditions

In order to simulate a biomolecule in its aqueous environment, it must be placed into a box of a sufficient size to allow it to adequately sample its conformational space. The remainder of the box is then filled with the solvent molecules. Periodic boundary conditions are used to avoid surface effects at the boundaries of this box and to allow a simulation of an infinite system, whilst explicitly simulating only a small part of that system. The boundary is required in order to control the temperature, pressure, and density of the system. However, the unit cell, i.e., the simulation box, is surrounded by an infinite number of images of that cell in all directions. The effect of this is that when a particle in the unit cell moves across the boundary, it re-enters the cell at the opposite side. Therefore, a particle is always interacting with its neighbours, even if the neighbour in question is on the other side of the simulation box.

Implicit Solvent Simulations

Implicit solvent models significantly speed up atomic simulation by representing the solvent as a continuous medium of specific properties, thus reducing the number of explicit particles in the system and the number of calculations required. A second speed-up results from the faster conformational sampling of the biomolecule in the frictionless environment [355–358]. Furthermore, there is no need to equilibrate the solvent, as is the case for explicit solvent models; there are no artefacts pertaining to the periodic boundary conditions; and cases involving solvent-exposed areas are more easily simulated without the need for a non-trivial, manual rearrangement of nearby solvent molecules [359]. In general, implicit solvent models omit the explicit solvent degrees of freedom by encompassing their thermodynamic properties into a solvation free energy, ΔG_{solv}. Thus, the only explicit degrees of freedom are within the solute.

Two types of implicit solvation models exist. The first is accessible surface models that define the free energy of solvation by its contributions from the accessible surface area of each atom i, ASA_i, and the solvation parameter of atom i, σ_i. The latter is based on the free energy of solvation of that atom per unit surface area (eq, 2.18) [360].

$$\Delta G_{solv} = \sum_i \sigma_i ASA_i \tag{2.18}$$

Alternatively, and currently, the most often chosen is the continuum electrostatic models that include only the enthalpic component of the free energy. The most commonly used model, and the one employed in this thesis, is the Generalised Born (GB) model. Here, the solvation free energy is the summed contributions from the electrostatic and non-electrostatic components (eq. 2.19). Sometimes, this is also referred to as polar and non-polar contributions.

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{non-el} \tag{2.19}$$

The non-electrostatic components describe the free energy of solvating a non-charged molecule—i.e., the partial charges of each atom have been set to zero—and it is the combined effect of the favourable van der Waals interactions between solute and solvent, and the unfavourable cost of breaking the solvent structure around the solute. In Amber, this is proportional to the total solvent accessible area, which is calculated analytically using the LCPO algorithm [361]. This uses the linear combination of terms describing the pairwise overlap of hard spheres.

The electrostatic component is the free energy when the charges are first removed in a vacuum and subsequently added back in the continuous solvent environment. The polar, or electrostatic, part dominates the solvation energy in water; therefore, most efforts concentrate on improving this term [362]. Calculation of this component was traditionally done by the Poisson-Boltzmann approach (eq. 2.20). Here, a differential equation is used to describe the distribution of the electric potential, ψ, in the direction normal to a charged surface. It considers the permittivity of the solvent, ε, and the mean concentration of each ion species, c_i^0.

$$\nabla^2\psi = -\frac{1}{\varepsilon}\sum_i c_i^0 q_i \exp\left(\frac{-q_i\psi(x,y,z)}{k_B T}\right) \tag{2.20}$$

The higher-order nature of the equation means that it is computationally intensive, as it must be solved each time the conformation of the molecule changes. In Amber, the analytical generalised Born (GB) approach is used instead (eq. 2.21). It provides a reasonable, computationally efficient estimate compared to the numerical solution of the Poisson-Boltzmann equation [359, 363, 364]. Within the Amber GB models, each solute atom is represented by a sphere of radius R_i and charge q_i. The charge is positioned at the centre of the atom, and its interior is uniformly filled with a material of dielectric constant 1. The solute is surrounded by a solvent of dielectric constant, ε, equal to 80 for water at 300 K.

$$\Delta G_{el} \approx -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)}\left(1 - \frac{exp[-\kappa f_{GB}]}{\varepsilon}\right) \tag{2.21}$$

The algorithm implicitly models monovalent salt due to the use of the Debye-Hückel screening parameter, κ, that is dependent on the ionic strength being modelled. S_{ionic} [364].

$$\kappa = \sqrt{\frac{8\pi S_{ionic}}{\varepsilon k_B T}} \tag{2.22}$$

The function $f_{GB}()$ is a smoothing function of the associated arguments that can take multiple forms [365, 366], although the common choice is given in eq. 2.23 [363]. It serves to interpolate between the extreme where the distance between two atoms, r_{ij}, tends to 0 and the atomic spheres merge into one, and the other extreme where r_{ij} tends to ∞ and the atoms can be treated as point charges obeying Coulomb's law [367].

$$f_{GB} = [r_{ij}^2 + R_i R_j \exp\Big(-r_{ij}^2/4R_i R_j\Big)]^{1/2} \tag{2.23}$$

The effective Born radius, R_i, of an atom describes how buried the atom is within a molecule. Therefore, for an isolated atom, it is equal to its van der Waals radius, ρ_i. Thus, by also assuming $\kappa = 0$ for pure water, the ΔG_{el} can be simplified (eq. 2.24).

$$\Delta G_{el} = \frac{q_i^2}{2\rho_i}\Big(1 - \frac{1}{\varepsilon}\Big) \tag{2.24}$$

However, for deeply buried atoms, $R_{ij} >> \rho_i$ and so the estimate that R_i is equal to the distance between the atom and the molecular surface is more appropriate. As the effective radii of each atom changes upon the conformational change of the biomolecule, it is easy to see how computational efficiency becomes critical here. To this end, Amber GB models use the Coulomb field approximation (eq.2.25) that integrates over the solute volume surrounding atom i to obtain the effective radii [367]. This approximation is calculated over the total volume, Ω, outside of atom i but inside the molecule. The reciprocal of this can be subtracted from the intrinsic radius of the atom in question to obtain the effective radius. Despite this approximation being well-known to over-estimate the effective radii [368, 369], it is simple and more computationally efficient than the alternatives and, therefore, is still widely used in MD simulations [362].

$$I_i = \frac{1}{4\pi}\int_{\Omega, r>\rho_i} \frac{1}{r^4} d^3r \tag{2.25}$$

Solving the solute boundary is non-trivial, and further approximations are required. The aim of an implicit solvent model is to reproduce high theory solvation free energies, typically benchmarked against Poisson-Boltzmann reference calculations [362] whilst minimising the computational expense. Thus, the handling of the solute boundary typically differentiates between the different implicit solvent models. The GB-neck model employed in this thesis (eq. 2.26) uses a "neck" correction [370] that overall brings the space defined by the van der Waals boundary (eq. 2.25, $I_i = I_{vdw}$) closer to that defined by the computationally-demanding molecular surface (MS) boundary calculations (eq. 2.26).

$$I_{MS} = I_{vdw} + \int_{neck} \frac{1}{r^4} d^3r \tag{2.26}$$

The effective radius can then be calculated using the MS boundary and a parameter set consisting of α, β and γ (eq. 2.27) where $\tilde{\rho_i}^{-1} = \rho_i -$ offset and $\psi = \tilde{\rho_i} I_{MS}$. The offset parameter serves to minimise the error between the Generalised Born model and experimentally-determined solvation energies [361].

$$R_i^{-1} = \widetilde{\rho}_i^{-1} - \rho^{-1} \tanh\Big(\alpha\psi - \beta\psi^2 + \gamma\psi^3\Big) \tag{2.27}$$

Thermostats and Barostats

The final parameters necessary to accurately model a biological system are the regulation of temperature and pressure. Akin to experimental counterparts, simulations are typically run in the NVT or NPT ensembles, named due to the parameters that are fixed during the simulation: the number of particles (N), the volume (V) or pressure (P), and the temperature (T).

Thermostats

Temperature control algorithms, in practice, rely on the relationship between kinetic energy, E_k, equilibrium temperature, T, and the number of degrees of freedom, N, (eq. 2.28). Whilst the total energy of the system must be conserved within a simulation, the kinetic energy is allowed to fluctuate as it converts between potential energy and back. Thus, thermostats maintain a constant temperature by adjusting the velocities of a subset of particles.

$$E_k = \frac{1}{2} N k_B T \tag{2.28}$$

There are four classifications of methods that maintain temperature within a simulation: strong coupling, weak coupling, stochastic methods, and extended system dynamics. The Berendsen thermostat [371] is one example of a weak-coupling algorithm (eq. 2.29). In this scheme, the system is weakly coupled to a heat bath of desired temperature, T_0. The effective temperature, T, can be calculated using the relationship to kinetic energy. The algorithm works to ensure that the fluctuations in temperature, as a result of kinetic energy fluctuations, exponentially decay with respect to a time constant, τ, and are, therefore, suppressed.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.29}$$

The time constant depends upon the total heat capacity of the system, C_V, a temperature coupling time constant, τ_T, the total number of degrees of freedom, N_{df}, and the Boltzmann constant, k (eq. 2.30). The fact that $\tau \neq \tau_T$ stems from the kinetic energy change as a result of velocity rescaling being distributed between both the kinetic and potential energies, and, therefore, the change in temperature being less than the scaling of the energy. The τ/τ_T ratio for water is 3.

$$\tau = \frac{2 C_V \tau_T}{N_{df} k} \tag{2.30}$$

Specifically, in order to suppress the changes in kinetic energy, the velocities are adjusted by a

scaling factor, λ, at every n_{TC} steps (eq. 2.31).

$$\lambda = \left[1 + \frac{n_{TC}\Delta t}{\tau_T}\left\{\frac{T_0}{T(t - \frac{1}{2}\Delta t)} - 1\right\}\right]^{1/2} \tag{2.31}$$

The change in kinetic energy as a result of the thermostat has a simple, quadratic relationship to the scaling factor (eq. 2.32).

$$\Delta E_k = (\lambda - 1)^2 E_k \tag{2.32}$$

The rescaling of velocities and suppression of kinetic energy fluctuations does not generate a proper canonical ensemble, as it cannot ensure that the temperature is consistent over all parts of the molecule. However, for large systems with numerous collisions, the temperature distribution improves, and the results converge, and the approximation yields correct results [372]. In practice, most ensemble averages will be unaffected by this, except for the distribution of kinetic energy itself. Furthermore, it is highly computationally efficient.

However, in implicit solvent simulations where the lack of discrete solvent molecules limits the temperature distribution under the Berendsen thermostat, this algorithm would be highly inappropriate. Here, the Langevin thermostat [373, 374] is a suitable stochastic thermostat method. All of the methods within this category use a Maxwell-Boltzmann distribution at the target temperature to assign new velocities to a subset of atoms. The Langevin equation of motion describes the total force acting on a solute by an implicit solvent (eq. 2.33). It contains a random force term—the last term of the equation—that combines the friction coefficient, γ, the Boltzmann constant, k_B, and the target temperature, T. This term describes the cases where a subset of solute-solvent collisions will be sufficient to cause a perturbation of the motion of the particles. Within the equation, there is also the friction term, γmv, and the interaction potential of particle X, $\nabla U(\mathbf{X})$. The friction term is derived from the force that would act on a particle during movement as a result of the velocity, v, and mass, m, of the particle in a solvent of viscosity γ. The frictional and random force terms are responsible for giving the correct canonical ensemble.

$$F = -\nabla U(\mathbf{X}) - \gamma mv + \sqrt{2\gamma k_B T}\mathbf{R}(t) \tag{2.33}$$

In real systems and in explicit solvent, collisions are responsible for the transfer of kinetic energy. In an implicit solvent model, collisions between solute and solvent are not possible. Furthermore, the generalised Born model only includes the effects of electrostatic screening and neglects this collision factor. Thus, Langevin dynamics explicitly include this with a collision frequency term, γ, that also represents the viscosity of the solvent.

Barostats

The pressure of a system is kept constant by scaling the volume of the unit cell at each time step. This is because, essentially, pressure is a force that results from the collision of solvent particles at the simulation boundary. Thus, the virial equation can be used to calculate the pressure (eq. 2.34). Here, the first term describes the pressure of an ideal gas where there is no interaction between the molecules, and the second term describes the force, F_{ij}, between particles i and j and the vector, r_{ij}, that connects the two.

$$P = \frac{NK_BT}{V} + \frac{1}{3V}\langle \sum r_{ij}F_{ij} \rangle \tag{2.34}$$

Analogous to the temperature control previously discussed, the pressure of the explicit solvent simulations can also be controlled using weak coupling methods. Here, the Berendsen barostat is conceptually similar to the corresponding thermostat (eq. 2.36) [371].

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \tag{2.35}$$

The scaling matrix, μ, depends upon the isothermal compressibility of the system, β, and the Kronecker delta, δ (eq. 2.36). In the same way as the thermostat, the scaling occurs every n_{PC} steps.

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_P}\beta_{ij}\{P_{0ij} - P_{ij}(t)\} \tag{2.36}$$

Overall, the Berendsen barostat achieves the correct target density; however, in practice, it does not strictly sample from the NPT ensemble. The only other barostat available in Amber, and which does sample from the isobaric-isothermal ensemble, is the Monte Carlo barostat [375, 376]. This also has the advantage that the virial does not need to be computed. It functions by allowing the simulation to continue for a certain number of time steps, then, after a user-defined interval, the volume is changed at random and the potential energy is evaluated. The trial volume is chosen due to some random number scaling such that $\Delta V = n_{rand}\Delta V_{max}$. The limit, V_{max}, is such that there is a typical Monte Carlo acceptance ratio of approximately 40-50%.

2.2 Single-molecule FRET

Single-molecule Förster Resonance Energy Transfer (smFRET) is a biophysical technique that is often described as a "molecular ruler" due to its ability to measure distances between 2.5 and 10 nm [377]. Here, the biomolecule of interest is labelled with two dyes, a donor and an acceptor. Upon excitation of the donor, a non-radiative energy transfer process may occur to the acceptor when the fluorophores are within the aforementioned distance range of one another.

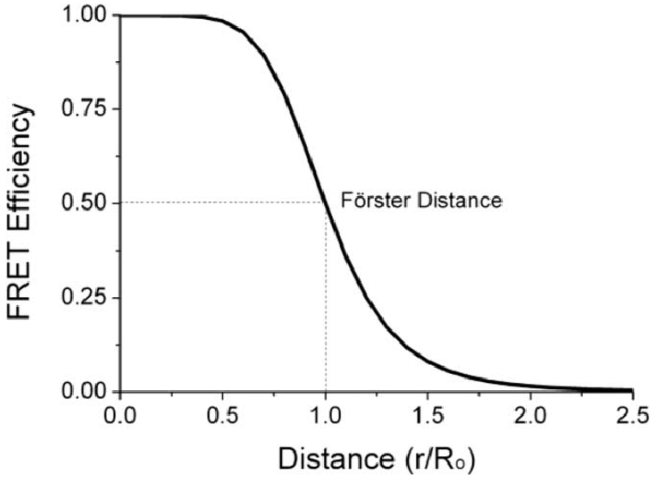

Figure 2.2: There is a sigmoidal relationship between the FRET efficiency and the distance between the dyes. Figure taken from reference [378].

As indicated by its name, this process occurs through electronic resonance. The acceptor relaxes via fluorescence, which can be measured to obtain the efficiency, E, of the energy transfer process (eq. 2.37). The FRET efficiency describes the quantum yield of the energy transfer or, in other words, the probability of FRET occurrence at each donor excitation event.

$$E = \frac{k_{ET}}{k_{ET} + k_f + k_i} \tag{2.37}$$

Where k_{ET} is the energy transfer rate, k_f is the radiative decay rate of the donor, and k_i is the rate of non-fluorescence de-excitation events in the absence of an acceptor fluorophore. The sigmoidal relationship (fig. 2.2) between the FRET efficiency and the distance between the fluorophores (eq. 2.38) allows the technique to follow the conformational dynamics of the system.

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{2.38}$$

Where R is the distance between the dye pair in the system being studied, and R_0 is the Förster radius. The latter is the distance at which the FRET efficiency is 50% and is constant for a given dye pair. It can be calculated using the quantum yield of the acceptor, ϕ_D, the dipole orientation factor, κ^2, the refractive index of the medium, n, and the overlap integral (eq. 2.39) [379]. The overlap integral refers to the degree of overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor. Therefore, a high degree of overlap is required to measure an experimentally appropriate FRET efficiency.

$${R_0}^6 = \frac{9\ ln10\ \phi_D\ \kappa^2}{128\pi^2\ N_A\ n^4} \int_0^\infty f_D(\lambda) E_A(\lambda) \lambda^4 d\lambda \tag{2.39}$$

The dipole orientation factor, κ^2, is a measure of the alignment of the acceptor excitation dipole with the electric field of the donor emission dipole. It depends upon the angle between the dipole moments of the donor and acceptor, θ_{DA}, and the angles between the separation vector of the two fluorophores and the dipole moments of the donor and acceptor, θ_D and θ_A, respectively. One of the major assumptions of smFRET is that the two dyes are freely rotating and so $\kappa^2 = \frac{2}{3}$; however, in reality, it can have any value between 0 and 4 (eq. 2.40).

$$\kappa^2 = cos(\theta_{DA} - 3cos\theta_D cos\theta_A)^2 \tag{2.40}$$

The approximation $\kappa^2 = \frac{2}{3}$ stems from the assumptions that the orientations of the fluorophores are able to fluctuate rapidly over the experimental integration time, and that the FRET event occurs between two static points. Thus, the FRET rate is much slower than the rotation of the dyes, yet much faster than the translational diffusion rate (eq. 2.41). This allows the dyes to sample a range of conformations within the experimental integration time such that their whole accessible volume — the region in which the dyes may be found due to their attachment to the biomolecule via flexible linkers—is sampled.

$$k_{rot} \gg k_{FRET} \gg k_{diff} \gg k_{int} \tag{2.41}$$

Practically, the method of labelling the biomolecules with the dyes is not 100% efficient. Thus, it may be the case that some molecules are only labelled with a donor—i.e. are "donor-only"—or an acceptor, which would be "acceptor-only". Corrections must be made in order to distinguish between donor-only molecules and low FRET species. This is done by the alternation of two lasers of different wavelengths that separately excite the donor and the acceptor. This defines the stoichiometry ratio, S, that is dependent on the total signal, Λ, and the fluorescence of the acceptor when directly excited by a laser, $F_{(A_{ex}A_{Em})}$ (equation 2.42).

$$S = \frac{\Lambda}{\Lambda + F_{(A_{ex}A_{Em})}} \tag{2.42}$$

$F_{(A_{ex}A_{Em})}$ is equal to 0 when the species is donor-only, and S would, therefore, be equal to 1. In the case of an acceptor-only species, S = 0 as the reverse is true. Overall, we can follow the conformational dynamics of a biomolecule via the movement of the two labelling positions with respect to one another over time. This allows for the construction of an ES plot, or ES histogram, that depicts both the FRET efficiency and the stoichiometry (figure 2.3).

Accessible Volume Modelling

The fluorophores are linked to the molecule of interest by a long flexible linker of either two or six carbons (fig. 2.4); this allows for a significant range of movement around the point of attachment.

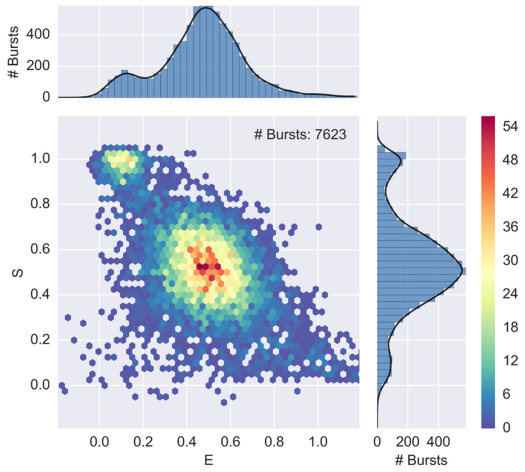

Figure 2.3: An example ES plot. Donor-only species lie at a high stoichiometry, whilst acceptor-only species lie at a low stoichiometry. The doubly labelled samples lie at the mid values. From the histogram, we can see that the majority of the sample is doubly labelled and that there are two main conformations of the biomolecule. A smaller population at lower FRET indicates the two dyes are further apart, and a more dominant population at higher FRET where the two dyes are further together. E: [FRET] Efficiency; S: Stoichiometry. Figure taken from reference [378]

Therefore, there is an inherent uncertainty in the exact distances between the two chromophore centres at the instant of energy transfer [380]. This is an important consideration within the context of the experiment as the distance between the fluorophores and the orientation of the two transition dipole moments are determining factors of the energy transfer efficiency [381–384]. Furthermore, the actual donor-acceptor distance distribution as a result of the dye linkers causes the mean dye position to be relatively distant from the dye attachment point on the biomolecule (fig. 2.5) [385]. Finally, the accessible space is large, resulting in many potential donor-acceptor distances. The distance between the mean position of the dye and that calculated from the experimental FRET efficiency may differ by up to 10 Å or 30% [385]. Thus, especially when the mean distance is below the R_0, the distance distribution must be carefully considered for quantitative FRET measurements.

In order to calculate the expected FRET efficiency, the spatial distribution of the fluorophores

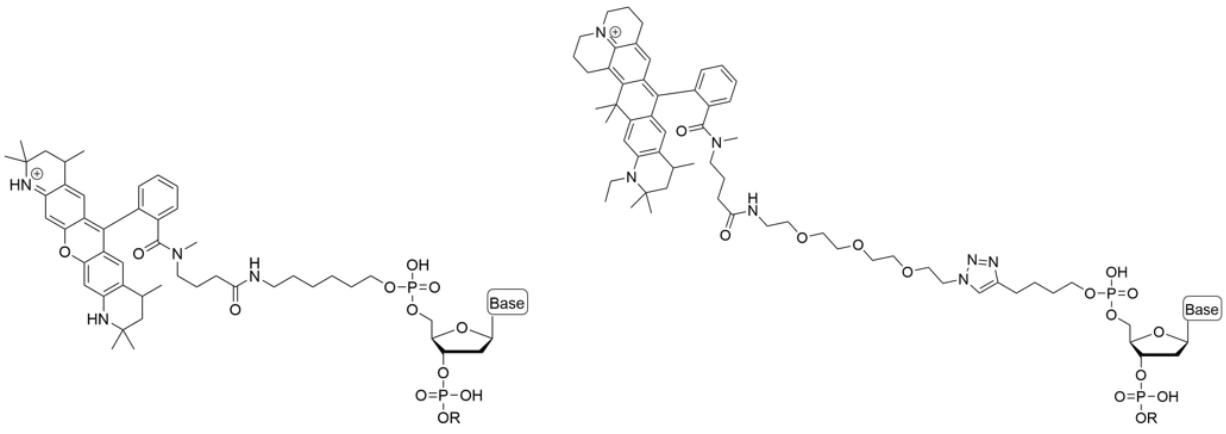

Figure 2.4: The chemical structures of Atto550 (**left**) and Atto647N (**right**) with their attachment point to the DNA shown. Normally, the DNA is labelled on thymines.

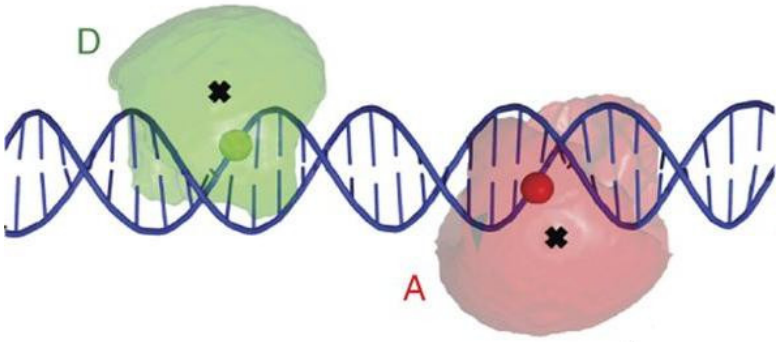

Figure 2.5: The mean distance between the two dyes is relatively far from the labelling positions. Figure taken from reference [385].

must be known. Molecular dynamics simulations are one such method of doing so; however, simulating the dye pair on every biomolecule of interest is unfeasible and time-consuming. Accessible volume (AV) modelling is one method of minimising the effect of these limitations. Here, the sterically accessible volume of the dye on the macromolecule is modelled to calculate the mean position of the dye and create an 'AV cloud' that describes the volume that is able to be sampled by the fluorophores (figure 2.6) [380]. In order to utilise this method, the local environment of the fluorophore must be well-defined [380]: for example, a double-stranded segment of DNA or RNA. This is due to the steric restriction the biomolecule has upon the accessible volume due to the microenvironment it creates. The main purpose of this method is experimental design. To efficiently follow the conformational dynamics of a biomolecule, two labelling positions must be chosen that are both informative due to the change in distance between them at the labelling positions, but that also elicit measurable changes in FRET efficiency. The latter must account for the R_0 of the selected dye pair—which typically ranges 50-70 Å [379]—and the sensitivity of the technique. Generally, distances measured across $0.6R_0$ and $1.6R_0$ are sufficiently accurate [379].

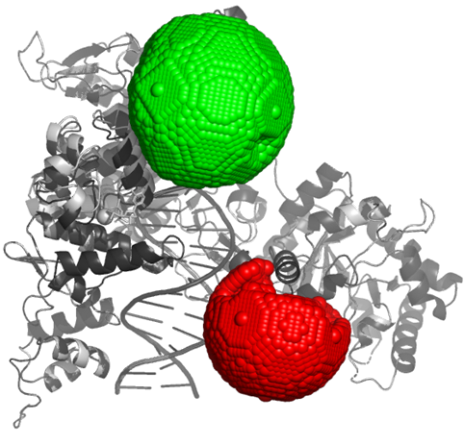

Figure 2.6: Two AV clouds of a donor (green) and acceptor (red) fluorophore shown on a DNA-protein complex (grey). The clouds represent the volumes in which the fluorophores can sample when labelled at two positions on the protein.

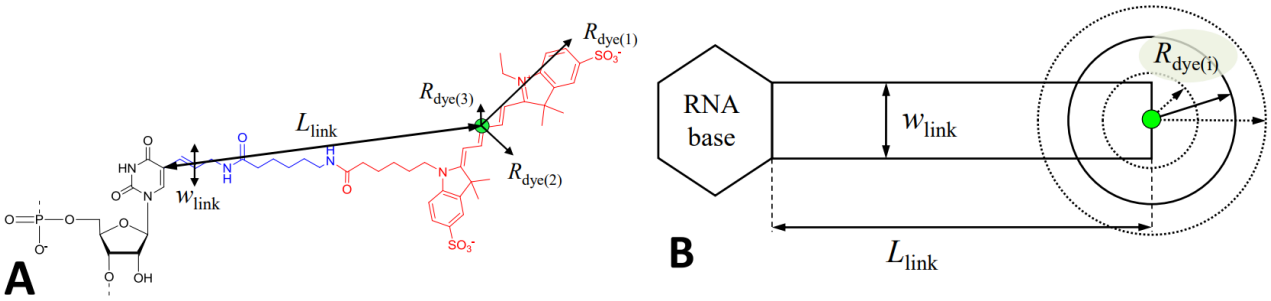

Figure 2.7: The Accessible Volume model functions by approximating the dye and linker as a sphere attached to the biomolecule by a long linker. The dimensions of the model are the three radii of the sphere, describing the asymmetry of the fluorophore, and the length and width of the linker. Figure taken from reference [385].

Olga

Olga is a program within the FRET positioning system (FPS 2.0) suite of tools used for AV modelling and experiment planning [380, 385, 386]. Here, the dye is modelled as a sphere of three radii $R_{dye(1-3)}$, with the central atom connected by a linker of length L_{link} and width w_{link} (fig. 2.7). A geometric search algorithm can then find all possible dye positions that do not cause steric clashes with the biomolecule and create the AV cloud. All allowable positions within the cloud are considered to be equally probable, and thus, the method is not applicable to cases with strong inter-dye interactions, such as dye sticking. Overall, the method closely agrees with MD simulations with deviations as small as 2 Å seen between the mean dye position over the simulation and that predicted by the AV modelling [380].

2.3 Predictive Algorithms

The work in this thesis would not be possible without the use of numerous external analysis methods and algorithms. Here, I briefly describe the methodology of the two predictive algorithms most crucial to this work: the SIDD model [113] for base pair opening prediction, and the Dekker model [41] for predicting the position of plectonemes.

SIDD

Later in this work, we use the Twist-DNA implementation [387, 388] of the Benham SIDD model [113] to predict bubble opening probabilities. The Twist-DNA implementation is a self-consistent linearization [389] of the model that is numerically efficient and allows the investigation of bubble statistics on the genome length scale. Overall, the model combines the thermodynamics of base pair interactions with the energetics of torsional strain [109, 390]. The contribution from writhe is omitted to describe the linking difference, α, by considering three different contributions to the overall twist (eq. 2.43). Firstly, there is a twist absorption equal to n_o/A, by the n_o number of denatured base pairs, where A = 10.4bp/turn, equal to the number of base pairs in a helical turn. Secondly, the high flexibility of the single-stranded denatured regions allows the

individual strands to intertwine to induce a global overtwist equal to $\mathcal{T}$. Finally, the residual linking number, α_r, is the traditional bending and twisting of the double-stranded regions.

$$\alpha = -\frac{n_o}{A} + \mathcal{T} + \alpha_r \tag{2.43}$$

The total chemical energy required to denature the base pairs, H_c, depends on both the initial energy to nucleate the bubble, a, the number of bubbles, r, and the energy required to denature each base pair, b_j (eq. 2.44). The variable n_j describes the state of each base pair j, where $n_j = 1$ when the base pair is denatured, and $n_j = 0$ when the base pair is bonded.

$$H_c = ar + \sum_{j=1}^{N} b_j n_j \tag{2.44}$$

Within the Benham model, a=10.8 kcal/mol, b_{AT}=0.26 kcal/mol, and b_{GC}=1.31 kcal/mol.

The energy associated with the helical twist within the denatured region, $H(\tau_j)$ at base pair j is a contribution from the torsional stiffness, C, and the helical twist τ_j (eq. 2.45). The torsional stiffness, $C \approx 3.09 k_B T$ is known from experiments.

$$H(\tau_j) = \frac{C n_j \tau_j^2}{2} \tag{2.45}$$

The total twist of the open regions, $\mathcal{T}$, is calculated by allowing the helical twist of each denatured base pair to independently fluctuate (eq. 2.46).

$$\mathcal{T} = \sum_{j=1}^{N} \frac{n_j \tau_j}{2\pi} \tag{2.46}$$

Finally, the energy of the residual twist, H_r, depends on an experimentally determined coefficient, K, which is inversely proportional to molecular length, N, where $K \approx 2220RT/N$.

$$H_r = \frac{K\alpha_r^2}{2} = \frac{K}{2}\left(\alpha + \frac{n}{A} - \mathcal{T}\right)^2 \tag{2.47}$$

Thus, the Hamiltonian for the case where each of the n separated base pairs is torsionally deformed at a rate of τ_j rad/bp is given by eq. 2.48.

$$H = \sum_{j=1}^{N} \frac{C n_j \tau_j^2}{2} + \frac{K}{2}\left(\alpha + \frac{n}{A} - \sum_{j=1}^{N} \frac{n_j \tau_j}{2\pi}\right)^2 + ar + \sum_{j=1}^{N} b_j n_j \tag{2.48}$$

The calculation of the partition function proceeds as usual by the summation and integration over the Boltzmann factor $e^{-\beta H}$, where $\beta = 1/(k_B T)$, for all possible states, each with a

Hamiltonian, H.

The self-consistent linearization of this model [387, 388] allows for the precise and more numerically efficient calculation of the total energy of a given DNA configuration of N base pairs at a given superhelical density, σ. It combines the torsional stress energetics of the Benham model with the nucleation penalty to open a bubble, Δg_{loop}, and the sequence-, temperature-, and salt-dependent nearest neighbour free energies, Δg_i within the base pair *(i,i+1)* (eq 2.49).

$$H_{tot} = \left(\frac{2\pi^2 CK}{4\pi^2 C + Kn_o} \left[\frac{\sigma N}{A} + \frac{n_o}{A} \right]^2 - \frac{k_B T}{2} log \left[\frac{4\pi^2 C}{4\pi^2 C + Kn_o} \right] \right) + \left(\sum_i \Delta g_i \theta_i \theta_{i+1} + \Delta g_{loop} \theta_i (1 - \theta_{i+1}) \right) \tag{2.49}$$

Where $\Delta g_{loop} = 20 k_B T$, and the temperature-dependence of Δg is the linear $\Delta g = \Delta h - T\Delta s$. The enthalpic, Δh, and entropic, Δs, contributions are those of the SantaLucia nearest neighbours [114].

Dekker plectoneme prediction

The Dekker algorithm [41] is, thus far, the only method of predicting plectoneme density within a stretch of DNA. It was developed to be consistent with the corresponding ISD experiments and so predicts the plectoneme density within linear DNA of length scale on the order of kilo-base pairs, and DNA with constrained ends. The plectoneme densities were further smoothed using a Gaussian filter to approximate the resulting curves to the spatial resolution of the experimental data.

Thus, the code was modified to remove the smoothing due to a much shorter length scale (339 bp) being investigated, to remove the length of DNA bound to the surface, such that plectonemes could form at the two extremes, and to reduce the average plectoneme length from 1000 to 20 nt. Different plectoneme lengths were tested for the modified code, but the resulting plectoneme densities were unchanged between them all; however, as you will see below, it increased the search space over the calculations. Due to the much smaller DNA under investigation, the speed and efficiency of the calculations were not an issue, and so loop sizes were calculated starting from 1 nt instead of 40 nt and increased in 8 nt increments. Overall, these changes had the effect of increasing the resolution in the much shorter DNA sequence but without changing the plectoneme positioning from that predicted by the original algorithm. Notably, the equations discussed below neglect any consideration towards superhelical density. This results in a simplified model that captures the key considerations of intrinsic curvature and flexibility on plectoneme positioning; however, it neglects the influence of key interactions. One notable omission is the formation of bubbles. This was likely not considered due to the model being developed for comparison to experiments conducted under positive supercoiling, where bubble formation is less prevalent, and for simplicity. However, for a full and accurate model, one must

consider bubble formation, plectoneme sizes, and the interactions between multiple plectonemes. Nevertheless, the model has a sufficient agreement with the current literature.

Plectoneme tip-loop size

The first component of the model is the determination of the energy required to bend the DNA at the plectoneme tip. By considering a circular, 360° loop formed under tension, the work required is dependent upon the base pair rise, $r = 0.334$ nm, and the tension across the DNA, F (eq. 2.50). This expression can be used to estimate the mean size of a plectoneme tip-loop, N.

$$W = rFN \tag{2.50}$$

The bending energy further depends on the bulk persistence length, A (eq. 2.51).

$$E_{bend} = \frac{2\pi^2 k_B T A}{rN} \tag{2.51}$$

Thus, both contributions can be combined to obtain an expression for the total energy, E_{total} (eq. 2.52).

$$E_{total} = rFN + \frac{2\pi^2 k_B T A}{rN} = k_B T(CN + \frac{B_{360}}{N}) \tag{2.52}$$

By taking the derivative of E_{total}, you can obtain an expression for N (eq. 2.53).

$$\begin{aligned} N &= \sqrt{\frac{B_{360}}{C}} \\ C &= \frac{F}{12.16pN} \\ B_{360} &= 2955 \end{aligned} \tag{2.53}$$

Thus, at 3 pN:

$$N = \sqrt{\frac{B_{360}}{C}} = 109nt \tag{2.54}$$

However, within a plectoneme, the length is held at the same value but only bent to form a partial circle. Thus, the work needed to accommodate the loop remains the same, but the overall bending energy is lower. Here, a loop of 240° is sufficient.

$$B_{240} = B_{360}\left(\frac{240^\circ}{340^\circ}\right)^2$$
$$N = \sqrt{\frac{B_{240}}{C}} = 73nt \tag{2.55}$$

However, the $3pN$ force assumed in the model is significantly larger than physiological conditions [391] and assumes that all plectonemes are an average of 1000 bp in size.

Predicting plectoneme density

The model itself predicts plectoneme density based on intrinsic curvature and flexibility. The model is entirely based on the hypothesis that as the plectoneme tip is sharply bent, those stretches of DNA with relatively high local curvature and flexibility will lower the energy barrier to plectoneme formation. Thus, the first step of the model is to use previously computed dinucleotide parameters [392] to calculate the winding ground state path of the DNA, determined by its sequence. The intrinsic curvature, $\theta(N, i)$, across N nucleotides centred at i can then be calculated using tangent vectors over a window of one helical turn (11 bp, ≈ 3.7 nm). The intrinsic curvature defines the preferred bend direction, $\phi = \phi_B$. The flexibility can be calculated by summing the roll and tilt covariance matrices for the stretch of nucleotides using values previously determined from molecular dynamics simulations [392]. The summed covariance is an estimate of local persistence length, $A(N, I, \phi)$. Thus, the bending energy can be described in a sequence-dependent manner (eq. 2.56)

$$\frac{E_{bend}(N, i, \theta)}{k_B T} = \left(\frac{2}{3}\right)^2 \frac{2\pi^2 A(N, i, \phi)}{0.334nm \cdot N}\left(1 + \left[\frac{\theta(N, i)}{240^\circ}\right]^2 - 2\left[\frac{\theta(N, i)}{240^\circ}\right] cos(\phi - \phi_B)\right) \tag{2.56}$$

Thus, due to the dependence of E_{bend} on sequence, the loop size will also be sequence dependent. In the interest of efficiency, rather than calculate the loop size at each position along the sequence, the relative probabilities of loops of a range of sizes can instead be calculated (2.57).

$$\frac{E_{total}(N, i, \phi)}{k_B T} = \frac{rFN}{k_B T} + \frac{E_{bend}(N, i, \phi)}{k_B T} \tag{2.57}$$

Each bending conformation can be assigned a Boltzmann weight, and the summation of all weights can inform the total weight of plectoneme formation at each position within the DNA. As the direction, ϕ, can span any vector normal to the path of the DNA, and the loop size can have a large variation, this results in a significant number of potential bending conformations. However, the maximum likelihood should occur at $N = 73$ and $\phi = \phi_B$; thus, the parameter values are only considered in the final model around this phase space.

3

Validating Implicit Solvent Simulations

In this chapter, I expand upon the present literature of sequence-dependent plectoneme pinning and present a molecular dynamics simulation methodology in implicit solvent. The method is validated, for the first time, against experimental results from existing literature along with novel smFRET experiments to show that all-atom implicit solvent simulations accurately and quantitatively capture sequence-dependent plectoneme pinning. I present insights into how bubble formation influences plectoneme density, and evaluate the effects of salt and temperature on bubble formation with validation against the trends reported in the literature.

The following people have contributed to this chapter:

- **Victoria Hill:** Experiment and construct design, MD simulations and analysis
- **Sophie Fountain:** All smFRET experiments and analysis of these
- **Quentin Smith**: Made the minicircle smFRET constructs

The publication arising from the work in this chapter is:

- **Victoria E. Hill**, Sophie E. Fountain, Quentin M. Smith, David S. Rueda, Agnes Noy, and Timothy D. Craggs. "Implicit Solvent Simulations Capture Sequence-Dependent Plectoneme Pinning". 2025. *Manuscript submitted for publication.*

3.1 Introduction

Minicircles are excellent models for studying DNA supercoiling. They are small enough to be studied using all-atom molecular dynamics simulations [40, 393, 394], yet still give good resolution when studied experimentally using methods such as AFM [40, 395], FRET [396] and smFRET [397], gel electrophoresis [395, 398], surface plasmon resonance (SPR) [40], cryo-EM [399], and electron cryo-tomography [101]. They have successfully been studied using coarse-grained simulations [97, 323]; however, in order to take advantage of theoretical methods at the atomistic level, a more computationally challenging approach must be taken [393].

Due to their size, minicircles larger than the DNA persistence length of 150 bp must first be simulated in implicit solvent [393]. This is necessary in order to speed up the relaxation from a planar starting structure to the writhed structures of interest, which occurs approximately 10-fold faster in the frictionless GB/SA implicit solvent model [101]. Improvements in the model's overlap screening and offset parameters have lessened the trade-off between computational efficiency and the inherent inaccuracies due to the neglect of explicit solvent atoms and improved the accuracy of interaction strengths [400]. The model is a continuum representation of solvation effects that is able to speed up conformational changes, compared to explicit solvents, due to the low viscosity of the medium. Electrostatic (Debye-Hückel) screening effects can be tuned to account for salt effects [364]. Overall, this proved a better agreement with both the explicit TIP3P water model and experimental data compared to earlier models [400]. Despite these advances, artifacts may still arise in the simulations. In the case of supercoiled DNA, implicit solvent does not sufficiently screen repulsion within the backbone which results in less compact writhed structures and narrower minor grooves, especially at the plectoneme tips [393]. The duplex is also less stable and may be more prone to the formation of defects and denaturation bubbles as a result.

As previously mentioned, implicit solvent simulations function by approximating the solvent as a continuum and omits the discrete water and ion particles seen in explicit solvent simulations. This has the advantage that the interactions between solvent molecules do not need to be calculated. Explicit solvent simulations are limited in practice to duplexes 70-100 bp in length as the size of the simulation box, and therefore the total number of particles to be simulated, scales approximately with the second power of solute particles [401] or third power of duplex length [324]. The calculation speed for the same length of DNA in implicit solvent is increased due to the omission of discrete water and ion particles and overall scales between approximately $nlog(n)$ and n^2, where n is the number of solute atoms [401], depending on the model and simulation parameters chosen. Furthermore, the lack of explicit particles increases the conformational sampling over the same timescale, due to the absence of solvent friction, by approximately 10-fold in the case of minicircles of approximately 300 bp [101]. But the scale of the speed-up is dependent on the scale of the system and the magnitude of conformational change under investigation, with estimates of up to ∼50-fold seen for nucleosome tail collapse and DNA

unwrapping [358]. In general, the larger the system and/or the larger the conformational change under investigation, the larger the speed-up. However, this is both system dependent as well as dependent upon the implicit solvent model chosen. This speed-up helps avoid issues of statistical error due to inadequate sampling of the conformational space. In the case of simulating supercoiled minicircles, the speed-up afforded by implicit solvent is paramount as the DNA stays "locked" into one conformation on the timescale accessible to explicit MD [393]. Even high-end supercomputers such as ARCHER were only able to produce around 5 ns/day of explicit solvent simulation of a 336 bp minicircle [393].

In the case of protein folding, it has previously been shown that different conformational landscapes are explored in implicit compared to explicit solvent [402]. This arose from overly strong salt-bridge effects which can be mitigated against by removing the non-bonded cutoff. This also results in a more accurate reproduction of molecular surfaces and solvation forces [118]. However, the conformational landscape of supercoiled minicircles explored in implicit solvent has been shown to be in agreement with the structures seen experimentally using AFM [40]. This is aided by improvements in force field parameters, such as in the parmbsc1 force field, which is able to correct the underestimation of DNA twist, excessive terminal fraying, and the challenges with simulating non-canonical DNA conformations often seen in previous force fields [344]. Crucially, DNA simulated with this forcefield exhibited sequence-dependent variations in persistence length that agreed with experimentally determined values. A small (100 bp) minicircle was simulated in explicit solvent and used as a benchmark to show that the secondary structure remains in-tact when relaxed, but significant denaturation occurred upon negative supercoiling. Furthermore, bsc1 performed the best in a comparison of modern DNA force fields when evaluating their performance in capturing sequence-dependent elasticity, flexibility, and other mechanical properties [403].

Sequence-Dependent Plectoneme Pinning

The locations of plectonemes have recently been shown to be highly sequence-dependent [41]. They locate to intrinsically curved sequences due to the lower energy barrier to forming the sharply bent plectoneme tip. The intrinsic curvature model is based strictly on the dinucleotide stacking parameters of roll, tilt, and twist; and, whilst it has strong qualitative accuracy, it occasionally suffers from false negative predictions. Furthermore, the model accurately identifies sequences of high and low pinning; but fails to accurately distinguish between relative magnitudes. Specifically, the model lacks the resolution to detect subtle differences within similarly high-density sequences and, in one case, incorrectly reverses the experimentally observed order of two high plectoneme density sequences. This suggests an unknown external factor also contributing to the pinning. One such explanation for this is the formation of flipped bases and bubbles even in the positive supercoiling regime (fig. 3.1). This has recently been shown to be possible via molecular dynamics simulations [106]. Here, bubbles formed during positive supercoiling were smaller and less stable than those formed under negative supercoiling. They were always found

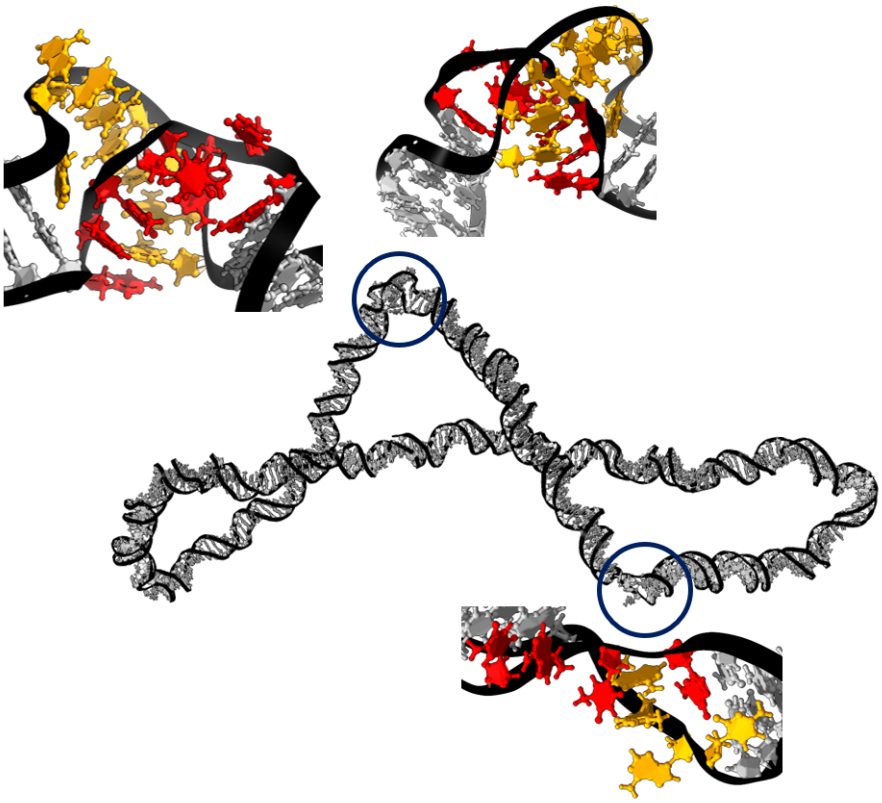

Figure 3.1: Bubbles are defined as the breaking of hydrogen bonds and significant structural deviation from canonical base pairing over a minimum of three base pairs.
A representative frame from a simulation at $\Delta Lk = -6$ with zoomed in inserts of two different bubbles. In the enlarged inserted structures, the two strands of the bubble are coloured separately. The top bubble results from accommodating a sharp bend in the DNA. Here, the bubble is much longer in length, but the base stacking of each strand is preserved as much as possible. The single stranded regions twist around one another in a manner separate to helical twist to induce the sharp bend. The inserted structures at the top are of the same bubble shown from either side to highlight the conserved base stacking, especially in the yellow strand. In the bottom bubble, there is a bubble arising from a local untwisting. There is no base stacking or hydrogen bonding preserved resulting in a greater amount of structural deformation; however, the size of the bubble is shorter.

near the plectoneme tip and were promoted by the high curvature. They were less efficient at absorbing excess twist than the bubbles formed during negative supercoiling and were always nucleated after plectoneme formation. The omission of bubble formation within this model is notable due to their ability to absorb twist and increase local flexibility.

Under negative supercoiling, plectoneme positioning is influenced by both intrinsic curvature and local melting. Whilst bubbles may form across the whole of the DNA under negative supercoiling [106], plectoneme-tip bubbles are highly prevalent [112, 288, 290] and the increased flexibility afforded by them facilitates the sharp bending at the plectoneme tip with no additional energetic penalty. The sequence-dependence of bubble formation under torsion has been shown to be complicated and sometimes counterintuitive [99]. The Stress-Induced DNA Duplex Destabilisation (SIDD) model has been developed to predict regions particularly prone to bubble formation based on thermal denaturation energies [113]. However, the SIDD model is limited in that it assumes complete extension of the DNA and so cannot model DNA that forms a

plectoneme. This is notable due to the influence plectoneme positioning has on bubble formation.

The effect of conditions

It is well-established that DNA twist decreases with increasing temperature [404] and it also softens stretching and bending [405]. However, the overall bending of DNA depends only weakly on temperature owing to mutual compensation of directional local bends [404], and the twist-stretch coupling that defines overall flexibility is nearly unaffected [405]. The sum of distances between base pair centres increases with increasing temperature which is absorbed by an increased helix radius [406].

Salt Dependence

The salt concentration is a critical factor in modulating plectoneme conformation and dynamics due to the screening effect that the ions provide. As the phosphate groups of DNA carry negative charges, and plectoneme formation involves bringing these charges closer together, the ionic shielding of this electrostatic repulsion is highly favourable. At higher salt concentrations, the radius of the plectonemic helix is decreased [407]. This has the effect of decreasing plectoneme nucleation rates [408] due to the higher energy penalty of forming the plectonemic helix. Additionally, the compactness stabilises the plectoneme resulting in a longer lifetime.

The point at which the DNA collapses into a plectoneme, rather than storing the torsional strain solely as twist, is known as the buckling point and it is dependent on salt concentration [104]. The various relationships can be understood by considering the relationship between the energy barrier to forming the initial loop (E1) and the subsequent barrier to elongating the superhelix with additional turns (E2). In general, there is a higher energy per writhe required to form the initial loop than for subsequent turns in the helix, i.e. $E1 >> E2$. Electrostatics have less of a contribution to E1 due to the large DNA-DNA distance within the loop. This means that the formation of a second plectoneme is normally less favourable and, at high ionic strengths, the writhe is stored as a single plectoneme. At lower ionic strengths, the increased repulsion between the DNA strands increases the energy to extend the helix (E2) which decreases the E1-E2 gap [104] and, along with positional entropy [409], increases the favourability of multiple plectoneme formation. The buckling transition is less abrupt at lower salt due to a lower energy barrier to forming the initial loop (E1) and at higher salt, the transition occurs at a lower superhelical density [104]. There is also increased torsional stiffness at the lower ionic conditions which allows the DNA to store more molecular torque as twist before undergoing the buckling transition [410]. Additionally, bubbles absorb torsional strain in the form of twist and so at increased ionic concentrations, there is a decreased probability of bubble opening due to a lower twist/writhe partitioning [117, 411].

Here, I present all-atom implicit solvent simulations coupled with experimental results showing that, despite the limitations of implicit solvent, it is a validated and powerful method to study

the sequence-dependent conformational dynamics of supercoiled DNA on the scale of hundreds of base pairs for microseconds of simulation time. Due to the omission of bubble formation in the Dekker model, and the omission of writhe in the SIDD model, as of yet we cannot completely understand sequence-driven plectoneme positioning. Until such time that an algorithm is developed that combines both models, we propose that implicit solvent simulations are an accessible tool to further investigate the sequence-driven properties that underpin plectoneme positioning and bridge the gap between the two existing predictive algorithms.

3.2 Methods

A total of 495 simulations representing approximately 50 µs simulation time are presented in this chapter. The simulations presented here of a 339 bp minicircle were run in implicit solvent on JADE2 at a rate of 12.5 ns/day. These were validated experimentally using single-molecule FRET.

Implicit Solvent Simulations

All DNA sequences can be found in A.1. The linear DNA starting structures for the simulations were built using the NAB module in AmberTools20 [412]. The two ends were then joined using an in-house program where the twist angle per base pair is changed to create the specified superhelical density. All 339-mers were simulated at $\Delta Lk = -6, -4, -3, -2, -1, 0, 2, 3$, and the Strong339, Mod339, and Control339 sequences under standard conditions were further simulated at $\Delta Lk = 1, 6$. The 126-mer was simulated at $\Delta Lk = 0, -1$. The initial topoisomers were in a planar starting structure with all of the superhelical stress in the twist of the duplex. The parmbsc1 force field [344] with improved ϵ, ζ and χ torsions was used with modifications by Agnes Noy, the University of York, to allow it to describe circular DNA systems. The Hawkins, Cramer, Truhlar pairwise generalized born model [413, 414] was used for energy minimization with parameters by Tsui and Case [415]. It was performed for a maximum of 10,000 cycles with the first 1,500 cycles being of the steepest descent method, and the remainder conjugate gradient minimization. The step length was 2 ps. For the following equilibrations and production run, the GB-neck2 model was used [370, 400] with refined parameters suitable for nucleic acid simulation [362]. The salt concentration was 200 mM and the temperature was 300 K unless otherwise stated. The simulations were performed using the pmemd.cuda module of Amber20 [416]. Long-range electrostatic cut-offs were 1000 Å and the time step was 2 fs. Temperature equilibration was performed for 10 ps with the minicircle restrained with a weight of 50 kcal mol^{-1} Å^{-2}. A further 100 ps was performed with restraint weights of 10 kcal mol^{-1} Å^{-2} followed by 200 ps with restraint weights of 1 kcal mol^{-1} Å^{-2}. The Langevin thermostat was used with a collision frequency of 0.01 ps^{-1}. All bonds to hydrogen were constrained using the SHAKE algorithm. The minicircles were relaxed for 10 ns with distance and angle restraints that maintain the Watson-Crick hydrogen bonds between base pairs following a protocol previously

described in the literature [40, 100, 101, 393]. The restraints were then removed, and each simulation was continued for 100 ns. Each of the simulations were run in triplicate by using different initial velocities. The input file for the simulation production run can be found in A.3. The writhe parameter and positioning of the plectoneme was determined to be stable at 100 ns for all simulations and so we considered these properties to be converged and adequately sampled on this timescale. The positioning of the plectoneme did not change over 100 ns and remained where it first nucleated.

Linear DNA Simulations

In order to investigate the relevant sequence-dependent structural and mechanical properties in the absence of torsion, the Control339, Mod339, Strong339, and 126-mer sequences were also simulated for 200 ns in the linear conformation. This would allow us to investigate the role of intrinsic curvature on plectoneme positioning, rather than the curvature induced by the minicircle. Three repeats were run of each construct following the same implicit solvent protocol, omitting the initial relaxation with hydrogen bond restraints.

Explicit Solvent Simulations

The starting structures for the explicit solvent simulations were selected from the most populated clusters of the implicit solvent simulations at each superhelical density. These were calculated using the average linkage algorithm within the cpptraj module of AmberTools20 [412]. The representative structures were solvated in a rectangular box with the TIP3P water model [417] with a 3 nm buffer and neutralised with Na^+ counterions. Additional Na^+/Cl^- ions were added at a concentration of 150 mM. A multi-stage equilibration protocol was used as follows. Energy minimization was performed for 10,000 cycles with the first 5,000 being of the steepest descent method and the rest conjugate gradient minimization. This was done first with the DNA restrained with a weight of 50 $\mathrm{kcal\,mol^{-1}\,Å^{-2}}$ in order to minimize the solvent and ions. This was repeated with no restraints in order to minimise the DNA structure. The long-range electrostatic cut-off was 8 Å and the timestep was 2 fs. The system was first brought to 100 K for 10 ps followed by a further 10 ps to bring the temperature to 300 K with the Berendsen thermostat. The pressure was brought to 1 bar using the Berendsen barostat with the DNA restrained with a weight of 50 $\mathrm{kcal\,mol^{-1}\,Å^{-2}}$ for 10 ps. Five short 10 ps equilibration runs were then performed with decreasing restraint weights of 25, 10, 5, 2.5, and 1 $\mathrm{kcal\,mol^{-1}\,Å^{-2}}$. A 100 ns production run was then carried out using the pmemd.cuda module of Amber20 [416] with no restraints apart from constraining all bonds to hydrogen using the SHAKE algorithm. The input file for the simulation production run can be found in A.3.

Analysis

Analysis was predominantly performed using AmberTools20 modules [412], SerraLINE to measure bending angles, and WrLINE to trace the molecular contour and thus calculate the twist and writhe [418]. The curvature was calculated using tangent vectors separated by 16 nucleotides, equal to approximately 1.5 helical turns, to neglect the influence of the helix. To avoid end-effects, values with curvatures greater than 1.5 standard deviations of the mean were not plotted for the linear constructs. Implicit solvent is known to systematically overestimate twist [419] and so calculations involving twist were corrected using the average twist over the three simulations of $\Delta Lk = 0$ for that construct. In the case of the increased salt and temperature simulations, the correction at standard conditions was used across all temperatures and salts. Analysis was performed after the removal of hydrogen bond restraints.

Bubbles were defined as the breaking of hydrogen bonding across a minimum of three base pairs, and the same base pairs having propeller twist, opening, and buckle values of, on average, at least 2 standard deviations away from the corresponding values in the same relaxed minicircle. This follows the protocol outlined in [106]. Bubbles that migrated due to the continuous denaturation of neighbouring base pairs, but not via repair and re-nucleation, were treated as the same bubble.

For the statistical analysis, bubbles were first removed that had a shorter lifetime than 0.5 ns in order to exclude transient denaturation that would not affect the minicircle conformation. The Brunner-Munzel test [420] was used to compare bubble size and lifetime between two groups: the two directions of supercoiling, or being located at the plectoneme tip or not. Where a bubble grew or migrated, the largest size of that specific bubble was recorded in the data. The Kruskal-Wallis test [421] and Dunn's post-hoc test [422, 423] with Bonferroni corrections were used to compare between multiple groups: sequence and the linking number. All statistical analysis was carried out using the scipy.stats [424] implementation in Python 3.12.7. Bubbles at the tip of the plectoneme were defined as being within 20 nt either side of the two base pairs with the largest distance from the centre of mass of the DNA, as calculated from the plectoneme propensity polar plots. The second highest peak was found by searching by distance at each frame and skipping over the base pairs that were adjacent to the previous base pair until the second plectoneme tip was identified.

Plectoneme Pinning Prediction

The IGOR Pro implementation of the Dekker model [41] with small modifications was used to predict plectoneme localisation. The original prediction model was written for predictions of DNA on the kilo-base scale and for DNA tethered at either end to a surface in order to replicate their ISD experiments. This originally included the effect of the handles limiting plectoneme growth at the attachment points, which is not necessary in this case. The code was modified such that no ends were bound to a surface by setting the BindLength to 0. The number of base

pairs used to calculate the local tangent vectors was reduced from 10 to 2 and the CurveWindow that defines the loop sizes over which to calculate the predictions was started from 2bp instead of 40bp to reflect the shorter DNA length. The effect of changing the window increased the sampling and was not computationally expensive owing to the short nature of the sequences. Overall, these changes created a more rugged prediction curve, which was necessary on the shorter length scale. We verified that making these changes did not otherwise change the results.

Bubble Prediction

The Twist-DNA implementation [387, 388] of the SIDD model [113] was used to predict bubble opening probabilities. The program is built to perform calculations on the genome (Mbp) scale and so the 126 bp minicircle was too short for the calculations. The sequence was repeated and the results restricted to one repeat in order to circumvent this. The 339 bp minicircles were large enough that this was not necessary. Bubbles with a probability of at least -10 were saved and the largest probability per base pair plotted.

AV Modelling and Experiment Design

The positioning of the dyes for the smFRET measurements were chosen via AV modelling using Olga [386]. The FRET efficiency between Atto550 and Atto647N was estimated using parameters from the literature (table 3.2) [379]. The R_0 was 62.6 Å. The representative structures resulting from the cluster analysis of each simulation at 0.2 M were used to predict the FRET efficiency between all thymines within the pinned region, and also all thymines within the region of low predicted plectoneme density. The pairs were chosen due to being the same distance apart and having a significant difference in predicted FRET efficiency, both compared to each other and across the relaxed and supercoiled regimes.

Table 3.2: *Atto550 and Atto647N AV parameters as calculated in the literature [379]. R values refer to the three radii of the dye when modelled as a sphere. All values are given in* Å.

Dye	Linker Length	Linker Width	R1	R2	R3
Atto550 (donor)	20.4	4.5	7.1	5.0	1.5
Atto647N (acceptor)	20.4	4.5	7.2	4.5	1.5

DNA Minicircle Assembly

Minicircles were assembled using an adapted protocol from [425]. Three main steps were used, first, the single-stranded DNA (126 nt) and splint DNA (29 nt) were heat annealed together at 95 °C for 3 minutes and subsequently cooled down on ice for 1 minute. The split contained the modified thymines that were labelled with the Atto550 (donor) and Atto647N (acceptor) dyes. The reaction was incubated with T4 DNA ligase (20 U) at 37 °C for 30 minutes in order to form a ssDNA minicircle. Secondly, the ssDNA minicircle was subjected to T4 DNA polymerase

fill-in, the ssDNA minicircle was diluted into a reaction containing T4 DNA polymerase (120 U) and T4 DNA ligase (16,000 U) in the presence of 100 mM dNTP mix and 10 mM ATP, the reaction was incubated at 12 °C for 1 hour. Finally, the reaction was treated with Exo V (20 U) and T5 Exo (20 U) in the presence of 1 mM ATP at 37 °C for 45 minutes, this was done to remove all incompletely assembled minicircles. To clean the reaction further, it was treated with Proteinase K to remove any remaining protein in the reaction before applying the reaction through a Monarch PCR & DNA Clean-up Kit (NEB).

Negatively supercoiling DNA minicircles

Minicircles were subjected to negative supercoiling using *E. coli* Gyrase (Inspiralis) in the presence of Ethidium Bromide (EtBr). A reaction of 10 nM mcDNA, 1X Gyrase Buffer, 0.1 mg/ml EtBr, 20 U/µL of *E. coli* Gyrase was incubated at 37 °C for ≥1 h. The reaction was then cleaned using a Monarch PCR & DNA Clean-up Kit (Bulk/Cryo-EM).

smFRET

Single-molecule FRET data was collected on the custom-built confocal microscope, smfBox. Full details of the construction and operation of the instrument are described elsewhere [426] including a step-by-step method protocol [427]. Briefly, the smfBox alternates two lasers (515 nm, ~220 μW and 635 nm, ~70 μW, Omicron LuxX Plus lasers, powers measured immediately before the excitation dichroic) by TTL-controlled modulation of electronic shutters. The beams are coupled into a single-mode fibre before being collimated to 10 mm and cropped by an iris to 5 mm, then directed into a custom-built anodized-aluminium microscope body.

Samples were diluted to ~5 pM in observation buffer (200 mM KCl, 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 0.1 mg/mL photobleached BSA) and measured on the smfBox. To assess the effect of ionic concentration, the concentration of KCl was increased to 1800 mM.

Data analysis was performed in Jupyter Notebooks using the FRETBursts Python module (version 0.7.1) [378]. The background for each channel was estimated via exponential fitting of inter-photon delays. Bursts were identified by performing a dual-channel burst search (DCBS) using a photon sliding window algorithm [428], with m=10, F=20 (for each channel) and DD+DA and AA thresholds =30, to extract doubly-labelled bursts from each acquisition.

Burst Variance Analysis (BVA)

We used burst variance analysis (BVA) [429] to analyse the FRET distributions. This method distinguishes between static components and interconverting states. The analysis involved plotting observed and expected standard deviations of the proximity ratio for each burst, alongside confidence intervals (CIs), against the chosen X-axis (proximity ratio or FRET efficiency). Bursts were segmented into consecutive windows with specified photon counts, and

standard deviations were binned along the X-axis with a 0.05 width. A threshold was set for bursts per bin to ensure statistical power by excluding bins with low counts.

3.3 Results

Minicircles of different sequences were simulated, using all-atom implicit solvent molecular dynamics simulations, to investigate the influences of plectoneme positioning under both positive and negative supercoiling. To design the sequences, the Dekker plectoneme prediction algorithm [41] was used to predict plectoneme positioning under positive supercoiling. Under negative supercoiling, it was assumed that bubble formation is the main driver of the pinning and so the SIDD algorithm [113, 387, 388] was used to predict bubble location and, therefore, plectoneme positioning under negative supercoiling.

The simulations use a protocol previously described in the literature [40, 100, 101, 393] where the starting structure is a planar minicircle with all of the superhelical strain stored as twist (fig. 3.3). The minicircle then relaxes for 110 ns, the first 10 ns of which there are restraints that maintain the hydrogen bonding between the base pairs. The first 10 ns is discarded and the positioning of the plectoneme over the rest of the simulation is used for analysis. We compare trends from the simulations against previous results in the literature and our own, novel in-house biophysical experiments to validate the use of these simulations to study sequence-dependent effects.

The analysis of these simulations, in this chapter, focuses on the plectoneme positioning and bubble formation within the different sequences. To determine the plectoneme positioning during a simulation, the average distance between each base pair and the centre of mass of the DNA is plotted on a polar plot. Here, each colour on the polar plot is a different replica of the simulation at the given superhelical density, and the average distance $\pm$ the standard deviation of the average distance for that replica is plotted. The positioning of the theta axis is such that the first base pair is at the mid-point on the top of the plot, and the numbering increases in a clockwise direction. Finally, the mean distance between the base pairs and the centre of mass of the DNA is averaged over all simulations at all superhelical densities of a given direction of supercoiling and plotted on a polar plot. Herein, this is called the plectoneme propensity. The average of the plectoneme propensity is plotted as a guide such that greater than this value signifies plectoneme positioning in this area, and less than this value signifies little plectoneme propensity in this region. Here, polar plots are often used such that the final, circular plot forms the average shape of the minicircle over all simulations and the location of the plectoneme can be easily identified (fig. 3.3). Low, global plectoneme propensity for the sequence overall can then be identified by a more regular circle in the end polar plot, signifying multiple positions within the minicircle where the plectoneme will form. The bubble densities are plotted in blue using the same theta axis describing the base pair position within the minicircle. Here, the number of frames where a bubble is found at the corresponding base pair is plotted. Both the

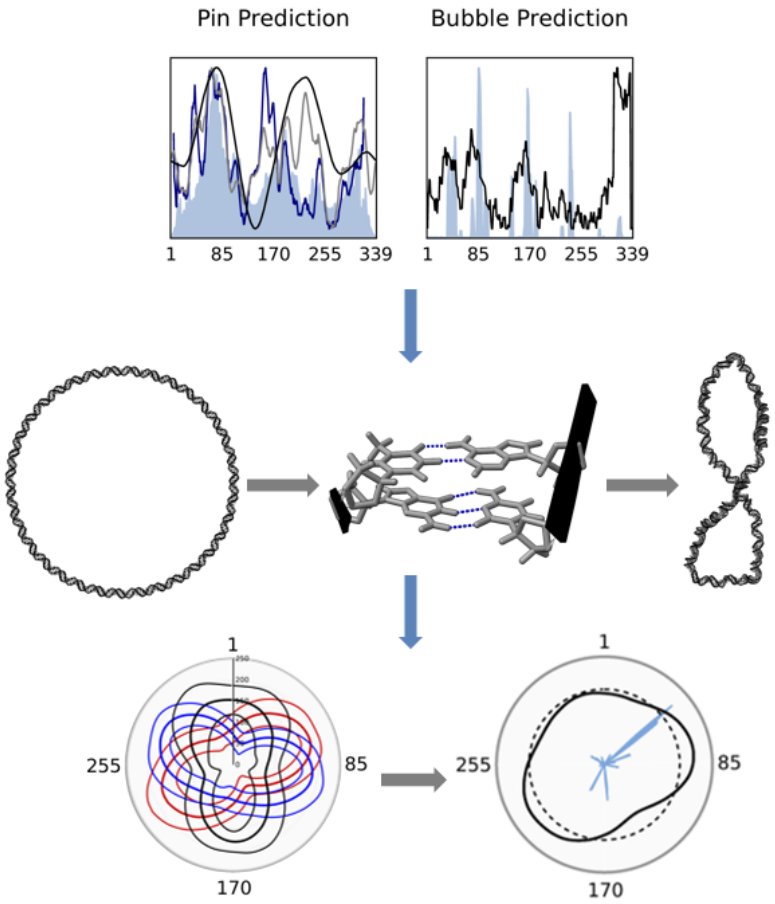

Figure 3.3: The workflow of the simulations and analysis presented in this thesis. The workflow for this research generally proceeds via three stages. First, the Dekker and SIDD algorithms are used to predict plectoneme and bubble positioning to design sequences with the desired properties. Secondly, the sequences are simulated as DNA minicircles with a planar starting structure. The minicircles relax for 10 ns with restraints that maintain the Watson-Crick hydrogen bonding. The restraints are removed and the simulations continue for 100 ns. The analysis of the simulations then involves mapping the plectoneme positioning and bubble densities over all simulations and plotting as polar plots. In order to compare to the Dekker and SIDD predictions, the flattened plots are combined with the SIDD and Dekker predictions. **Top, left:** The Dekker prediction (blue, filled), curvature of the minicircle during positive supercoiling (grey), curvature of the linear construct (navy), and flattened pinning propensity (black). **Top, right:** The SIDD prediction (black) and flattened bubble density plot (blue, filled). **Bottom, left:** The average distance between each base pair and the centre of the mass of the DNA where each colour represents a different repeat of a simulation. The average distance ± the standard deviation of this average is plotted. **Bottom, right:** The individual mean distances over all simulations is averaged across each direction of supercoiling (black). Herein, this is called the plectoneme propensity. The average of the plectoneme propensity is plotted as a guide (black, dash), along with the bubble densities (blue, filled). They are normalised such that, for both the plectoneme and bubble density, the maximum value of each is equal to 1 on the plots.

plectoneme propensity and bubble density are normalised such that the maximum value for each plot is equal to 1.

Finally, to compare the results of our simulations to those predicted by the Dekker and SIDD algorithms, we return to the original predictions made. Here, the flattened polar plot on traditional Cartesian axes is used to represent the plectoneme propensity. The pinning

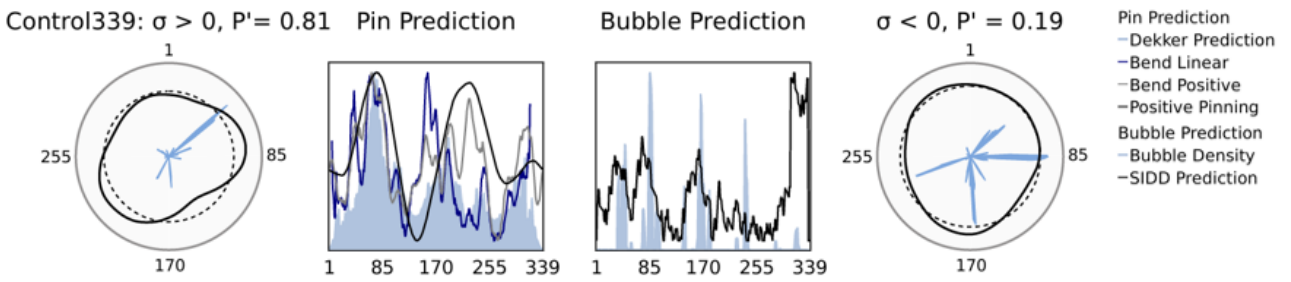

Figure 3.4: The Control339 sequence was designed to not have a strong pinning propensity in either direction of supercoiling.
Polar plots of the pinning propensity over the Control339 minicircle under positive **(far left)** and negative **(far right)** supercoiling. They are normalised such that, for both parameters, the maximum value is equal to 1 on the plots. The ratiometric pinning propensity, P', describes the relative pinning propensities between the two directions of supercoiling as the sum of both equals one. **Middle left:** The Dekker prediction of plectoneme positioning under positive supercoiling is plotted (blue, filled) along with the flattened plectoneme propensity polar plot under positive supercoiling (black), the bend angle over all simulations of positively supercoiled minicircles (grey) and the bend angle over all repeats of the linear Control339 simulation (navy). **Middle right:** The SIDD prediction (black) of bubble density and the actual bubble densities over all simulations under negative supercoiling (blue, filled) are plotted.

propensity under positive supercoiling is then compared to the Dekker predictions and the bending angle of the minicircle averaged base pair-wise over all simulations under positive supercoiling is used as a proxy for intrinsic curvature. As this bend angle is enforced by the plectoneme positioning, for some sequences, the linear construct is also simulated and the bend angle averaged over these simulations plotted. The SIDD predictions are then compared to the bubble densities across all simulations under negative supercoiling. Throughout this thesis, analysis has been separated for the two directions of supercoiling to elucidate any differences between the two.

3.4 Pinning Propensity

We defined a metric: pinning propensity, P, to quantify pinning strength based on the partition of sums of squares concept that describes the variance of the polar plots (eq 3.1). It compares the variance of the mean distance of each base pair, x_{bp}, with the mean value of the overall minicircle, $\bar{x}$. It is the latter that is plotted as a guide on the polar plots. Note the distinction between our pinning propensity metric, P, sometimes referred to as a P value or P metric, and p, referring to a p-value during the statistical tests later presented in this thesis.

$$P = \sum_{bp=1}^{n} (x_{bp} - \bar{x})^2 \tag{3.1}$$

A weakly pinning sequence would form the plectoneme apices, and therefore have a large distance x, across different regions of the minicircle. This would have the effect of a more uniform distribution of the overall mean, x_{bp}, across the whole minicircle, creating less variance,

and lowering the P value. Conversely, a strongly pinning sequence would have two regions with a large increase from the mean, representing the plectoneme tips, and two regions with a lower value compared to the mean, representing the crossing points, creating large variance and a higher P value.

A 339 bp low pinning sequence was designed as a baseline against which to measure our results. In this sequence, the plectoneme tip can be positioned in multiple places across the DNA and shows no strong pinning propensity, especially in the negative supercoiling regime (fig. 3.4). Herein, we have called this sequence Control339.

3.5 Validating Relative Pinning Propensities

To test the ability of molecular dynamics simulations to accurately capture sequence-dependent plectoneme positioning, two intrinsically curved 75-mers and two flat 75-mers were inserted into Control339 that have previously been experimentally validated in the literature under positive supercoiling [41]. Using a plectoneme prediction algorithm (see section 2.3) based on intrinsic curvature and local flexibility, the Curve75-1 and Curve75-2 sequences were predicted to be of high plectoneme density in the positive supercoiling regime, whereas the Flat75-1 and Flat75-2 were not (fig. 3.5). Using ISD, these predictions were experimentally verified with Curve75-2 being the strongest pinner of the two curved sequences, shown by having a higher relative plectoneme density, despite Curve75-1 being predicted to be the stronger of the two by the intrinsic curvature algorithm. Our simulations in the positive supercoiling regime replicate the experimental results with Curve75-2 having a higher P value of 2.33 compared to Curve75-1's

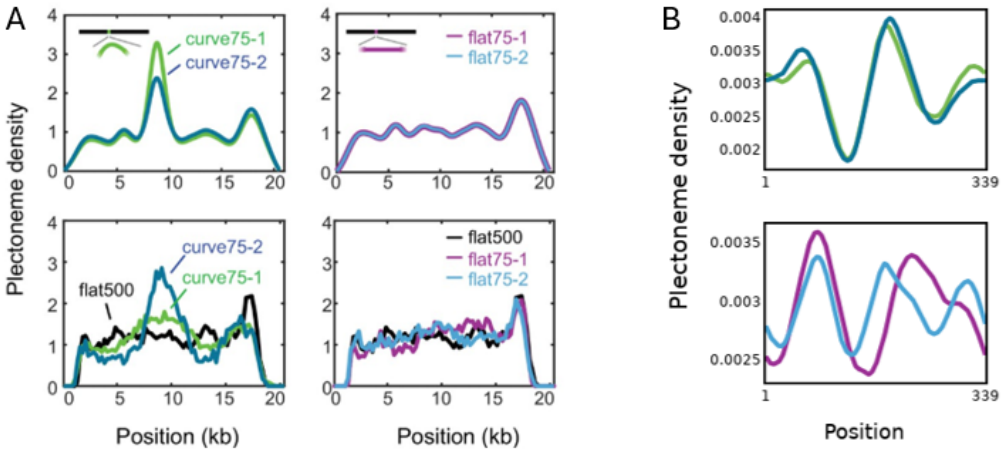

Figure 3.5: Our simulations quantitatively replicate the relative pinning propensities of the curved and flat Dekker sequences.
A: The predicted **(top)** and experimentally determined **(bottom)** plectoneme densities of the curved **(left)** and flat **(right)** Dekker sequences. Figure taken directly from Kim *et al.* [41]. **B:** The flattened polar plots of the simulations of the curved and flat Dekker sequences under positive supercoiling. Curve75-2 (green) is a more strongly pinning sequence than Curve75-1 (blue) as seen in both the simulations and the ISD experiments **(A, bottom, left)** This was the opposite way round to that predicted by the intrinsic curvature algorithm **(A, top, left)**.

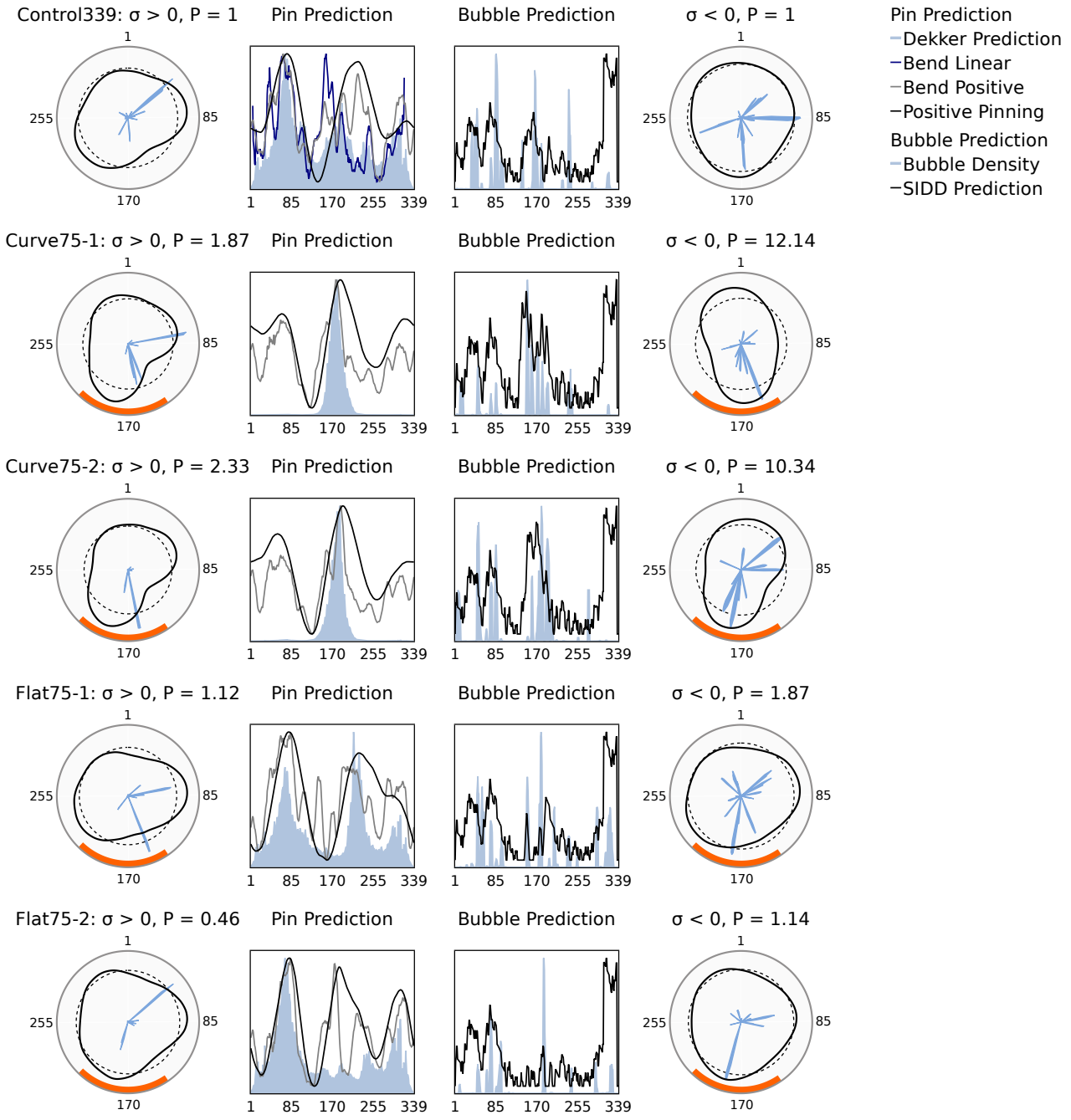

Figure 3.6: In the positive supercoiling regime, the flat sequences show no apparent plectoneme pinning propensity whilst Curve75-2 has a higher P value than Curve75-1 in line with experimental results from Kim *et al.* [41].
Polar plots of the pinning propensities under positive **(far left)** and negative **(far right)** supercoiling. They are normalised such that, for both parameters, the maximum value is equal to 1 on the plots. The plectoneme pinning propensity, P, is calculated individually for the different supercoiling regimes and normalised such that the control sequence is equal to 1. The location of the inserted curved and flat 75-mers within the Control339 sequence are shown by an orange band on the polar plots. **Middle left:** The Dekker prediction of plectoneme positioning under positive supercoiling is plotted (blue, filled) along with the flattened plectoneme propensity polar plot under positive supercoiling (black), the bend angle over all simulations of positively supercoiled minicircles (grey) and the bend angle over all repeats of the linear Control339 simulation (navy). **Middle right:** The SIDD prediction (black) of bubble density and the actual bubble densities over all simulations under negative supercoiling (blue, filled) are plotted.

1.87 (fig. 3.6). This could be explained by the higher degree of bubble localisation—herein meaning the specific nucleation of bubbles at one region with comparatively low to no bubble density elsewhere in the minicircle—in the Curve75-2 construct under positive supercoiling. This suggests that the Dekker model's omission of bubble formation may limit its ability to quantify subtle differences in relative pinning propensity. Under negative supercoiling, stronger bubble localisation in the Curve75-1 construct leads it to have a stronger plectoneme positioning effect compared to Curve75-2, as judged by their P values. Overall, as expected, bubbles are significantly more prevalent across the minicircle under negative supercoiling.

To evaluate whether the sequence-dependent landscapes of these motifs are in line with what we would expect, we compared the curvature seen in our simulations and the plectoneme positioning in the positive supercoiled simulations to that predicted by the Dekker algorithm (fig. 3.6). High curvature values and plectoneme positioning are seen to coincide with Dekker predictions. The bending angles in the supercoiled simulations are guided by plectoneme and bubble formation, and whilst they describe the bending across the simulations, they do not necessarily describe *intrinsic* curvature. To verify that our simulations capture this, we simulated the linear construct of the control sequence and calculated the bending angles in the same way. We found that the bend angle distribution in the linear construct is largely in agreement with the bend angle distribution across the supercoiled minicircle simulations. Furthermore, the largest peak predicted by the Dekker algorithm indeed coincides with a high bending angle in the linear construct and high plectoneme density in our simulations. The nature of the minicircle means that when one region of the circle is pinned, modulated through bubble formation and intrinsic curvature, the other side of the circle is inherently pinned no matter its preference for curvature and bubble formation. We later refer to this as the double pinning effect. The effect of this can be seen in the control sequence where the positioning of one plectoneme tip is dominated by the high intrinsic curvature in this region. The second tip forms slightly off-set from the preferred high bend region. There is likely a balance between the favourability of positioning the plectoneme apex in a given region and the inherent pinning of the other side that is not seen in linear constructs.

Where there are multiple peaks in the Dekker predictions, the positioning of the plectoneme most often coincides with the largest peak. The Dekker prediction algorithm was designed for linear sequences and will not account for this "double pinning" effect of circular DNA. This can be seen most clearly in Flat75-2, which has the lowest P value (fig. 3.6). Three distinct peaks are seen in the plectoneme positioning plots. This likely occurs due to the minicircle oscillating between the two most favourable peaks, as seen in the Dekker prediction, with the middle peak arising due to the plectoneme positioning opposite.

Where the sequence is strongly-pinning, such as is the case for the curved sequences, there is a single predicted peak despite the weakly pinning sequences having peaks in the surrounding sequence, and the surrounding sequence being the same across all constructs. These other

peaks are lost to the baseline in the strongly pinning sequences. In the curved sequences, there is a two-fold effect in that the inserted sequences are strongly pinning and that the opposite plectoneme apices are positioned at the location of these peaks predicted in the flat sequences. This shows that the curved sequences were inserted in the optimal position for the double pin and the favourable pinning in the opposite side likely aids the strong pinning.

We further compared the location of the bubbles predicted by SIDD to their conformational landscape in our negatively supercoiled simulations. In general, the locations predicted by SIDD coincide with those seen in our simulations, but there is a variation in the relative probabilities. This is seen especially in the 310-339 region where SIDD has predicted an especially high probability of bubble formation that is not seen in our simulations. The rest of the sequence is more in agreement. However, this can be explained by considering the nature of bubble positioning. SIDD is unable to model DNA that forms plectonemes, and it is well established that the location of bubbles and plectoneme tips are highly correlated [106, 110–112]. The bubble nucleation energy may be lowered by the high curvature at the apex of the plectoneme which increases bubble densities in this area. Furthermore, the SIDD algorithm models DNA as completely extended and thus unaffected by DNA curvature. This is known to overestimate bubble predictions in some cases [117, 430]. Elsewhere in the minicircle, SIDD also predicts bubble formation in regions of high curvature where the plectoneme tip preferentially forms. The combination of curvature and favourable bubble formation locates the plectoneme tip at this region and localises bubble formation. In the cases of the strongly pinning curved sequences, very little bubble formation is seen at this highly predicted region. However, in the control and, especially, the flat sequences, the plectoneme tip may position itself here and, as a result, some bubbles do form in this region and bubble positioning is less localised.

Whilst there is reasonable agreement between the intrinsic curvature and the predicted plectoneme positioning with the Dekker algorithm, it is not unsurprising that there are some small discrepancies. Firstly, the Dekker algorithm was developed to predict plectoneme positioning on the scale of kilobase pairs, and the original model has a tip-loop region on the order of 73 bp. Secondly, the Dekker model combines curvature with the influences of local persistence length, calculated using roll-tilt covariance matrices, that we have not considered. However, intrinsic curvature remains the dominant predictor in the algorithm, and we are confident that the regions of increased curvature seen in our simulations are not artifacts of implicit solvent and show the correct sequence-dependence.

3.6 Validation Against Explicit Solvent

The sequence-dependent pinning in the positive supercoiling regime has been validated experimentally; however, bubbles are a strong determinant of pinning, especially under negative supercoiling. It was therefore necessary to verify that bubbles were sequence-dependent in nature and not artifacts of the reduced stability in implicit solvent or simply positioned at the

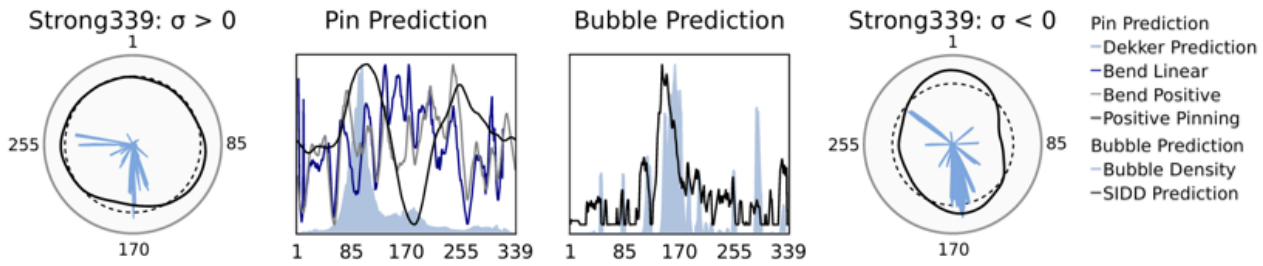

Figure 3.7: The Strong339 sequence was designed to have a high pinning propensity driven by significant bubble formation in the negative supercoiling regime.
Polar plots of the pinning propensities of Strong339 under positive **(far left)** and negative **(far right)** supercoiling. They are normalised such that, for both parameters, the maximum value is equal to 1 on the plots. **Middle left:** The Dekker prediction of plectoneme positioning under positive supercoiling is plotted (blue, filled) along with the flattened plectoneme propensity polar plot under positive supercoiling (black), the bend angle over all simulations of positively supercoiled minicircles (grey) and the bend angle over all repeats of the linear Control339 simulation (navy). **Middle right:** The SIDD prediction (black) of bubble density and the actual bubble densities over all simulations under negative supercoiling (blue, filled) are plotted.

plectoneme tip due to the increased strain. Thus, a strongly pinning construct, Strong339, was designed that has a large predicted bubble density in a region separate to the Dekker-predicted regions of high plectoneme density (fig. 3.7).

In this construct, it is expected that the positive supercoiling regime would have a higher plectoneme density in the 50-100 nucleotide region, and this would shift to the 100-150 nt region under negative supercoiling due to the influence of the bubbles. This resulted in distinct plectoneme positioning landscapes across positive and negative supercoiling of comparably high propensities. Interestingly, despite also having strong bubble formation at this area in the positive supercoiling regime, the plectoneme, as predicted, does not pin here. This suggests that the flexibility and intrinsic curvature remains the main driver of plectoneme positioning under positive supercoiling, even when bubbles do form. This could be explained by results from the literature that show that, under positive supercoiling, bubbles nucleate after plectoneme formation rather than drive the positioning [106].

We further simulated the negative supercoiling regime in explicit solvent to evaluate the effect of the discrete, stabilising ions on the bubbles. Despite the bubbles being larger ($p = 5.34 \times 10^{-60}$) and shorter lived ($p = 1.58 \times 10^{-4}$) in implicit solvent (fig. A.1), the plectoneme positioning in the negatively supercoiled simulations is the same across both implicit and explicit solvent (fig. 3.8). Furthermore, the positioning of the bubbles across the simulations are the same and both agree with SIDD predictions. The nature of the solvent has a greater influence on bubble size, where post-hoc testing showed a significant difference in the distribution of sizes at each superhelical density ($p < 0.01$ for each). When comparing bubble lifetime, there is only a significant difference ($p < 0.01$) at the highest superhelical density, $\Delta Lk = -6$. Evidently, the utilisation of explicit solvent to study the sequence-dependent conformational dynamics of supercoiled DNA is not necessary as the limitations of implicit solvent do not impede its ability to accurately capture plectoneme pinning landscapes. In fact, the use of implicit solvent

is enforced in minicircles of this size as the conformations of the DNA did not significantly change over the course of the 100 ns and each simulation took approximately 3 months to run on JADE2. Note that the pinning propensity landscape appears to be stronger in explicit solvent; however, the starting structure for the explicit solvent simulations is taken directly from the implicit solvent simulations and thus the plectoneme is already formed. This, coupled with the little structural change over the simulation, leads to the appearance of an artificially stronger pinning propensity.

3.7 Validation Against The Literature

Finally, Control339 was modified in order to obtain a construct with a similar plectoneme density profile across both positive and negative directions of supercoiling - Mod339 (fig. 3.9). This was done as the Mod339 sequence now acts as a "mid-strength" pinning sequence under negative supercoiling. In this regime, Control339 has a low plectoneme propensity, whilst Strong339 has a high pinning propensity. Further, having a similar plectoneme density profile would be beneficial later in the thesis as comparisons can more easily be made between the positive and negative directions of supercoiling. Control339 has three main peaks in both the predicted plectoneme density plot, and in the predicted bubble density plots using SIDD. However, they were of three different magnitudes with the largest peak in the Dekker algorithm being orthogonal to the largest peak in the SIDD prediction, thus creating different plectoneme pinning landscapes across the positive and negative superhelical regimes. In order to create the same plectoneme densities across both positive and negative supercoiling, the pinning due to intrinsic curvature had to be balanced with the preferential location of bubble formation. To do this, a total of 14 base pair mutations were made which overall increased the AT% from 46.3% to 50.4%. This was achieved by the creation of an A-tract within the largest, left-most peak in the Dekker predictions to increase bubble density in this region. Plectoneme density in the middle region was increased

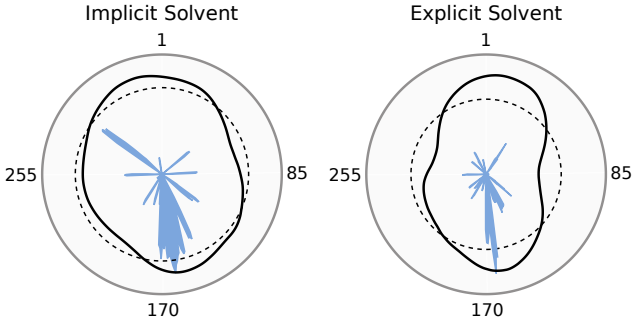

Figure 3.8: The bubble density and plectoneme positioning landscapes are similar across the implicit and explicit solvent simulations of Strong339 under negative supercoiling. Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the Strong339 sequence under negative supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The simulations in implicit solvent **(left)** and explicit solvent **(right)** are compared.

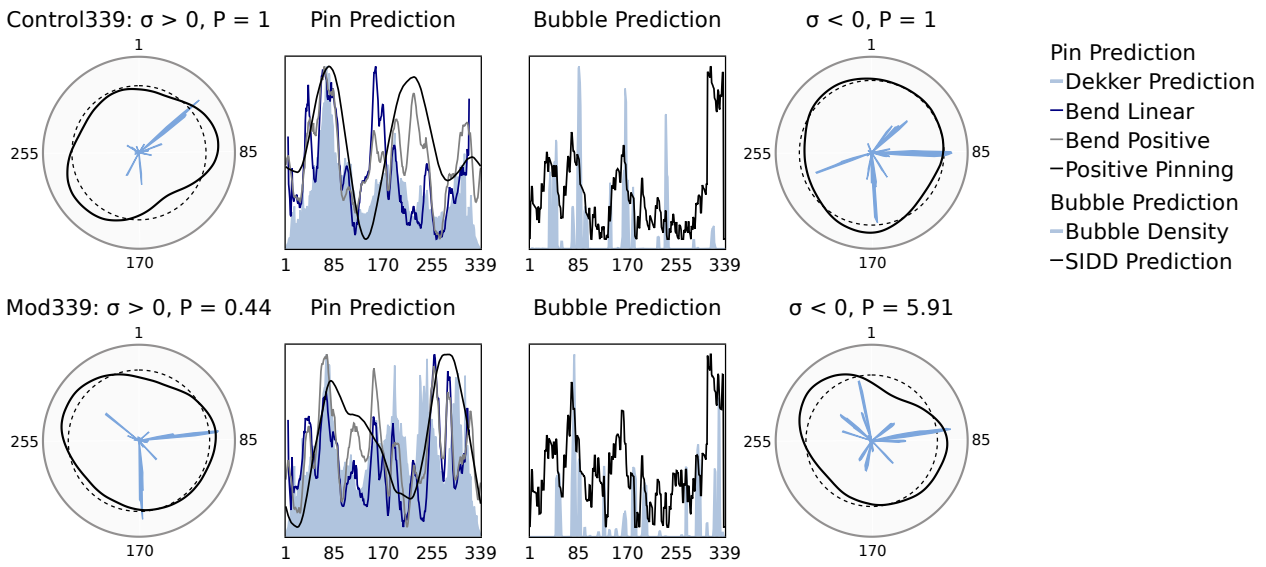

Figure 3.9: Mod339 was designed such that the positioning of the plectoneme was the same in both directions of supercoiling.
Polar plots of the pinning propensities of Control339 **(top)** and Mod339 (bottom) under positive **(far left)** and negative **(far right)** supercoiling. They are normalised such that, for both parameters, the maximum value is equal to 1 on the plots. **Middle left:** The Dekker prediction of plectoneme positioning under positive supercoiling is plotted (blue, filled) along with the flattened plectoneme propensity polar plot under positive supercoiling (black), the bend angle over all simulations of positively supercoiled minicircles (grey) and the bend angle over all repeats of the linear Control339 simulation (navy). **Middle right:** The SIDD prediction (black) of bubble density and the actual bubble densities over all simulations under negative supercoiling (blue, filled) are plotted.

by increasing the AT% content from 38% to 50% across the mutated region, and from 45.5% to 50.5% across the middle peak overall. Finally, the AT content was increased from 45% to 52.5% in the highly predicted bubble region and from 39.2% to 51% across the mutated bases. The sequence retained 95.9% similarity to Control339. Overall, this had the effect of decreasing the pinning propensity in the positively supercoiled minicircle as the plectoneme localisation due to intrinsic curvature is less dominated by one region; and increasing the pinning propensity under negative supercoiling compared to the control sequence due to the intrinsic curvature being more in line with bubble preferences. As a result, the positioning of the plectoneme is the same across both regimes in this mutated sequence. We used our Control399, Mod339, and Strong339 sequences, together with the curved and flat Dekker sequences, to evaluate the performance of implicit solvent simulations in capturing various properties from the literature.

Bubbles

Validating and understanding bubble properties such as size, lifetime, and their effect on sequence is crucial to uncovering the drivers of plectoneme positioning. The P values obtained are much higher in the negatively supercoiled regime, especially in the more strongly-pinning sequences, where the influence of bubbles is greater. Bubble formation dominates the plectoneme pinning landscape and a strong localisation of bubbles would greatly influence results. Thus, we sought to further understand the role of bubbles in pinning a plectoneme.

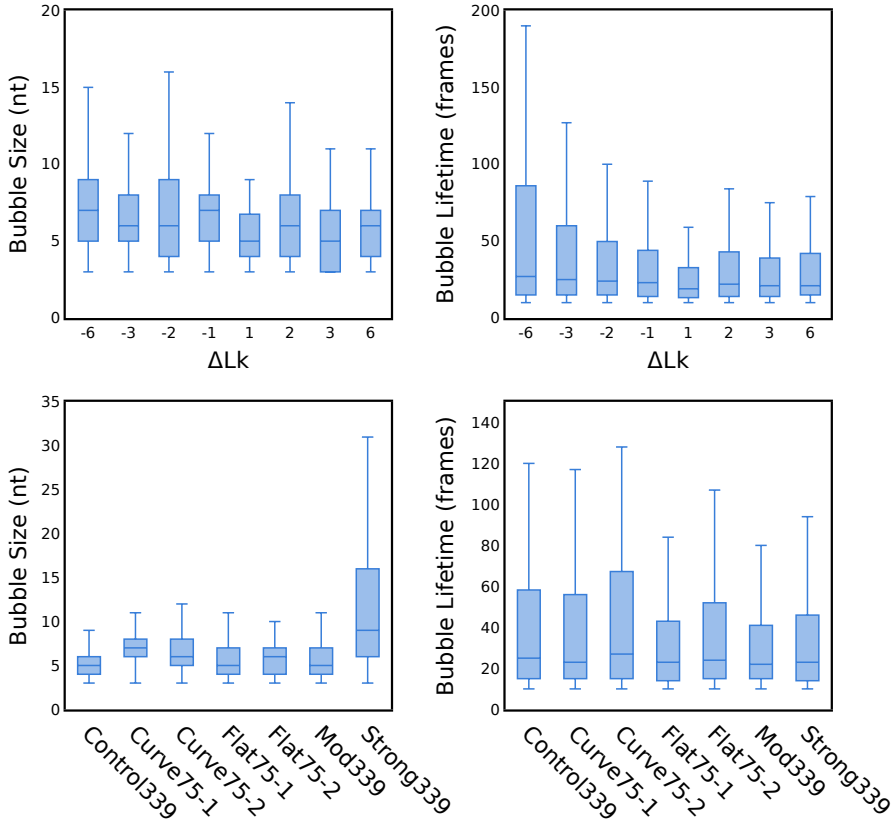

Figure 3.10: Both bubble size and lifetime are influenced by the degree of supercoiling; however, sequence only influences the size of bubbles and not their lifetime.
Box plots of the bubble sizes **(left)** and bubble lifetimes **(right)** compared across superhelical density **(top)** and across sequence **(bottom)**. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There are a possible total of 2000 frames. For comparison by superhelical density, the data was plotted for all sequences at each specified number of turns. For comparison by sequence, the data was plotted for all superhelical densities of the corresponding sequence. For clarity, the outliers (values $>$ Q3 + 1.5IQR) have been removed from the plots, where Q3 is the 75th percentile of the data, and IQR is the interquartile range, or the range between the 75th and 25th percentiles. The full data can be seen in fig. A.3. See fig. A.4 for pairwise comparisons.

By comparing the bubbles formed across all sequences at the different superhelical densities, the Kruskal-Wallis test revealed that both the size ($H = 113; p = 2.12 \times 10^{-21}$) and the lifetime ($H = 40.4; p = 1.03 \times 10^{-6}$) of bubbles are influenced by the degree of supercoiling (fig. 3.10). Similarly, by comparing across all degrees of supercoiling for the different sequences, the sequence influenced bubble size ($H = 487; p = 4.62^{-102}$) but not bubble lifetime ($H = 11.4; p = 0.077$).

Here, Dunn's post-hoc test (fig. A.2) revealed interesting results suggesting that the bubble size could be linked to the pinning propensity. The pairwise comparisons showed Control339, Mod339, and the two flat 75-mers to be statistically different to the Curve75 sequences and to Strong339. The two curved sequences were statistically different to all other sequences apart from each other. Finally, Strong339 was statistically different to every other sequence. Thus, this suggests that within our data, there lies three groups with significantly different bubble sizes:

the first is Strong339; Curve75-1 and Curve75-2 are in the second group; finally, Control339, Mod339, Flat75-1, and Flat75-2 make up the third group. These are in line with the pinning propensities of the sequences: Curve75-1 and Curve75-2 have strong pinning propensities that were designed to be driven primarily by intrinsic curvature. The curvature and flexibility in this region thus facilitated bubble formation and leads to relatively larger bubbles in these sequences compared to the Control339, Mod339, and the two flat 75-mers. However, the bubbles are not as large in the curved 75-mers as they are in Strong339 where the bubbles are the dominant driving force of plectoneme positioning. Together, as bubble lifetime is not influenced by sequence, this suggests that bubble size, and not stability, is the sole driver of plectoneme positioning. In reference to our bubble density plots, this signifies that it is the width of the peak, and not the height, that is the driving force. This agrees with the previous literature, where increasing the length of a permanent bubble in coarse-grained simulations deterministically increased the pinning probability in that region [288].

However, despite both bubble size and lifetime being influenced by superhelical density, Dunn's post-hoc test (fig. A.4) revealed no clear link between the degree of supercoiling and either property. The box plots appear to show a linear increase in bubble lifetime under negative supercoiling as the superhelical density increases. However, pairwise comparison reveals limited statistical difference. This may be due to the same median value across all superhelical densities, both positive and negative, despite larger variations around the 75th percentile and beyond. This suggests that bubble stability is an inherent property of the bubble, and likely bubbles in implicit solvent, that is unaffected by overall duplex destabilization, determined by the magnitude of the superhelical strain. Under positive supercoiling, visual inspection of the box plots suggests that there is little difference in bubble lifetime across the different superhelical densities. This is further confirmed in the pairwise comparisons (fig. A.4). Thus, we reduced our analysis to solely compare across the two directions of supercoiling. The Brunner-Munzel test revealed, in agreement with previous literature [106], that bubbles in the negatively supercoiled regime are both larger ($p = 1.66\times^{-15}$) and longer lived ($p = 1.46 \times 10^{-6}$) than in the positively supercoiled regime (fig. 3.11).

Finally, as bubble size seems to be indicative of pinning strength, we tested for statistical differences between bubbles at the tip of the plectoneme and elsewhere within the minicircles. By combining across all sequences and superhelical densities, it was found that bubbles at the tip of the plectoneme, surprisingly, tended to be smaller ($p = 7.45 \times 10^{-6}$) despite having the same median value of size, but are significantly longer lived ($p = 2.42 \times 10^{-10}$) than those positioned elsewhere in the minicircle (fig. 3.11).

This, at face value, disagrees with our results suggesting that bubble size, and not stability, drives plectoneme positioning. However, this may be an artifact of the double pinning effect, as the formation of the plectonemes in a minicircle creates two plectoneme tips. If the positioning of one tip is due to bubble formation, and the positioning of the second tip is as a consequence of

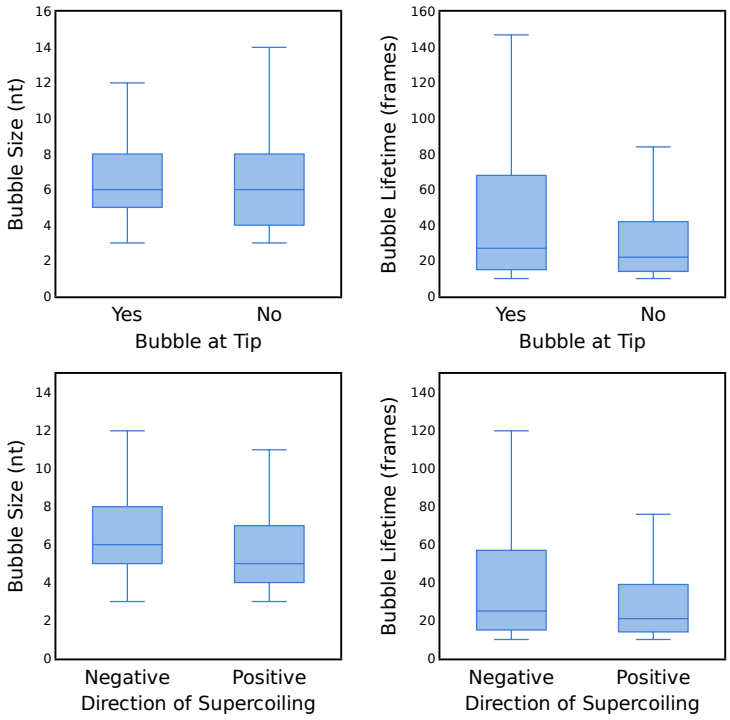

Figure 3.11: Bubbles at the tip of the plectoneme are shorter and longer lived than those nucleated elsewhere, whilst bubbles under negative supercoiling are larger and longer lived than those under positive supercoiling.
Box plots of the bubble sizes **(left)** and bubble lifetimes **(right)** compared across location **(top)** and across direction of supercoiling **(bottom)**. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There are a possible total of 2000 frames. For comparison by bubble location, a bubble was considered to be at the tip of the plectoneme if it was within 20 nucleotides of the two base pairs with the maximum average distance from the centre of mass of the DNA over the frames in which it exists. For comparison by direction of supercoiling, the data was plotted for all sequences at all superhelical densities of the corresponding supercoiling direction. For clarity, the outliers (values $> Q3 + 1.5IQR$) have been removed from the plots, where Q3 is the 75th percentile of the data, and IQR is the interquartile range, or the range between the 75th and 25th percentiles. The full data can be seen in fig. A.5.

the positioning of the first and not bubble formation, this may somewhat confound the analysis. Furthermore, the size of the bubble generally depends on the degree of deformity it induces (fig. 3.12). For example, sharp bends are often accompanied by a significant deviation from canonical structure over a few base pairs whilst larger bubbles result from locally unwound segments and often retain some base stacking. The latter tends to aid in smoother bends and flexible regions positioned away from the plectoneme tip, whilst generally, sharp bends and small bubbles aid the formation of the highly curved apex. As the global conformation of the supercoiled minicircles are sequence-independent, this will be true for all sequences. Thus, the relative size of the smaller, plectoneme positioning bubbles may still vary with sequence and drive the pinning effect, in agreement with both sets of results.

Primarily, the role of bubble formation is to absorb twist and dissipate the torsional strain as they form preferentially over the higher-energy increase in writhe [40]. Additionally, bubbles act

Figure 3.12: Shorter, significantly deformed bubbles form at the plectoneme tip whilst longer, locally unwound bubbles form elsewhere.
A snapshot of the Control339 sequence at $\Delta Lk = -6$ with zoomed in inserts of two different bubbles. In the enlarged inserted structures, the two strands of the bubble are coloured separately. The top bubble is positioned away from the plectoneme tip and is a long, locally unwound section. The bottom bubble is formed as a result of sharp bending at the apex and is relatively shorter.

as a flexible hinge to facilitate formation of the sharply bent plectoneme tip. The larger and more stable nature of bubbles under negative supercoiling improves their ability to absorb twist in this regime, in line with previous results [106]. The degree of twist that a bubble absorbs is proportional to the bubble density in that region (fig. 3.13). Their increased lifetime at the tip of a plectoneme reinforces their ability to pin a plectoneme, with the combined effect of reducing strain and increasing flexibility at the highly bent plectoneme tip. Despite the reduced stability in implicit solvent and the overestimation of twist and bubble formation, the simulated

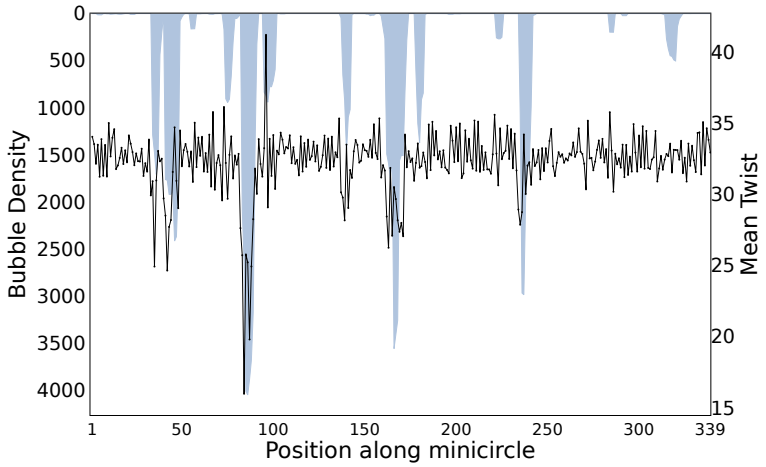

Figure 3.13: The magnitude of twist absorption by a bubble correlates with the relative bubble density in that region.
The twist (black) averaged per base pair for the Control339 sequence over all simulations under negative supercoiling. The bubble density (blue, filled) is summed for the same simulations.

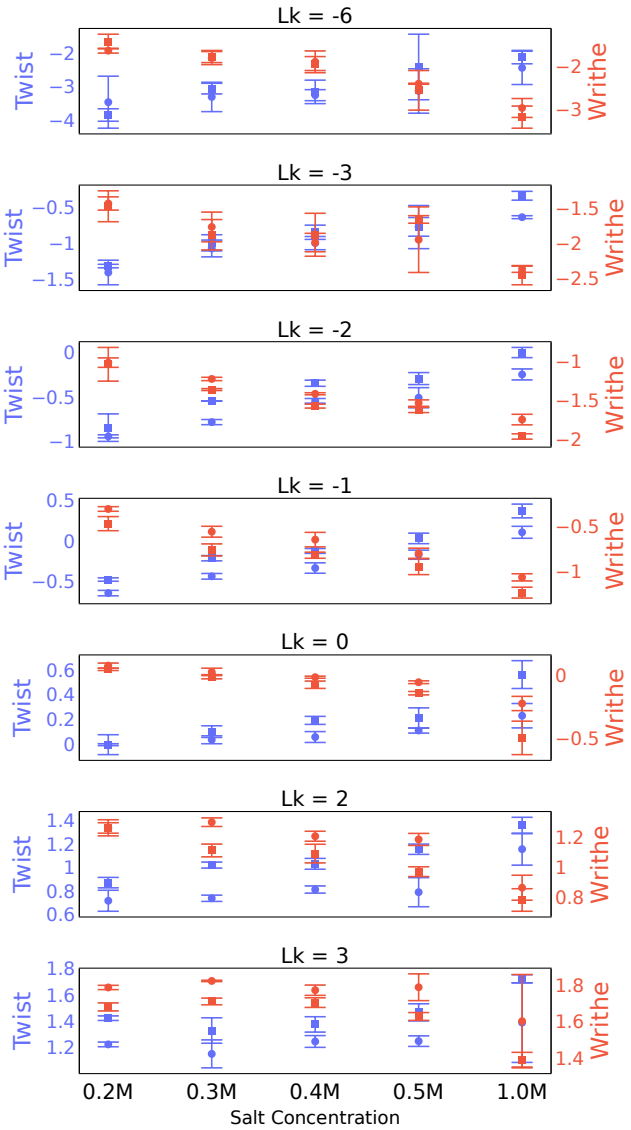

Figure 3.14: As the salt concentration increases, the twist becomes more positive with a complementary decrease in writhe.
The twist (blue) and writhe (red) averaged over the simulations of Mod339 (square) and Control339 (circle) at each superhelical density for each salt concentration. The error bars represent the standard deviation of the twist and writhe over the three simulation repeats of each sequence at each superhelical density and salt concentration.

bubble properties remain in line with the literature and experimentally determined properties.

Salt

Our results into the investigation of bubble properties revealed that bubble size may be a key driver in the strength of plectoneme pinning. As bubble formation is directly influenced by twist, to investigate this hypothesis further, we studied the effect of changing the simulation conditions and, thus, the overall twist of the DNA, on the pinning propensity. We evaluated the effect of increasing the salt concentration and the temperature as these are both known to influence helical structure. We would expect that there is a greater degree of bubble opening

upon increases in temperature and at lower ionic strength [117].

In agreement with current literature (see section 3.1), upon increasing the salt concentration, the sharp bending within the plectoneme is increasingly stabilised and the writhe becomes more negative. This was tested across both the Control339 and Mod339 sequences under both positive and negative supercoiling (fig. 3.14).

The literature states that, due to less of the torsional strain being stored as twist, bubbles are less able to form in order to absorb that twist [117]. This means that, in our simulations, the pinning landscapes in the negatively supercoiled regime are more in line with the Dekker predictions of plectoneme density under positive supercoiling conditions (fig. 3.15). This is more obviously seen in Control339, as Mod339 was designed to have a similar pinning landscape across both the positive and negative directions of supercoiling (fig. A.6). In Control339, the pinning density peaks in the negative supercoiling regime shift as salt increases until they are more in line with the Dekker predictions (fig. 3.15). Furthermore, the absolute P values move in line with that of the positive supercoiling regime. As Control339 has a higher pinning propensity under positive supercoiling, the P value increases as salt concentration increases. Conversely, Mod339 has a significantly higher pinning propensity under negative supercoiling and, therefore, the P value decreases as the ionic concentration increases.

As the writhe becomes more negative, the conformations accessible to the minicircle change (fig. 3.16). Increased writhe is coupled with more turns in the helix and a more highly strained and compact global conformation. The trend between superhelical density, salt concentration, and minicircle conformation appears qualitatively linear—as the salt concentration increases, the minicircle at the given superhelical density tends to adopt the conformation representative of the higher superhelical density at lower ionic concentration. This is in agreement with the linear trend between increasingly negative writhe and increases in salt concentration (fig. 3.14).

Qualitatively—through visual inspection of the polar plots—the shift in the plectoneme positioning landscape agrees the literature evidence surrounding decreased bubble formation at increased salt concentration [117]. Thus, we sought to quantitatively confirm this link. For each value of σ, there was no statistically significant ($p > 0.05$) difference in the number of bubbles formed at each salt concentration. There was some statistical difference in the size of the bubbles formed ($p < 0.05$ for $\Delta Lk = -2, -1, 1, 2$); however, inspection of the box plot (fig. 3.17) and Dunn's post-hoc analysis reveals no clear trend. Due to the large sample of data, small differences can result in significant p-values, and we propose that this has occurred here. Visualisation of the data suggests that there is no practical, biological effect of salt concentration on bubble size. There is a small increase in bubble lifetime as a result of increasing ionic concentration; however, the effect is small ($H = 9.73; p = 0.045$).

These results are not completely unexpected as the writhe of the minicircles increases with increasing salt concentration to afford conformations similar to that at higher superhelical densities. Previously, we found that bubble lifetime, but not bubble size, is dependent on

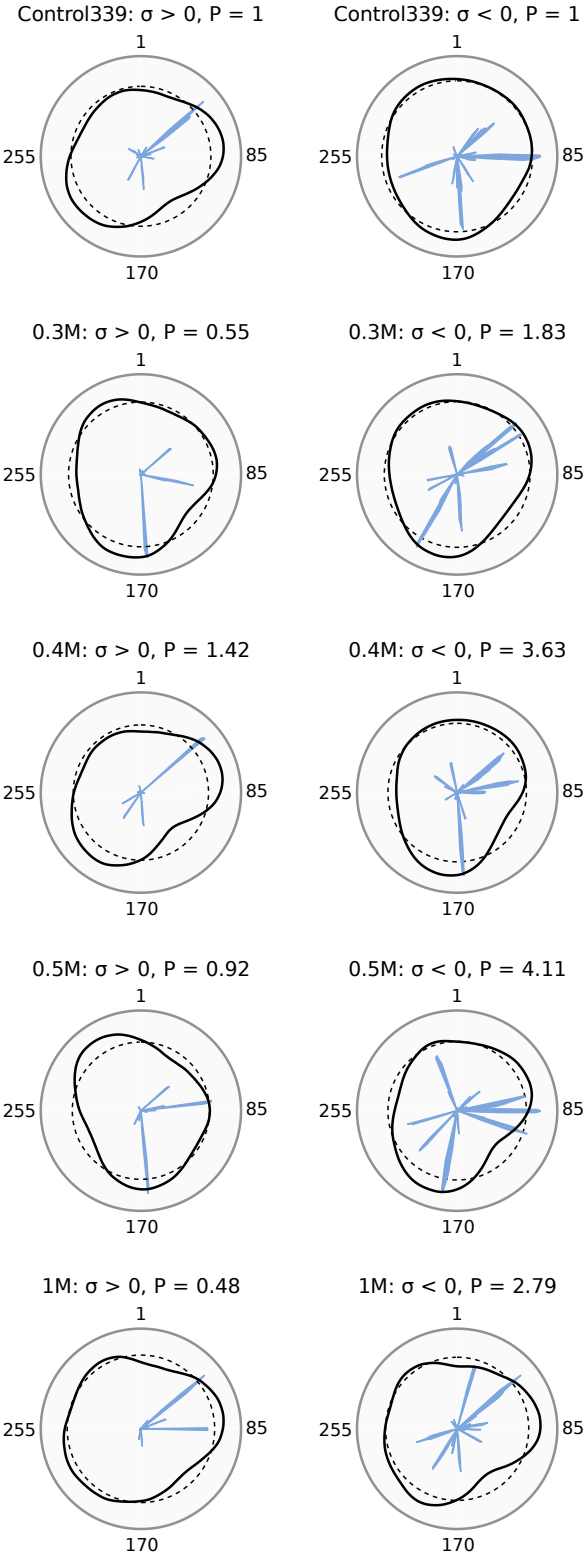

Figure 3.15: As salt concentration increases, the pinning landscapes under negative supercoiling become more in line with that under positive supercoiling.
The plectoneme propensity (black), average plectoneme propensity (black, dashed) and bubble density (blue, filled) of the Control339 sequence under positive **(left)** and negative **(right)** supercoiling at increased salt concentration. The bubble and plectoneme densities are normalised such that the maximum value for each parameter is 1.

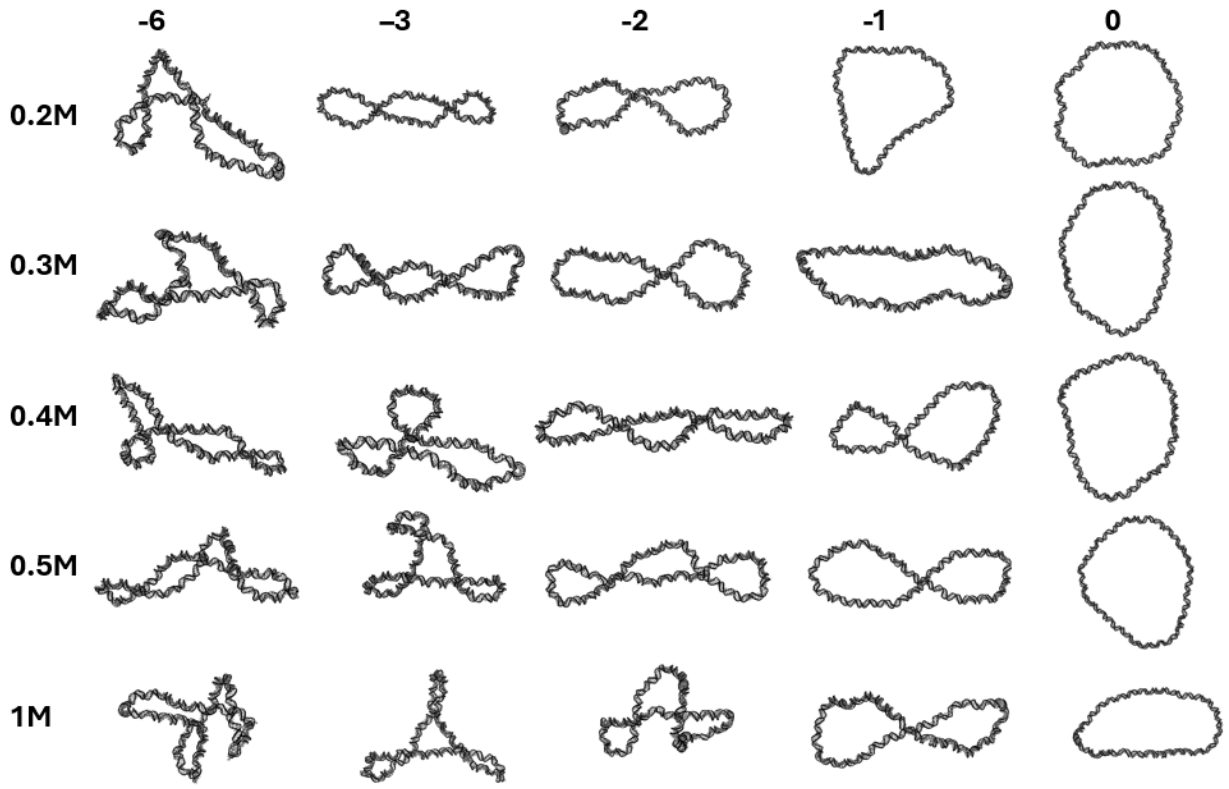

Figure 3.16: As salt concentration increases, the writhe increases and the minicircles adopt different conformations.
Representative snapshots from the simulations at each superhelical density and salt concentration of the Control339 construct.

the superhelical density of the DNA (fig. 3.10). This, combined with the increased stability afforded from the shielding of the backbone, it stands to reason that the stability of the bubble increases with increasing ionic strength. However, these results indicate that the changes in the plectoneme pinning landscape upon increased salt concentration are not as a result of changing bubble properties. We propose that changes in plectoneme pinning propensity are mediated by the increase in writhe. One hypothesis in line with the results presented here is that, upon increasing writhe, more sharply bent regions form and the favourability of this is dependent upon the flexibility of the DNA sequence. A more highly pinning sequence under positive supercoiling may, therefore, elicit more favourable highly-writhed conformations at high ionic concentration.

Our simulations indicate that bubble properties are not greatly affected by salt concentration, despite the clear dependence of twist on the conditions. This directly disagrees with previous literature that shows a decreased probability of bubble opening upon increasing ionic strength [117]. However, the literature investigated this effect on plasmids on the order of ∼2000 bp. Thus, our significantly smaller minicircle may not be able to increase the writhe sufficiently in order to supress the same degree of bubble formation seen in the experiments. Alternatively, implicit solvent may not be sufficiently able to capture the nuances of salt effects. Nevertheless, this does provide us with further insight into the role of bubbles on plectoneme positioning.

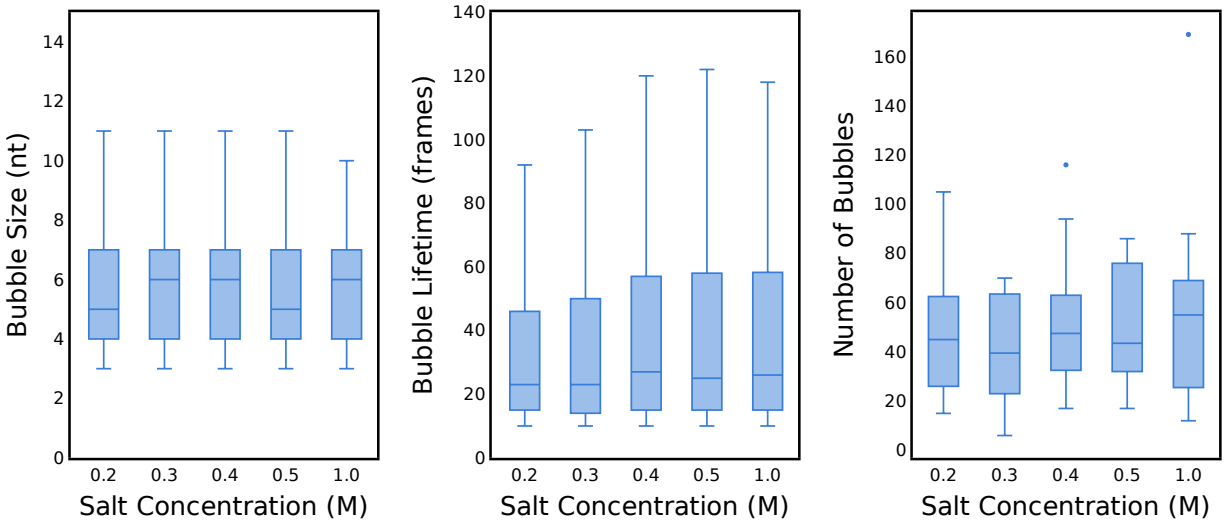

Figure 3.17: Increasing salt concentration has little effect on bubble properties, despite directly influencing twist.
Box plots of the bubble sizes **(left)**, bubble lifetimes **(middle)**, and bubble count **(right)** compared across salt concentration. The data is combined for both Control339 and Mod339. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There are a possible total of 2000 frames. The number of bubbles is the number of unique bubbles across all simulations. For clarity, the outliers (values > Q3 + 1.5IQR) have been removed from the left and middle plots, where Q3 is the 75th percentile of the data, and IQR is the interquartile range, or the range between the 75th and 25th percentiles. The full data can be seen in fig. A.7.

Specifically, these results suggest that there are other contributing factors than just bubble size to be considered.

Temperature

In the absence of supercoiling, DNA is known to unwind with increasing temperature [404]. This is seen clearly in both the negatively supercoiled and the relaxed simulations, where further unwinding decreases the twist in the same direction as supercoiling.

Under positive supercoiling, twist decreases as expected with compensatory increases in bubble formation. This is shown in the plectoneme density landscapes where the positive supercoiling landscape is similar to that seen in the negative, driven by bubble formation (fig. A.8). This occurs until 330K, where there is a sudden increase in twist (fig. 3.18) and the pinning landscape is similar to that at 300K. This may be due to the decrease in twist being preferentially dissipated by an increasing writhe, which is more favourable than further bubble formation in the unfavourable positive supercoiling regime. Following this, there is the expected trend downwards for twist with increasing temperature, accompanied by further bubble formation driving the pinning landscape. However, the effect is subtle as the change in writhe spans across only approximately 0.15 over the 50 K temperature increase at the most positively supercoiled

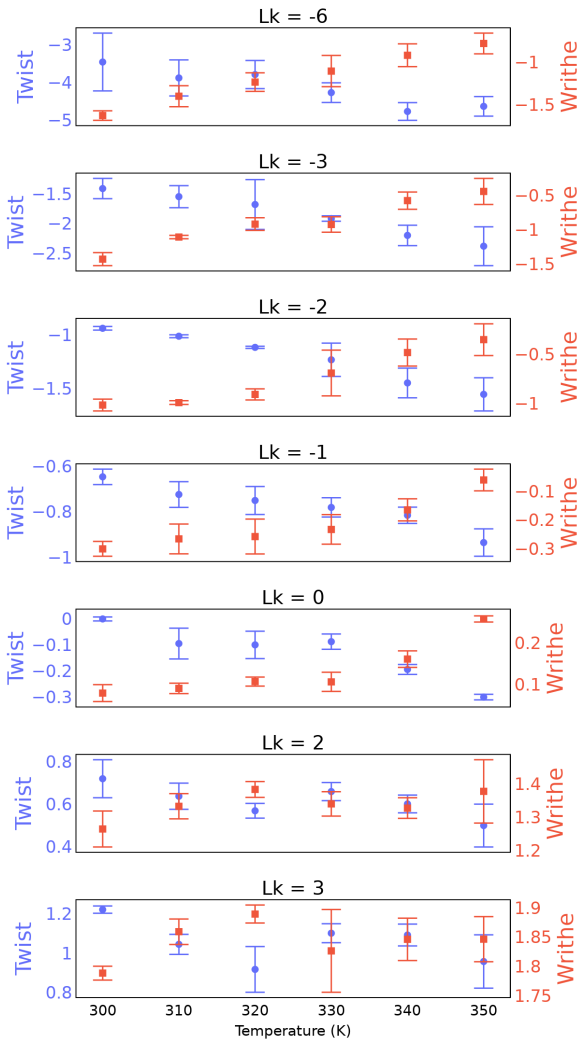

Figure 3.18: As the temperature increases, the twist becomes more negative with a complementary decrease in writhe.
The twist (blue) and writhe (red) averaged over the simulations of Control339 at each superhelical density for each temperature. The error bars represent the standard deviation of the twist and writhe over the three simulation repeats of each sequence at each superhelical density and salt concentration.

topoisomer of $\Delta Lk = 3$. Contrastingly, the corresponding negatively supercoiled topoisomer, $\Delta Lk = -3$, spans a writhe change of 1 over the 50 K temperature increase.

As the temperature is increased, under negative supercoiling, the minicircle stores significantly more torsional strain as twist. This is coupled with increased bubble formation in order to dissipate some of the torsional strain. The result is that the conformation of the minicircle becomes more open, and the pinning propensity decreases as, now, plectonemes do not form (fig. 3.19). The location of the bubbles remains in line with that at physiological conditions as, at increased temperatures, the highest bubble densities are seen in the same bubble-forming regions at 300 K. However, significantly more bubbles must form due to the high strain on the twist of the minicircle and no one region can be said to be "pinning".

Quantitatively, the bubble properties upon increasing temperature are in much better agreement

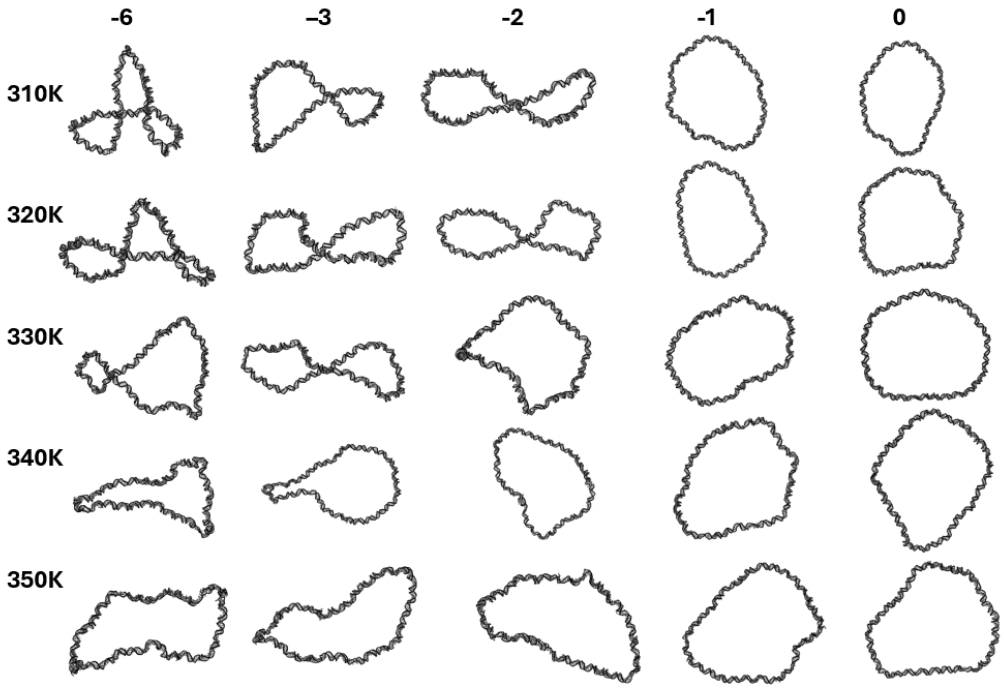

Figure 3.19: As salt concentration increases, the writhe decreases and the minicircles adopt increasingly open conformations.
Representative snapshots from the simulations at each superhelical density and temperature.

with the literature. Previously, it has been shown that there is an increased probability of bubble opening as temperature is increased [117]. Here, the size of the bubble increases linearly with temperature ($H = 654; p = 3.33 \times 10^{-139}$), as does the number of bubbles formed ($H = 28.7; p = 3.2.66 \times 10^{-15}$). There is no effect on the stability of the bubbles ($p > 0.05$) (fig. 3.20).

In general, under both increased salt and increased temperature conditions, the conformation of the minicircles under positive supercoiling remains largely unchanged (fig. A.9). The ability of positive supercoiling to buffer against changes in twist, both increases and decreases, may be linked a biological role: thermophilic bacteria maintain their genome in a positively supercoiled state to prevent thermally-induced denaturation in their high-temperature environment [38, 39]. Here, given the link between DNA structure and function, the ability to maintain a positive supercoil and resist global conformational change at increased temperatures is crucial. The temperatures studied here are physiologically relevant to these thermophiles as they can be broadly categorised depending on their optimal growth temperature: simple thermophiles (50-64 °C; 323-337 K), extreme thermophiles (65-79 °C; 338-352 K), and hyperthermophiles (>80 °C; >353 K) [431].

Overall, the effect of salt and temperature on DNA modelled with implicit solvent has been validated against trends seen in the literature. In general, increasing salt concentration increases the writhe and increasing temperature increases the twist; however, the effect of this on pinning propensity varies depending on sequence due to the twist-absorbing nature of denaturation

bubbles, which themselves are sequence-dependent. This is the first description, to date, of the effect of changing twist on the pinning strength of DNA.

3.8 Validation Against Single-Molecule Experiments

Due to the Dekker paper only considering positive supercoiling, a smaller, 126 bp minicircle was simulated in order to experimentally validate the ability of all-atom implicit solvent simulations to replicate sequence effects in the negative supercoiling regime. It was simulated in both the relaxed and $\Delta Lk = -1$ conformations. At the low superhelical density in such a small minicircle, it does not writhe and instead forms a teardrop shape with two regions of relatively high curvature. The simulations were run at five different salt concentrations and as expected, as the salt concentration is increased, there are less bubbles formed (fig. 3.21) and an increasingly negative writhe (fig. 3.22) due to the favourable shielding of the high curvature tips by the salt and decreased torsional stiffness. The lower twist partitioning increases the localisation of bubbles and thus increases the pinning propensity.

Increasing the salt concentration and, thus, the writhe, changes the conformational space sampled by the minicircle (fig. 3.23). Furthermore, inducing writhe in relaxed topoisomers allows bubble formation even at $\Delta Lk = 0$. At 0.2 M salt, the apex of the teardrop or collapsed conformation is a smooth bend. As the salt concentration increases to 0.3 M, the smooth bend is sometimes increased to a sharp kink. At a concentration of 0.4 M and above, the smooth bend/teardrop conformation is no longer seen. Here, the sharp kink dominates along with

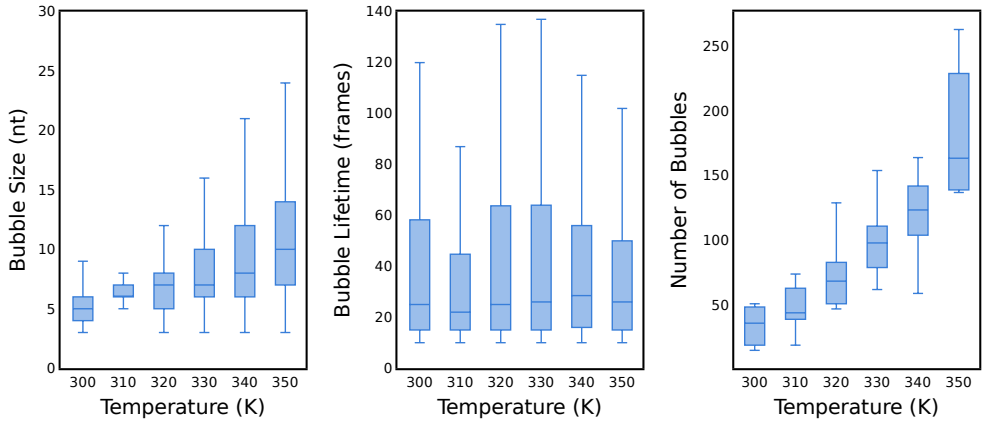

Figure 3.20: Increasing the temperature increases the size and number of bubbles formed. Box plots of the bubble sizes **(left)**, bubble lifetimes **(middle)**, and bubble count **(right)** compared across temperature. The data is combined for both Control339 and Mod339. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There is a possible total of 2000 frames. The number of bubbles is the number of unique bubbles across all simulations. For clarity, the outliers (values > Q3 + 1.5IQR) have been removed from the plots, where Q3 is the 75th percentile of the data, and IQR is the interquartile range, or the range between the 75th and 25th percentiles. The full data can be seen in fig. A.10.

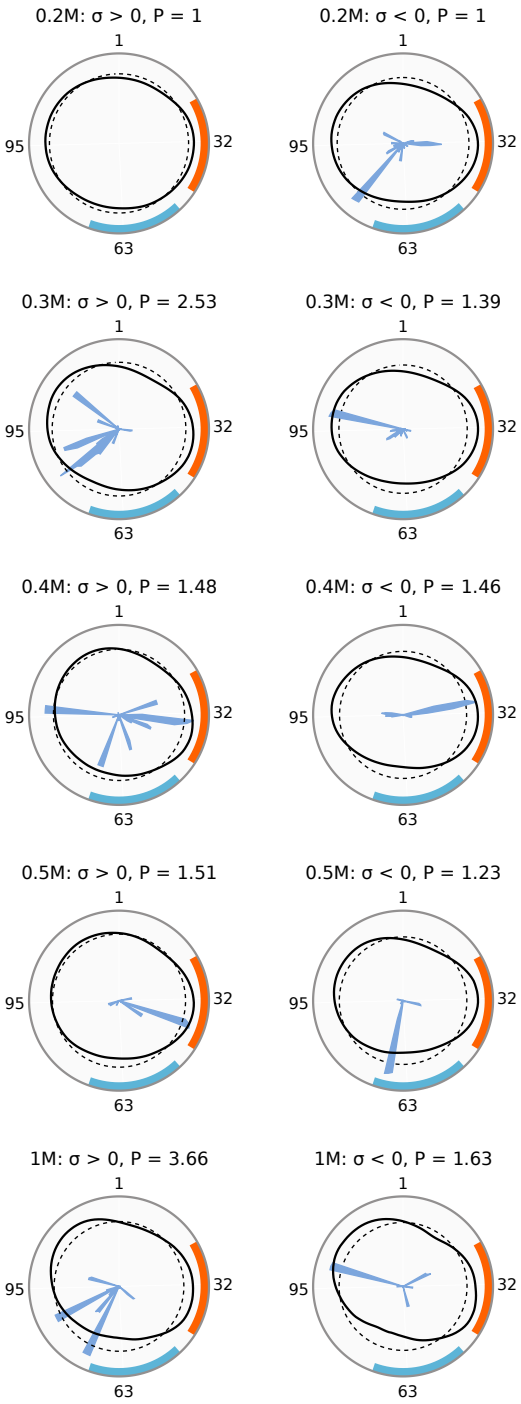

Figure 3.21: As salt concentration increases, the pinning propensity generally increases as bubble formation becomes more localised.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the 126-mer in the relaxed **(left)** and negatively supercoiled **(right)** states at increasing salt concentration. The densities are normalised such that the maximum value for both parameters is 1. The position of the pin dye pair (orange) and control dye pair (blue) are displayed on the polar plots.

plectoneme-like conformations. The latter appears to writhe from one angle; however, a second angle reveals that the strands do not completely intertwine. At 0.5 M, multiple sharp bends can form to afford a square-like structure. At 1 M, this square-like structure exists with significant

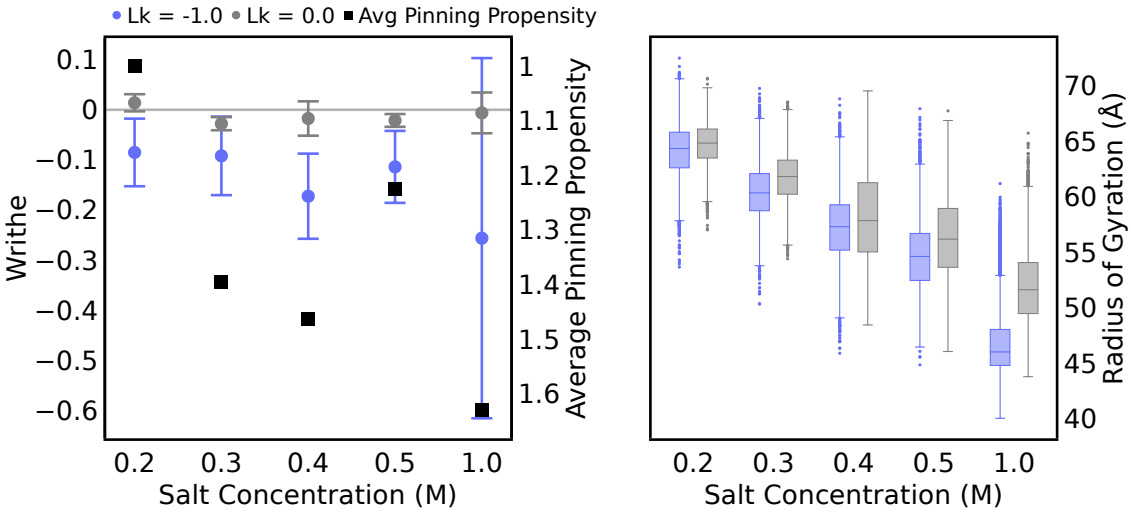

Figure 3.22: As salt concentration increases, the writhe, radius of gyration, and pinning propensity of the 126-mer generally increases.
Left: The writhe (left axis) of the relaxed (grey, circle) and negatively supercoiled (blue, circle) topoisomers is plotted. The error bars represent the standard deviation of the writhe over the corresponding simulations. The pinning propensity (black, square) calculated over all simulations of the corresponding salt concentration is plotted on the right, inverted y-axis such that the trend follows the same direction as the writhe. **Right:** Box plots of the radius of gyration of the relaxed (grey) and negatively supercoiled (blue) topoisomers over the corresponding simulations at increasing salt concentration.

deformation at the two corners, along with the sharp bend-containing-teardrop conformation, and a fully writhed plectoneme. This progression in deformation is likely the mediator of the increased pinning propensity as the stronger bubble localisation is enforced by the higher degree of bending at the apex as the salt concentration increases. In the relaxed topoisomer, increasing the salt concentration shields the backbone and increases the compaction of the minicircle (fig. 3.22) leading to a teardrop shape similar to that under negative supercoiling at physiological salt conditions.

Due to experimental constraints, only the $\Delta Lk = -2$ topoisomer is able to be synthesised; however, it is too strained to be simulated with the current methodology. However, in this sequence, as salt and writhe increase, the pinning propensity increases. Therefore, we would expect the supercoiled constructs at $\Delta Lk = -2$ to behave in a similar fashion to the $\Delta Lk = -1$ construct at high salt conditions, and to have a strong pinning propensity.

Two pairs of labelling positions were chosen on the minicircle to evaluate the pinning strength in single-molecule experiments. The first pair is positioned across the segment that preferentially locates to the collapsed tip, named the pin dye pair, and the other labelling pair is placed in the region that preferentially locates to lower curvature. We have called this our control dye pair. They were additionally positioned to be the same distance apart in terms of the number of base pairs. smFRET measurements were taken for both the relaxed and $\Delta Lk = -2$ topoisomer. At 200 mM KCl, the FRET efficiencies—judged by the median of the FRET histograms—are similar across all four constructs. The histograms are centred around 0.2 in all

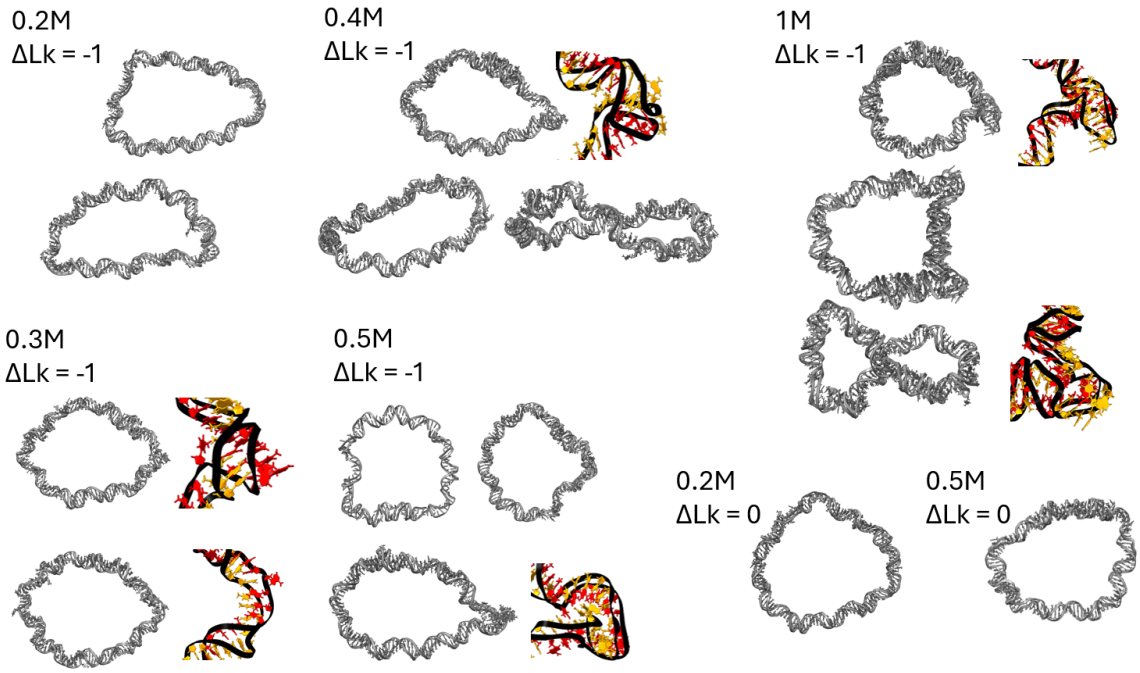

Figure 3.23: As salt concentration increases, the kink at the apex of the 126-mer becomes sharper until, at high salt, there are multiple sharp kinks or the possibility of writhe. Snapshots from the simulations of the 126-mer at increasing salt concentration. The most dominant conformations are shown for each salt concentration. Where there are two snapshots of the overall conformation side-by-side, this is the same conformation/snapshot from two different perspectives. Zoomed-in inserts are shown of selected apices. Here, the bases of each strand are coloured separately for clarity.

cases with a slightly larger tail seen for the pinning dye pair under negative supercoiling. This tail corresponds to a shorter distance between the dye pairs, signifying the apex of the collapsed minicircle, and is not present in the control dye pair. Upon increasing the salt concentration to 1800 mM KCl, the FRET histograms remain centred around 0.2; however, the pinning dye pair under negative supercoiling now has a much more prominent tail. This increase in population of lower distance between the two dyes is not seen in the relaxed topoisomer and is in line with the expected increase in pinning propensity as seen in our simulations.

The results from the FRET experiments at 200 mM under no torsional strain are in qualitative agreement with the results from our simulations: at 0.2 M, the predicted FRET efficiency is similar for both the control and pin dye pairs in the absence of supercoiling. Quantitatively, they differ by only 0.1 in favour of the control dye pair both experimentally at 200 mM KCl and in the simulations at 0.2 M ionic concentration (table 3.24). At increased salt, the FRET efficiencies measured at 1800 mM KCl and 0.5 M ionic concentration are also in qualitative agreement, with the control region having a FRET efficiency of only 0.05 and 0.03 higher than the pin region, respectively. Our simulations correctly predicted that these FRET efficiencies would be similar. Under negative supercoiling, the experimental FRET efficiencies for both the control and pinning dye pairs are also similar. This was not seen in our simulations where there was a significant decrease in FRET efficiency upon supercoiling in the pin dye pair. However, this is not unexpected as the simulations are of the $\Delta Lk = -1$ topoisomer, and the FRET

efficiency is measured at $\Delta Lk = -2$.

The difference plots (fig. 3.25) reveal a consistent shift toward lower FRET efficiencies for both dye pairs in the absence of torsional strain. At low salt concentrations, the overall conformations of the minicircle—both under negative supercoiling and in the relaxed topoisomer—are more rounded. As the salt concentration increases, the overall conformation of the relaxed minicircle resembles the more collapsed or teardrop-like shape seen under negative supercoiling at low salt. Here, linear regions resembling canonical B-DNA are adopted along the 'sides' of the teardrop to aid the sharper bending at the apex (fig. 3.23). These linear regions have a lower FRET efficiency compared to the smooth global curvature seen in the more rounded minicircle. This is reflected in the FRET histograms of both dye pairs in the relaxed topoisomer. The same effect is seen for the control dye pair under supercoiling; however, here, the effect is only subtle.

From the difference plots, the greatest effect for the supercoiled control pair is the decrease in mid-FRET population and complementary increase in high-FRET population. This high-FRET population is low in density yet is broader and higher in FRET efficiency than the peak corresponding to the apex in the supercoiled pinning dye pair. However, this effect is only notable in the difference plots and is lost in the raw FRET histograms, again signifying only a subtle effect. This suggests that the high-FRET population is likely due to bubble formation in this region. This is not unexpected as, in our simulations at 0.5 M salt concentration, bubbles form between the positions of the control dye pairs. This suggests that the conformation of the $\Delta Lk = -2$ topoisomer is structurally similar to that of the $\Delta Lk = -1$ at 0.5 M ionic concentration. This is further supported by the salt-dependent writhes measured for both the Control339 and Mod339 sequences (fig. 3.14). Here, at $\Delta Lk = -1$, the writhe becomes increasingly negative as salt concentration increases until, at 0.5 M ionic concentration, the writhe reaches ∼-1. This is the equivalent writhe for the $\Delta Lk = -2$ construct at 0.2 M salt concentration. If we assume the same magnitude of twist decrease upon the same increase in

Table 3.24: ***The simulations well-predict the trends seen experimentally using smFRET and suggest that the*** $Lk = -2$ ***topoisomer is conformationally similar to the*** $\Delta Lk = -1$ ***topoisomer at 0.5 M ionic concentration.***
The experimental FRET efficiencies of the control and pin dye pairs in the relaxed and negatively supercoiled states. The experimental FRET efficiencies were measured with 200 mM and 1800 mM KCl at $\Delta Lk = -2$*. The simulations in implicit solvent of the supercoiled topoisomer were at* $\Delta Lk = -1$*.*

		Experimental		Simulation	
Linking Number	Dye Pair	200 mM	1800 mM	0.2 M	0.5 M
Relaxed	Control	0.22	0.21	0.35	0.60
	Pin	0.21	0.16	0.34	0.57
Supercoiled	Control	0.20	0.20	0.47	0.74
	Pin	0.21	0.21	0.29	0.71

salt concentration across the two minicircle sizes, this may indicate that the minicircle does not fully writhe at $\Delta Lk = -2$ under low salt, as it does not do so at $\Delta Lk = -1$ and 0.5 M ionic concentration. This is further supported by the increased strain writhing would place on the 126-mer compared to the 339-mer. This would also explain why there are only subtle differences in FRET efficiency at 200 mM KCl between the relaxed and supercoiled constructs, and between the control and pinning dye pairs. In the absence of writhe, the conformational differences between these labelling positions are minimal. This is reflected in the predicted FRET efficiencies of the two supercoiled dye pairs from the simulations at 0.5 M ionic concentration being similar in magnitude to one another (table 3.24).

Across all conditions and superhelical densities, the most prominent change observed in the

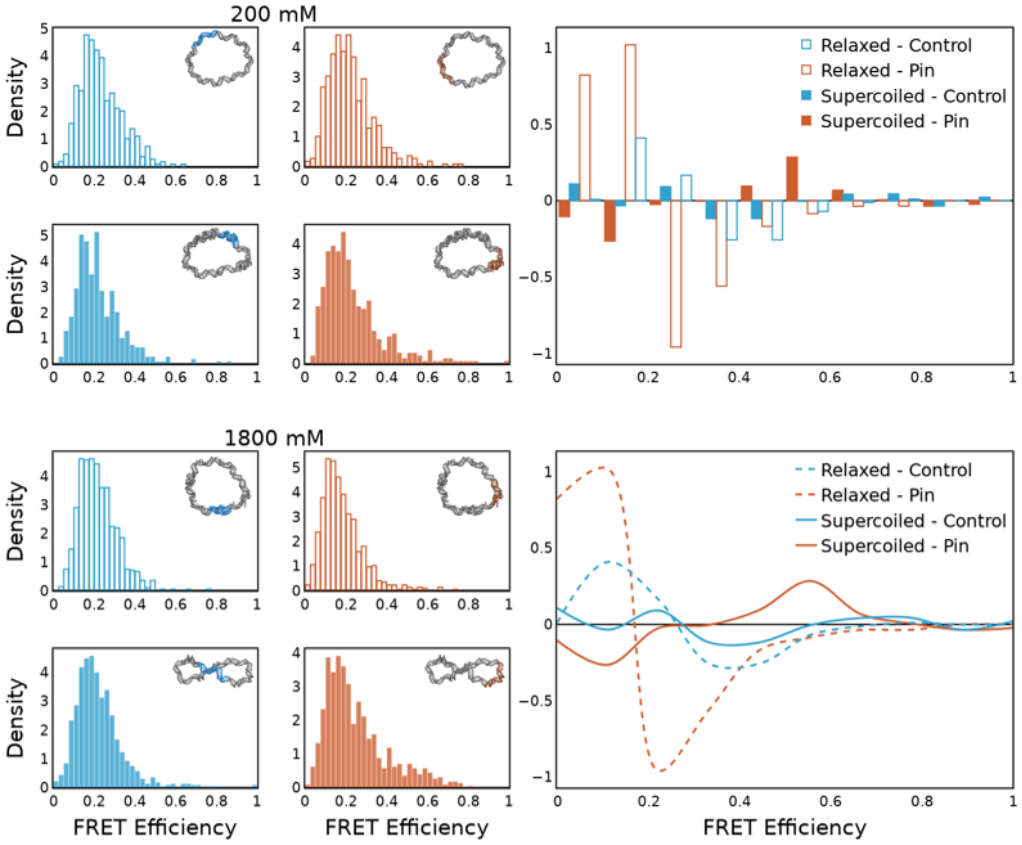

Figure 3.25: As salt concentration is increased, the FRET histograms show a larger high-FRET population for the dye pair under supercoiling in line with the expected stronger pinning propensity from our simulations.
Top 2 rows, left: FRET histograms at 200 mM KCl for the control **(left)** and pin **(right)** dye pairs at $\Delta Lk = -1$ **(right)** and $\Delta Lk = 0$ **(left)**. **Bottom two rows, right:** The FRET histograms are shown for the same constructs at 1800 mM KCl. The labelling positions are shown on a representative minicircle structure. The pinning region is shown in orange, and the control region in blue. **Right, top:** A difference plot of the difference in the FRET histograms at 1800 mM KCl compared to 200 mM KCl. The control is plotted in blue, and the pin dye pair is plotted in orange. For both, the relaxed topoisomer is plotted as an outline whilst the supercoiled topoisomer is plotted as a filled bar. **Right, bottom:** The same difference plot is plotted as a curve. The control is plotted in blue, and the pin dye pair is plotted in orange. For both, the relaxed topoisomer is plotted as a dashed line whilst the supercoiled topoisomer is plotted as a solid line.

raw FRET histograms is the expected increase in high-FRET population in the pinning dye pair upon supercoiling. This effect becomes more pronounced at higher salt concentrations. At these concentrations, such as 1800 mM KCl, it is likely that the minicircle becomes sufficiently stabilised by the increased screening by the high ionic strength to accommodate writhe. Under these conditions, the marked increase in the mid-to-high FRET population observed for the pinning dye pair may reflect the formation of the highly curved plectoneme tip. This is coupled with bubble localisation to the control dye region, evidenced by the increased density of the high-FRET population. However, because increasing writhe is coupled with a compensatory decrease in twist, there is only a modest increase in FRET signal from the control region bubbles. It is possible that in minicircles of this size, the writhed state is not fully stabilized, and the DNA continues to fluctuate between plectonemic and alternative conformations—explaining the continued dominance of the low-to-mid FRET populations.

Due to the random nature of photon emission and detection events, there is an inherent variability in the number of detected photons within a discrete time period. This time period is called a 'burst' and is partitioned such that each burst contains a consistent number of consecutive photons. This has the effect of broadening a FRET distribution such that there is an expected standard deviation due to this shot noise. If a molecule is static within the timeframe of a burst, the observed distribution of the resulting FRET efficiency will be broadened only by shot noise. Thus, we can compare the standard deviation of each burst to that expected, whereby a larger standard deviation would signify a more dynamic species. This is known as burst variance analysis (BVA) and was performed on all of our minicircle constructs (fig. A.11). In general, all constructs were more dynamic at an increased ionic concentration; and, at each concentration, the supercoiled constructs were more dynamic than the relaxed topoisomers. This is in line with expectations from our simulations, as increasing the ionic concentration increases the writhe partitioning, which, in turn, would increase overall dynamics and conformations accessible to the minicircle.

The Dekker algorithm did not predict any particular region to be of high plectoneme density in this construct (fig. A.12). Likewise, the SIDD algorithm had a very broad prediction of bubble localisation with two main peaks corresponding roughly to the regions of highest average distance from the centre of mass. However, both Dekker and SIDD predictions were of low density in the 110-25 bp region, and especially across the first 25 nt. Sequence dependence in the minicircle may, therefore, arise due to the preference of positioning this region in a position of low curvature and bubble formation, rather than for a strong preference for any particular segment for the tip. The low intrinsic curvature here is somewhat masked by spikes in the curvature plot (fig. A.12) as these are the two ends of the DNA in the linear construct that are, therefore, subject to end effects; however, there is a higher intrinsic curvature in the middle region of the linear construct consistent with these predictions. Overall, this has the effect that plectoneme positioning and bubble formation is not disfavoured across much of the minicircle. As high FRET efficiencies are indicative of regions of the DNA being in close proximity, which

itself can describe the apex of the minicircle, or highly kinked regions or bubbles, combining experiment with simulation is crucial to understanding the nuances of the experimental results.

3.9 Discussion

In this chapter, we have validated the use of all-atom implicit solvent simulations for studying the sequence-dependence of plectoneme positioning in DNA minicircles. We have shown, for the first time, that simulations perform better than existing algorithms in predicting the location of plectonemes and exemplified the role of denaturation bubbles in plectoneme positioning. To date, no one algorithm exists than can combine the effects of intrinsic curvature, flexibility, and bubble formation to predict plectoneme localisation across both directions of supercoiling. Thus, we have proven the ability of this simulation workflow to replicate experimental results on plectoneme density and salt and temperature effects. We have expanded this further to show the ability of positively supercoiled DNA to buffer against changes in twist, how changes in twist mediate plectoneme pinning strength via bubble formation, and we have extended the current understanding of bubble mechanics to describe their size-driven influence on plectoneme positioning.

We have used a small 126 bp minicircle as a proof-of-concept for a novel system that can be probed experimentally, using single-molecule FRET, to study DNA supercoiling. Minicircles of this size are limited in that, at least at physiological conditions, they don't seem to writhe. Thus, the pinned region is not a well-defined plectoneme and is instead a segment of increased curvature into a tear-drop shape. The FRET histograms, therefore, show only a subtle conformational change in line with the results from our simulations. This is further affected by the "control" labelling position being only 5 nt away from the "pin" labelling position, a restriction necessary due to thymine positioning within the sequence and the size of the minicircle. The other inherently pinned region, due to the double-pin nature of the minicircle, is only approximately 35 nt away from the "pinned" dye pair. Choosing labelling positions required a balance between measuring over a distance large enough to measure a conformational change, and small enough that we were able to place two dye pairs on a relatively short stretch of DNA. Furthermore, the circular nature of the DNA effectively limited labelling positions to half of its total length. The proximity of the "control" labelling position to the apex means that there is often a slightly increased curvature here to accommodate the larger bend only a few base pairs away. The $\Delta Lk = -2$ topoisomer was studied in these experiments, which was too strained to be simulated using the current protocol. We would expect that this would not change our results, especially as simulations in the 339 bp minicircle show to be consistent across different superhelical densities under the same direction of supercoiling; however, further work would be needed to verify this. This is especially important as, although we have provided evidence of the proposed conformation of the minicircle at $\Delta Lk = -2$, we do not have confirmation of the experimentally observed structure. Nevertheless, we have presented, for the first time, a novel biophysical

methodology to probe the sequence-dependent conformational dynamics of DNA under torsional strain.

Plectoneme positioning in our simulations is driven by cooperative bending and bubble formation. There is a required balance between the favourability of forming the plectoneme tip in one region, and the inherent pinning of the opposite side, due to the circular nature of our minicircles, that would not be seen in linear constructs. Despite the Dekker algorithm not being designed to model circular DNA constructs, which would not take into account the double-pin effect; and both the Dekker and SIDD algorithms not being intended for sequences of such a small length scale, our simulations are in remarkable agreement with predictions. In using minicircles, we are restricting our simulations to modelling just the plectoneme, rather than the formation of a plectoneme within a longer sequence. Nevertheless, owing to the comparison to the Dekker constructs, we have shown that the results from our simulations can be easily propagated to biologically relevant length scales whilst still benefitting from atomic-level resolution. Studying the relative lack of pinning, such as the comparison between the Flat75 constructs, becomes more complex, as the pinning landscape becomes dominated by the surrounding sequence. Predicting and understanding the stochastic positioning of a plectoneme is inherently more difficult than for a strongly-pinning sequence. The predictions by the Dekker algorithm in the sequence surrounding the two flat constructs are not the same, despite being of the same sequence. Unsurprisingly, this shows that small changes in sequence can propagate to effects on a much larger length scale. This is not seen in the strongly curved sequences where the large intrinsic curvature dominates the landscape, and these effects are lost to the baseline. However, this limitation is likely negligible at biologically relevant length scales and is an artifact of using a minicircle, which forces the formation of the plectoneme no matter the favourability across such a short sequence. Furthermore, the P value of the flat sequences under negative supercoiling are the same despite them having different preferences for plectoneme positioning at the 75-mer. This is a limitation of our P metric that describes the overall pinning propensity of the sequence, but it is not able to describe any directionality of where that pinning lies.

A second limitation of our P metric is its inability to be a relative comparison across widely different sequences. The P value for Strong339, normalised to Control339 = 1, is 5.04 under negative supercoiling. This is less than Mod339, which has a P value of 5.91 under negative supercoiling, despite Strong339 having a higher true pinning propensity. The strongly bubble-driven region pins the plectoneme in every simulation, as determined by the polar plots being positioned at this region for each repeat. This is compared to the two peaks in Mod339 that pin the plectoneme in approximately two-thirds of simulations under negative supercoiling. However, in Strong339, more highly strained structures—such as those with multiple sharp bends—arose. Despite having a well-defined peak and trends in line with a strong pinning propensity, these structures increased the $\bar{x}$ term of the pinning propensity due to having more regions of the minicircle a larger distance from the centre of mass of the DNA. Thus, the P metric did not well-quantify the relative pinning strength of the Strong339 construct in

comparison to Control339 and Mod339. We conclude that our P metric functions as a relative, quantitative comparison of pinning across related sequences, such as those where one sequence has been mutated to afford the other, or where a segment of DNA has been inserted into an existing sequence—such as is the case for our Dekker and Mod339 sequences. Thus the pinning propensity metric is only capable of describing the relative change in pinning propensity upon the sequence modifications where, due to the sequence similarity in the surrounding sequence, there would be a similar $\bar{x}$ value. Nevertheless, comparison between our calculated P values and the Dekker calculated plectoneme density values [41] allows us to infer the biological relevance of our sequences. The predicted plectoneme density of Curve75-1 is approximately 3, and the predicted value for Curve75-2 is approximately 2.5. The predicted values of the plectoneme density upstream of transcription start sites in actual genomes were lower than this: *V. cholerae* was approximately 1.75, *E. coli* was approximately 1.6, and *B. methanolicus* was approximately 1.4. Thus, our weaker pinning Control339 and Mod339 sequences may actually be of biologically relevant pinning strengths on this length scale.

Furthermore, the bubble-forming region in Strong339 has an AT content of approximately 85%. The *plasmodium* species is the causal agent of malaria. It has two genome-containing organelles that are important for its survival: a mitochondrion and an apicoplast. The latter has been identified as one of the most A/T rich genomic sequences to date with a total A/T content of 86.9% [432]. The genome itself is circular—and thus, topologically constrained—and the replication is known to be sensitive to topoisomerase and has been suggested to occur via the formation of loops [433]. The plastid has attracted attention as a therapeutic target to combat malaria [434–436]. Thus, understanding the behaviour of such highly A/T rich sequences within the genome could be crucial, and our Strong339 sequence could provide further insights relevant to this field.

The effect of salt on plectoneme positioning varies by sequence. There is also a different effect between the 339-mers and the 126-mer; however, this may be a consequence of sequence rather than size, which we cannot say with only one sequence being studied on the 126 bp scale. The reasoning behind this is beyond the scope of this work, and an understanding of salt on pinning propensity would require an in-depth study with more sequences than presented here. All statements on these differences are purely speculative and beyond the scope of this project. Furthermore, despite unquestionably proving that implicit solvent is appropriate for the work in this thesis; we are not suggesting that we have validated it for a comprehensive study into salt effects. Salt concentration was, in this thesis, merely a tool to influence twist and bubble formation. However, one explanation for the different salt effects on pinning propensity depending on sequence is that the more negative writhe upon increasing salt affords a balance between tip-bubbles lowering the energetics required to form the more highly strained plectoneme tip, and less bubble formation in order to absorb less of the twist. This shifts the plectoneme positioning landscape more in line with that of the positive supercoiling regime. The effect of this varies across the different sequences, depending on the balance between bubble positioning and

intrinsic curvature in driving the plectoneme positioning landscape. In the case of Control339, the P value generally increased as the sequence was more strongly pinning under the positive regime and bubble formation was not significantly driving the pinning strength. The opposite is true for Mod339 where an increase in salt lowers the P value under negative supercoiling, and the P value for this sequence was smaller under positive supercoiling, suggesting that bubble formation was a key driver of plectoneme positioning in this sequence. This also offers further proof that our P metric is able to quantify relative plectoneme propensity in line with expected trends.

Finally, a minicircle of size 339 bp was chosen for practical considerations as, at the beginning of this project, it was thought that collaborators were able to synthesise this. Unfortunately, this was not the case. A minicircle of this size does not have an integer number of turns; and the relaxed topoisomer when simulated with 32 turns is actually negatively supercoiled with an effective ΔLk of -0.29. The superhelical density, σ, of -0.009 for a minicircle of this size is negligible and should not negatively affect our results.

3.10 Summary and Future Work

In summary, we have shown that all-atom molecular dynamics simulations in implicit solvent are able to capture sequence-dependent plectoneme pinning in line with experimental results. We have described a pinning propensity metric that is able to quantify the relative pinning strengths of similar sequences, and we have used this to show that simulations correctly capture subtle variations in pinning propensity. We have further verified that the simulations replicate expected bubble, salt, and temperature effects relevant to DNA supercoiling in a sequence-dependent manner. We have expanded on current knowledge to present novel results supporting the idea that bubble size a key driver of plectoneme pinning strength. Despite small differences in the size and stability of bubbles across implicit and explicit solvent simulations, the positioning of the bubbles and the resulting plectoneme location remains the same across both. Further, on the timescale accessible by explicit solvent simulations currently, the minicircle stayed locked in one conformation. Therefore, it is essential to exploit both the frictional speed up from implicit solvent as well as the computational speed up afforded by the lack of discrete solvent ions.

We have used our simulations to design a novel, dye-labelled minicircle construct that, as a proof-of-concept, can be studied using smFRET and against which we can further validate results from our simulations. In the future, we should expand on this system in order to study larger minicircles that are capable of writhing. This should create more of a pronounced effect in the results of our biophysical experiments, as well as offer an opportunity to study more exciting properties relevant to DNA supercoiling. We could also refine our simulation protocol in order to simulate the 126-mer at $\Delta Lk = -2$ in line with our current experiment.

It was our intention to use the simulations presented in this chapter to verify that, despite the

limitations of implicit solvent, it correctly captures the necessary properties to study sequence-dependent plectoneme pinning. We can then use the presented methodology and sequences to study the effect of DNA damage as seen in following chapter. However, to this end, we have obtained simulation time on the order of microseconds across a range of DNA sequences and superhelical densities, all with varying properties and features. This, to our knowledge, is the most comprehensive and varied set of simulation data of this nature. Future work could exploit this data to reveal as-of-yet unknown influences driving sequence dependence plectoneme pinning.

4

DNA Mismatches Compete with Sequence to Pin a Plectoneme

In this chapter, I describe the current literature surrounding the sequence- and lesion-dependent nature of mismatch repair efficiency. It has previously been shown that mismatches will preferentially locate to the plectoneme tip. Here I present, for the first time, results that show that the 12 different single mismatches have different propensities for the plectoneme tip and that these are further affected by the direction of supercoiling and the surrounding sequence context. I present insights into what is driving this mechanism by evaluating the pinning propensities and bubble formation of multiple tandem mismatches, and of the single mismatches under different conditions. Based on the hypothesis that the pinning of a plectoneme tip by a mismatch facilitates recognition by MutS, I reveal a novel trend between mismatch pinning propensity and the literature-reported efficiencies of mismatch repair.

4.1 Introduction

The mismatch repair (MMR) pathway is crucial to preventing the development and progression of tumours; however, because it must compete with the replication machinery to complete the repair, it is not fully efficient. Many details of how the pathway works are still not understood.

One such example of this is the varying efficiency of repair across the genome that has been suggested to reflect some aspect of chromosome structure [437]. Using population genetics data to investigate relative MMR rates is non-trivial, as mutation profiles are a combination of non-uniform mismatch generation as well as non-uniform mismatch repair, in addition to other processes [438]. As such, MMR rates can vary from little apparent repair [439], to a greater than 99.999% repair [440].

The efficiency of mismatch repair differs depending on both the nature of the mismatch itself and the surrounding sequence context [295, 441–446]. MMR-seq is a single-molecule mismatch repair-tracking assay used to evaluate sequence dependence on the scale of pentamers [442]. It was found that TT, TC, and AG mismatches had an especially strong sequence-dependent variability of repair efficiency, whilst CC is poorly repaired regardless of local sequence. In the case of TT, the variability of the efficiency of the repair of a specific mismatch was attributed to whether the lesion within that specific sequence context would lead to a change in the encoded amino acid or not. Whilst this is the largest dataset to date, owing to the high-throughput nature of the assay, evidence suggests that DNA sequence exerts an influence on much larger length scales. For example, in the case of CC, it has been shown that the thermodynamic impact of the mismatch can span up to 9 bp [447].

The full mechanism of mismatch recognition and repair, with the associated intricacies surrounding the hypervariable repair efficiencies, is not yet fully understood. The general hypothesis is that MutS binds to a mismatch and forms a kinked initiation complex. The kinking widens the groove around the mismatch, likely to facilitate stacking of MutS's phenylalanine, Phe36, with the flipped, mismatched base (fig. 4.1) [448]. MutS then adopts the sliding clamp conformation,

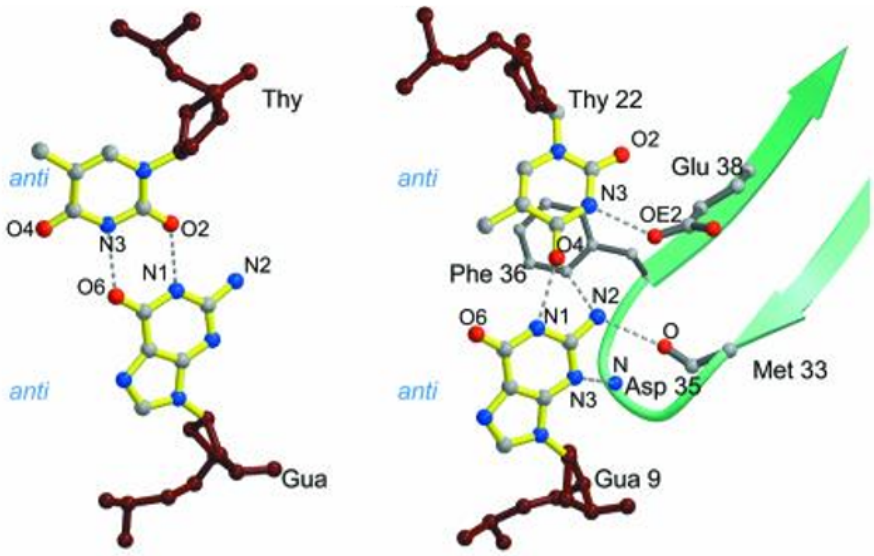

Figure 4.1: The phenylalanine of MutS interacts with a flipped mismatch base to stabilise the protein-DNA complex.
Left: Hydrogen bonding within the G/T *anti/anti* mismatch. **Right:** The interactions between the same mismatch when bound to MutS. Image taken from Natrajan *et al.* [448].

and the DNA unbends and signals for MutL recruitment for mismatch repair [444, 449]. This is supported by evidence that the sliding clamp conformation of MutS interacts with DNA that is not bent [450]. Thus, an overly stable bent state compared to the unbent conformation may stall repair due to low interconversion rates [444]. The strong repair efficiencies are attributed to an increased rate of sliding clamp formation [442, 444] and, especially, an increased rate of association between MutS and the mismatch-containing DNA [444]. The mismatches themselves have differing affinities for MutS and its homologs that can be further influenced by the sequence surrounding the mismatch [451–453]. Highly stable MutS-mismatch complexes are associated with a stable unbent state and a low barrier to the bent-to-unbent transition. The rate of conversion between the two states has been found to be mismatch-dependent [454]. It has been shown that trapping a GT mismatch in an intermediatory bent state at an angle of approximately 15° is sufficient to lose the repair of the mismatch [449].

However, understanding of the pathway is further complicated by the mismatch-dependent kinetics not being in full agreement with the relative efficiencies of mismatch repair. For example, it is well-established that the CC mismatch is the most flexible and has a highly stable kinked state [455]; equally, the CC mismatch has been shown to be poorly repaired in multiple different sequence contexts [441–443] (table 4.2). Conversely, the AC mismatch is also highly stable in the kinked state [455] but has shown to be efficiently repaired across a range of sequence contexts

Table 4.2: *The contributors to mismatch repair efficiency. All citations treated non-symmetrical mismatches and their reciprocal as equivalent, for example, AC equal to CA. Thus, non-symmetrical mismatches have the same values.*

Mismatch	Repair efficiency [442]	Flexibility [455]	Apparent $K_d(\mu M)^a$ [451]
AA	High 0.92±0.07	Mid	Mid 1.0±0.1
AC	High 0.93±0.04	High	High 3.4±0.4
AG	Variable 0.81±0.17	Low	High 4.8±0.9
CA	High 0.93±0.04	High	High 3.4±0.4
CC	Low 0.35±0.13	High	High 6.9±2.4
CT	Variable 0.82±0.19	High	Mid 1.3±0.3
GA	Variable 0.81±0.17	Low	High 4.8±0.9
GG	High 0.94±0.04	Low	Low 0.62±0.07
GT	High 0.93±0.04	Mid	Low 0.19±0.03
TC	Variable 0.82±0.19	High	Mid 1.3±0.3
TG	High 0.93±0.04	Mid	Low 0.19±0.03
TT	Variable 0.75±0.22	High	Mid 1.0±0.1

[a]High/mid/low are with respect to Mid = 1 μM

in *E. coli* [442]. However, as the stability of the kinked state is dependent on the destabilisation of the mismatch relative to the complementary strand, the stabilisation of this conformation would be dependent on the surrounding sequence. Further, a stable kinked conformation does not necessarily imply a slow interconversion rate as the relative stability of the bent and unbent states has not been considered. More work is required to determine the bent-to-unbent transition rates.

Even combining the stability of the bent conformation with affinity of the mismatch for MutS cannot fully explain the variations in repair efficiencies. For example, the AC mismatch has a highly stable kinked state, suggesting a high barrier to the bent-to-unbent transition, as well as a low affinity for MutS. This would indicate that AC has a low repair efficiency; however, this is not the case. Therefore, further factors must play a role in determining the overall repair efficiencies. We suggest that DNA supercoiling is one such factor that must be taken into consideration.

The efficiency and fidelity of repair depend largely on the stage of the cell cycle, with the highest efficiency and fidelity seen in actively replicating cells [456] where DNA supercoiling is especially prevalent due to its role in replication. It has been shown that the areas to which MutL preferentially locates within the genome are AT-rich and have low thermal stability [457], which are the same regions that plectonemes preferentially locate to. The repair efficiency is further affected by the position of the lesion within the genome [437], and the nature of the lesion itself [441]. Accessibility of the mismatch to the MMR machinery has been shown to be important for efficient mismatch repair as DNA positioned at nucleosomes are poorly repaired [458, 459]. This phenomenon is also seen across other repair systems, where, in the BER pathway, lesions closer to or rotated towards the nucleosome are repaired slower than distal or outwardly rotated lesions [243, 244, 460–462]. This is likely due to the fact that MutS encounters DNA via 3D diffusion [463]. The indirect readout mechanism by which mismatch repair proteins recognise a lesion can be greatly facilitated by preorganization - arrangement of the DNA in a way that is similar to the protein-bound conformation [295]. The tip of the plectoneme is very sharply bent and often contains bubbles, flipped bases, kinks, and other defects to accommodate this bend [40, 97, 112, 287, 299, 330, 399]. Location of the mismatch to the plectoneme tip would not only lower the energy barrier to the kink required by mismatch-recognising proteins, but also increase the accessibility of the mismatch as the plectoneme protrudes out of the densely packed genome. In support of this hypothesis, supercoiling is known to enhance MutS binding to a GT mismatch with little effect on the rate of MutS sliding clamp formation, suggesting that supercoiling plays a role in the initial recognition step [289]. However, many of the previously discussed MMR efficiency studies neglect the involvement of DNA supercoiling, which may limit the understanding of both the pathway and the variable repair efficiencies.

In eukaryotes, MutS homologs recruit MutLα endonuclease to nick the lesion-containing DNA strand as part of the repair process. This nick occurs at a distal site, typically up to several

hundreds of base pairs away [464–466]. MutLα has been shown to enhance the relaxation of supercoiled DNA by topoisomerases, and DNA supercoiling increases nuclease activity at high MutLα concentrations [464]. An attractive hypothesis is that as MutLα works across distant sites, supercoiling and the formation of plectonemes bring the separated DNA bases into close spatial proximity to facilitate this. This would also offer an explanation as to why MutLα nicks a linear DNA molecule with less efficiency [467] and provide a solution to the reduced repair efficiency as the distance between the nick and the mismatch increases [468]. Further, MutLα binds supercoiled DNA with a higher affinity than the same linear or relaxed DNA [469], and incises mismatch-containing supercoiled DNA more efficiently than linear DNA [467]. The ability of MutLα to alter DNA topology would allow it to reposition crossing points within plectonemes to remove topological barriers between the mismatch and nicking point [464].

MutS has been shown to be sensitive towards the direction of supercoiling as the dissociation from MutS from a GT mismatch was slower under positive supercoiling than in relaxed DNA, and slower still under negative supercoiling [289]. Negative supercoiling showed faster MutS binding to the mismatch with positive supercoiling being similar to the relaxed circular DNA. Overall, both positive and negative supercoiling increase the stability of the MutS-mismatch interaction [289]. The difference in kinetics as a result of the direction of supercoiling suggests that MutS recognises changes in mechanical properties of the DNA in order to recognise the mismatch, and these same properties may explain why mismatches are repaired with varying efficiency depending on the sequence context.

The foundation of mismatches locating to the plectoneme tip had been previously well-established by showing that pre-kinking or bubble formation through defects can pin a plectoneme. Magnetic tweezer experiments have shown plectonemes to be positioned at a permanently kinked point in the DNA [104], statistical mechanics calculations have shown that a short, base-unpaired region—later defined by the field as a 'bubble'—will pin a plectoneme [470], and coarse-grained simulations modelling a mismatch by removing the hydrogen bonding within a base pair found that a bubble 2 nt in length will pin a plectoneme under high salt-high force conditions [288]. The same simulations revealed that under physiological conditions, a 2 nt bubble has only a stochastic probability of pinning the plectoneme, whilst increasing bubble size up to 16 nt deterministically increases the probability; however, the probability never reaches 1 [288]. Building upon this, magnetic tweezers showed DNA to be pinned by a single AA mismatch at high force and salt under positive supercoiling [287], ISD has been used to show a high plectoneme density in the region of a 10 nucleotide run of mismatches [111], and coarse-grained modelling has shown that thymine dimers can pin a plectoneme due to the induction of tip-bubbles by the lesion [290].

In linear, relaxed DNA, the distortion of DNA conformation as a result of the mismatch varies depending on which mismatch it is [471]. Purine-purine mismatches widen the base pair and can flip into the *syn* conformation, whilst pyrimidine-pyrimidine mismatches narrow the base pair. The ability of the bases to rotate around the glycosidic bond (fig. 4.3) affords a wide variety of

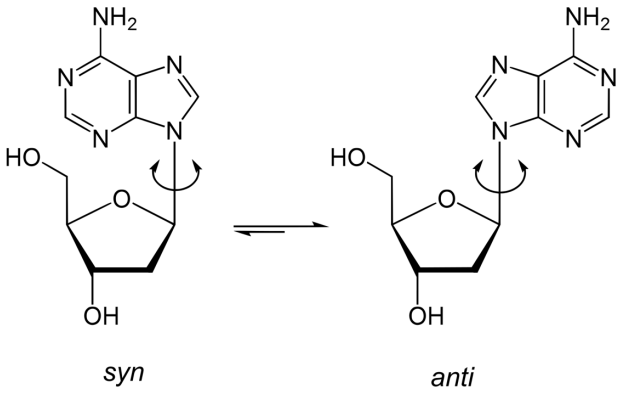

Figure 4.3: The *syn* to *anti* transition describes the rotation of the base around the glycosidic bond. Image adapted from WikiLectures [474].

hydrogen bonding schemes within in the literature. These Hoogsteen base pairs, defined as an *anti-syn* conformation, allow for previously inaccessible atoms to contribute to the hydrogen bonding within a base pair. For some mismatches, such as the GG mismatch, this is more stable than the traditional Watson-Crick conformation [472, 473]. Furthermore, wobble TT and GT mismatches increase the shear, and the reduction of hydrogen bonding between AA and CC can afford a variety of conformations including partially melted states. The energetic cost of the distortion of the duplex as a result of the mismatch (3.5-10 k_BT) is comparable to the distortion upon protein binding (3-8 k_BT), reinforcing the preorganization hypothesis. However, on the global scale, mismatches exert only moderate changes in DNA conformation [154]. The ability of DNA to buffer the global conformation against local distortions is unsurprising, as this allows DNA to maintain its integrity whilst containing lesions. The majority of mismatches induce a bend of only 19-25°, and even the most extreme cases of AC at 27.3° and CT at 30.9° do not reach the sharp 45-60° kink seen in mismatch-containing DNA upon MutS binding [154, 448, 449]. The same can be said for twist, where global levels lie within the variability of undamaged constructs despite deviations in local twist at the site of the mismatch [154]. However, local base step geometries at the mismatch are significantly affected due to the alteration in base pair interactions, primarily due to hydrogen bonding schemes. In general, pyrimidine-pyrimidine pairs have one pairing scheme, whilst purine-purine pairs have a range of bonding schemes allowing for large rotations and translations with respect to one another. The stretch, shear, opening, roll, and tilt are particularly affected. This results in breathing frequencies—defined as a large opening compared to the value of canonical B-DNA—of up to 67%, as seen for GG in molecular dynamics simulations, whilst some, such as CC have low breathing frequencies in line with the canonical Watson-Crick base pairs [154].

The resulting groove width and depth changes due to these local variations in DNA structure induced by the mismatch change the ion environment around the lesion. The effect of this also varies depending on the nature of the mismatch, with GT having a similar ionic distribution to canonical AT; whilst with GG, CA, and CC, the change in ionic environment is significant enough to affect the recognition of the mismatched DNA by a protein [154]. This is especially

important as mismatch recognition requires extensive interactions between MutS and the minor groove [285] and poorly repaired mismatches have been previously associated with narrower minor grooves [442].

We propose that the variable distortion in DNA due to mismatches may result in differing mismatch pinning propensities that further depends on the surrounding sequence. This, should DNA supercoiling play a role within the DNA damage recognition pathway, would offer an explanation as to why there are hypervariable repair efficiencies. Furthermore, the distortion of duplex conformation due to mismatch incorporation is likely to be highly sequence-dependent due to nearest neighbour effects and the effect of stacking. As GG is efficiently repaired and CC is poorly repaired [441–443], the increased breathing may facilitate bubble formation and promote plectoneme pinning. However, to date, the structure and properties of mismatches has only been considered within the context of two globally flexible and one globally rigid segments of DNA [154], or within the nearest neighbour's context [472] and it cannot be ruled out that there are effects beyond these. All structural understanding is further complicated when considering double and triple mismatches, where there is little structural information in current literature beyond stacking and hydrogen bonding parameters at the nearest neighbour level [472].

The mismatch repair pathway, and especially the details of the search and recognition process by MutS, are still unknown. This is primarily due to the lack of atomistic resolution studies [463]. Whilst some trends appear straightforward—such as increased breathing increasing mismatch repair efficiency [442]—as of yet, it is impossible to unambiguously determine what factors contribute to the variations in mismatch repair efficiency. It is likely that the overall efficiency of repair is a result of the combined variations in the rate and strength of association of MutS, the relative stabilities of the kinked and unbent states, and the rate of formation of the sliding clamp. We hypothesise that DNA supercoiling and the differing propensities of mismatches for the plectoneme tip play a role in the variation of MutS association rate and could also explain one factor in the overall hyper-variation of MMR repair due to both the nature of the lesion and the sequence context.

Here, I present all-atom implicit solvent simulations, using the methodology validated in the previous chapter, of the 12 single mismatch types in the Control339 and Mod339 sequences. This is combined with an investigation into what is driving the plectoneme pinning by simulating the same mismatches within two 55-mers of the same surrounding sequence contexts in explicit solvent. Finally, the competition between sequence and damage is further investigated by studying double, triple, and up to 8 tandem mismatches across four different sequences, including when directly inserted into a region of high pinning propensity.

4.2 Methods

A total of 1050 simulations of mismatch-containing minicircles are presented in this chapter, representing over 100 µs of simulation time. This is combined with 78 simulations of mismatched duplex DNA in explicit solvent totalling nearly 8 µs of simulation time. The simulations presented here of a 339 bp minicircle were run in implicit solvent on JADE2 at a rate of 12.5 ns/day, whilst the simulations of the 55-mer were run at a rate of 15 ns/day, also on JADE2.

Implicit Solvent Simulations

Implicit solvent simulations were carried out using the same protocol as previously described in section 3.2. In the Control339 and Mod339 sequences, all twelve single mismatches were inserted such that they were flanked by the same base pairs: A/T and C/G. This would negate nearest neighbour effects and focus the analysis on the competition with sequence on a larger length scale and allow for a more direct comparison between the single mismatches. Double and triple mismatches were also inserted into the Control339, Mod339, and Strong339 sequences, with the G/G single mismatch as well as 4, 6, 7, and 8 consecutive mismatches further inserted into the Strong339 sequence (table 4.4). The mismatches were placed at a position of low plectoneme density across both positive and negative supercoiling regimes (fig. 4.5). Finally, the G/G single mismatch, as well as the double and triple mismatches were inserted into a region of high plectoneme density in the Mod339 construct. This final construct is named inserted399. Unless otherwise stated, the ionic concentration was 0.2 M. The cAC and cGG mismatches were further simulated at 0.3 M, 0.4 M, 0.5 M, and 1 M salt concentrations.

Table 4.4: *The positions of the mismatches within the various sequences. All multiple mismatches were symmetrical.*

	Sequence			
Mismatch	Control339	Mod339	Strong339	inserted339
Single mismatches	107	25	85	82
Double (GG)	107-108	25-26	85-86	81-82
Double (TG)	106-107	24-25	84-85	82-83
Triple (CTG)	105-107	23-25	84-86	82-84
4nt (TGGC)			84-87	
6nt (ATGGCT)			83-88	
7nt (ATGGCTG)			83-89	
8nt (TATGGCTG)			82-89	

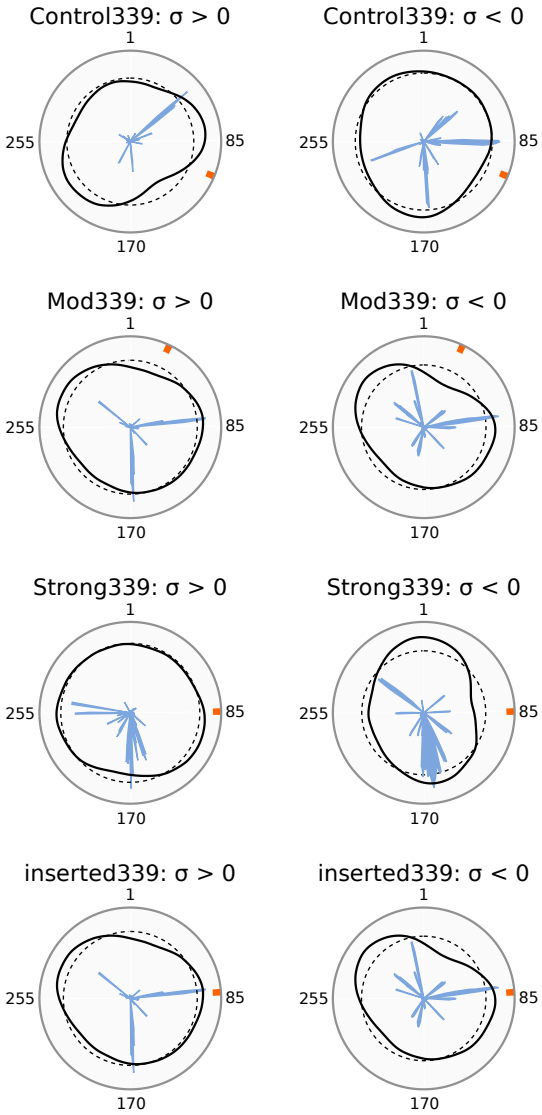

Figure 4.5: The positions of the mismatches within thhe corresponding control circle plots. Circle plots of the control simulations under both positive **(left)** and negative **(right)** supercoiling regimes. The location of the mismatch is shown by an orange band at the periphery of the circle plot.

Starting structures

Labile hydrogen bonding schemes within mismatches are the source of the large amount of variation seen in the structures of mismatches. These variations in conformation may lend themselves to a single mismatch having different propensities for the plectoneme tip. Thus, where multiple conformations are possible, the starting conformation was chosen to be the hydrogen bonding scheme with the most literature evidence of being recognised by MutS or, where this was not available, the lowest energy conformation found most often in the literature (table 4.6). The hydrogen bonds and corresponding distances for the restraints were taken from specific studies measuring and quantifying these or, where this was not available, found and measured from the crystal structure using the UCSF Chimera [475] tool FindHBond. All equivalent single mismatches had the same starting structure irrespective of sequence context

Table 4.6: *The starting structures for the mismatch simulations.*

Mismatch	Structure	Hydrogen Bonding Restraints	Notes
AA	2WTU [476]	2WTU [476]	*anti/syn* [477]
AC	1OH5 [448]	1OH5 [448]	
AG	3DNB [478]	[479, 480]	
CA	1OH5 [448]	1OH5 [448]	
CC	1FKZ [481]	[481, 482]	Wobble conformations interconvert on MD timescales [481]
CT	1FKZ [481]	[483]	
GA	3DNB [478, 481]	[479, 480]	
GG	10H7 [448]4	[484]	*anti/syn* [472, 477, 484]
GT	1D92 [485]	[486]	
TC	1FKZ [481]	[483]	
TG	1D92 [485]	[486]	
TT	[487]	[487, 488]	Wobble conformations interconvert when flanked by one pyrimidine and one purine [487].

and multiple mismatches were built in the same conformation as the corresponding single mismatch.

Implicit Solvent Simulation Analysis

Whether a mismatch pinned the plectoneme or not was determined by whether it was within 20 nt of either peak in the plectoneme propensity curve. Peaks were defined as the two positions with the highest plectoneme propensity values, subject to the constraint that the second peak was located at least 40 nucleotides from the first. In cases where more than two peaks were present, the two with the highest magnitudes of plectoneme propensity were selected for analysis. The WrLINE [418] program was used to calculate the twist of each base pair throughout the simulations. The average twist per base pair over all the simulations was plotted. The twist absorption was calculated by subtracting the twist at the mismatch from the baseline twist in this average plot. The baseline twist was determined by calculating the average twist over all base pairs whose twist was less than 2 standard deviations away from the mean. Due to the peak not always being centred directly on the mismatch, the lowest twist value for the nucleotide within 5 nt of the mismatch was considered. The width of the bubble for each mismatch was calculated by identifying the peak in the bubble density plot under negative supercoiling that

corresponds to the location of the mismatch and determining the first and last nucleotides within that peak. The peak is defined as being within two nucleotides with a bubble density of 0.

Explicit Solvent Simulations

Explicit solvent simulations were run of two 55-mers, Control55 and Mod55, containing the 12 mismatches in order to investigate the effect of mismatches within our specific sequences in the absence of torsional strain. See A.1 for sequences. These sequences contain the mismatch at nucleotide 31 and the surrounding sequence was that of the corresponding Control339 and Mod339 constructs. There was 64% total sequence similarity across the two 55-mers.

The linear DNA starting structures for the simulations were built using the NAB module in AmberTools20 [412]. The DNA was solvated in a truncated octahedral box with the TIP3P water model [417] with a 15 nm buffer. The DNA was neutralised with Ca^{2+} counterions and additional Ca^{2+}/Cl^{-} ions were added at a concentration of 200 mM. A multi-stage equilibration protocol was used as follows. Energy minimization was performed for 10,000 cycles with the first 5,000 being of the steepest descent method and the rest conjugate gradient minimization. This was done first with the DNA restrained with a weight of 50 kcal mol^{-1} Å^{-2} in order to minimize the solvent and ions. This was repeated with no restraints in order to minimise the DNA structure. The long-range electrostatic cut-off was 12 Å and the timestep was 2 fs. The system was first brought to 100 K for 10 ps followed by a further 10 ps to bring the temperature to 300 K with the Langevin thermostat and a collision frequency of 1 ps^{-1}. The pressure was brought to 1 bar using the Monte Carlo barostat with the DNA restrained with a weight of 50 kcal mol^{-1} Å^{-2} for 10 ps. Five short 10 ps equilibration runs were then performed with decreasing DNA restraint weights of 25, 10, 5, 2.5, and 1 kcal mol^{-1} Å^{-2}. A 100 ns production run was then carried out using the pmemd.cuda module of Amber20 [416] with no restraints apart from constraining all bonds to hydrogen using the SHAKE algorithm. The input file for the simulation production run can be found in A.3.

Explicit Solvent Simulation Analysis

Analysis was predominantly performed using AmberTools20 modules [412]. The bend angle induced at the site of the mismatch was calculated by measuring the angle between the mismatch and the nucleotides 10 nt away in either direction. SerraLINE was used to measure curvature using the molecular contour traced by WrLINE [418]. The curvature was calculated using tangent vectors separated by 16 nucleotides, equal to approximately 1.5 helical turns, to neglect the influence of the helix. To avoid end-effects, only nucleotides 15-32 were considered which span a total curvature over 15-48 nt.

The *nastruct* module was used to calculate base pair and base step geometries. This module is an Amber implementation of the 3DNA procedure [9, 489] using reference frame coordinates by

Olson *et al.* [12] and algorithms adapted from Babcock *et al.* [490]. Breathing was defined as in previous studies [154] by a deviation of ±20° from the value of canonical B-DNA (-3.6°) and expressed as a total percentage of frames where breathing occurs at the mismatch site.

MDAnalysis was used to measure the persistence length [491, 492]. It was calculated by measuring the autocorrelation, $C(n)$ of two bond vectors, (a_i, a_{a+n}) separated by n bonds (eq 4.1).

$$C(n) = \langle cos\theta_{i,i+n} \rangle = \langle a_i \cdot a_{i+n} \rangle \tag{4.1}$$

The persistence length, l_P, and average bond length, l_B^-, were calculated by fitting the exponential decay (eq 4.2).

$$C(n) \approx exp(-\frac{nl_B^-}{l_P}) \tag{4.2}$$

The Morse potential and elastic constants for the mismatched sequences were taken from the literature using the Search Mismatch Parameters app [472].

4.3 Results

The 12 possible mismatch types (table 4.6), accounting for strand direction, were inserted into the Control339 and Mod339 sequences in regions of low plectoneme densities. In both sequences, they were placed with the same flanking base pairs, but the sequences varied beyond the nearest neighbours (Control339: TTT**X**CGA; Mod339: ACT**X**CAC). Herein, they are referred to by the mismatched base pair preceded by c or m for insertion into Control339 or Mod339, respectively.

The AA and GG mismatches were simulated in the *anti/syn* conformation, whilst the rest were all-*anti*. This should not negatively impact our results as the mismatches are able to flip between conformations on the timescale of our simulations in explicit solvent (fig. 4.7). Two of the three cAA replicas underwent this spontaneous transition; however, none of the GG nor the mAA replicas did. This suggests a rare, but possible pathway. Furthermore, the GG mismatch is known to be highly stable in the *anti/syn* configuration.

4.4 Single mismatches have sequence-dependent pinning propensities

Following on from the convention established in the previous chapter, the pinning propensity was plotted as a function of the base pair position. Thus, a mismatch was recorded as pinning the plectoneme if it was within 20 nucleotides of the two greatest peaks in the respective plot

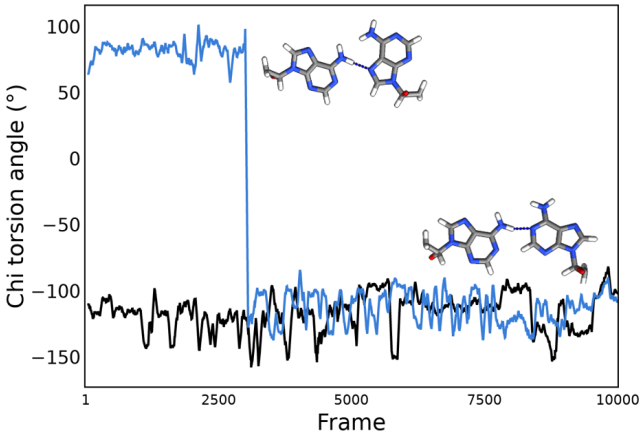

Figure 4.7: Mismatches can flip between *syn* and *anti* on the timescale of our explicit solvent simulations.
A plot of the chi torsion angle (°), which defines the *anti* or *syn* conformation of the base. Example shown for cAA where the *syn* base (blue) spontaneously flips to *anti*, whilst the *anti* base (black) remains in that conformation. The *syn* conformation has a torsion angle between -90° and 90°. The *syn* to *anti* transition can be seen clearly at approximately the 3,000$^{\text{th}}$ frame by a sharp high-to-low torsion angle transition. The conformation of the base at the corresponding chi angle is shown on the plot.

(table 4.8). All of the single mismatch-containing circle plots can be found in the appendix (fig. A.13 - A.20); however, specific plots will be included and discussed in the main text where appropriate. Throughout the text, mismatches are sometimes referred to as "damage", especially

Table 4.8: *The propensity for a single mismatch to pin a plectoneme is both lesion- and sequence-dependent.*
The mismatches that do (1) and don't (0) pin the plectoneme in the corresponding sequence under the given direction of supercoiling. Mismatches that do pin are further highlighted with a grey cell.

	Control339		Mod339	
Mismatch	Positive	Negative	Positive	Negative
AA	0	1	1	0
AC	0	1	0	0
AG	1	1	0	0
CA	1	0	0	0
CC	0	1	0	0
CT	1	0	0	0
GA	1	0	0	0
GG	0	0	1	1
GT	0	0	1	0
TC	1	0	1	0
TG	0	0	0	0
TT	0	0	0	0

in the context of the competition between sequence and damage for the plectoneme tip. Whilst we hypothesise that our results may be applicable to other forms of DNA damage, this is as of yet unconfirmed. Thus, I must stress, that any references to DNA damage within these results are, currently, applicable only to DNA mismatches.

Only the TC mismatch consistently pins the plectoneme in both sequences in the same direction of supercoiling. This suggests that a mismatch does not have an inherent pinning propensity that drives its localization to the plectoneme tip when it exceeds that of the surrounding sequence. Instead, the ability of a mismatch to pin must be greatly influenced by the sequence context and, potentially, existing in balance with numerous factors that drive overall plectoneme positioning. The consideration resulting from this is that in studies involving only a singular mismatch type or sequence, as discussed in the introduction, the results cannot necessarily be extrapolated to mismatches of a different nature or within different sequence contexts.

Due to the competition between the mismatch and the sequence for pinning the plectoneme tip, the resulting pinning landscape can become more complex than those seen in the undamaged constructs. In some cases, three or more peaks can appear in the pinning propensity plots, with their relative heights reflecting this competition. For ease, we have excluded cases where the mismatch is situated in a region of high pinning propensity but is not at the tip, as well as cases where the mismatch increases local pinning but not sufficiently to overcome the inherent sequence-driven pinning (fig. 4.9). Therefore, we define 'pinning' strictly as those cases in which the mismatch consistently localises to within 20 nucleotides of the plectoneme tip, reflected in the overall plectoneme pinning landscape.

The mismatches that increase local pinning propensity but do not pin the plectoneme overall can broadly be defined as one of two cases (fig. 4.10). Most obviously, there are occasions in which the mismatch pins in some, but not the majority, of simulation replicas. This is reflected

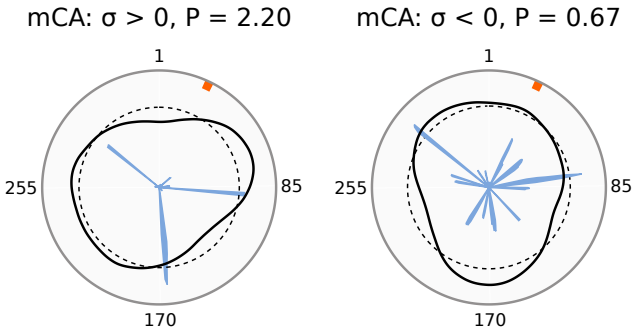

Figure 4.9: We define mismatch 'pinning' strictly as cases where the mismatch consistently localises to the plectoneme tip, reflected in being within 20 nucleotides of the two largest peaks in the overall pinning propensity plots.
Here we show two examples that do not qualify as mismatch-driven plectoneme pinning. In both directions of supercoiling, the location of mCA is not within 20 nucleotides of the peak despite local increases in pinning propensity. Thus, the pinning is likely driven by other factors, or the strength of the mismatch-driven pinning is not sufficient to overcome that by sequence.

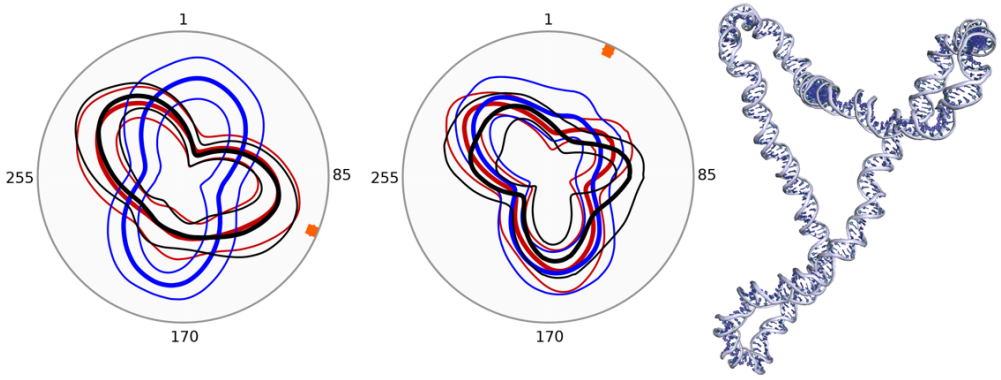

Figure 4.10: Mismatches that increase local pinning propensity may do so by pinning the plectoneme in some, but not all of the simulation replicas, or by aiding the mismatch to adopt more or broader peaks at higher values of writhe.
Polar plots showing the average distance from each base pair to the centre of mass of the DNA over the course of a simulation. Each replica is represented by a different colour with the mean per base pair, and the mean ± the standard deviation plotted. **Left:** In some cases, the mismatches pin the plectoneme in some simulations but show no local pinning propensity in other repeats. Example shown is cAG at $\Delta Lk = -3$. **Right:** At high writhe, the minicircles can adopt conformations that result in broader or more than two plectoneme tips. The example shown is cCT at $\Delta Lk = -6$. An example conformation is shown to the right, spatially aligned with the corresponding circle plot.

in small, defined plectoneme propensity peaks at the mismatch. Secondly, the mismatches can adopt conformations with broader plectoneme peaks or with three peaks. This is aided by positioning the mismatch in a region of high curvature without compromising the preferential pinning of the sequence. This is seen more often in the higher superhelical densities as the writhe increases, and it is reflected by more complicated pinning landscapes with less defined peaks. The latter case is likely an artefact of using minicircles as a model and the proximity of the mismatch to the pinning region; however, it does reflect the weaker pinning propensity of the mismatch compared to the sequence in that case.

Positive Supercoiling

Under the positive supercoiling regime, the undamaged Mod339 has a pinning propensity of approximately half that of Control339. This offers an explanation as to why the P values depicting mismatch-driven plectoneme pinning are smaller in the control sequence: they have greater competition for the plectoneme tip (fig. 4.11). Only cCT has two clear peaks in the pinning propensity landscapes, similar to those more often seen in the undamaged constructs, which leads to small P values and describes weak overall pinning. The other mismatches have approximately three peaks, defined as regions at or above the dashed average plectoneme propensity for the corresponding sequence plot. For reference, the Flat75-2 construct increased pinning propensity to 1.12 compared to Control339. Here, under positive supercoiling, only cCT and cCT supersede this value. Conversely, two well-defined peaks are seen across all of the pinning propensity landscapes in Mod339. mTC is the strongest pinning mismatch in this sequence and, interestingly, is the only mismatch to pin across both sequences under positive supercoiling.

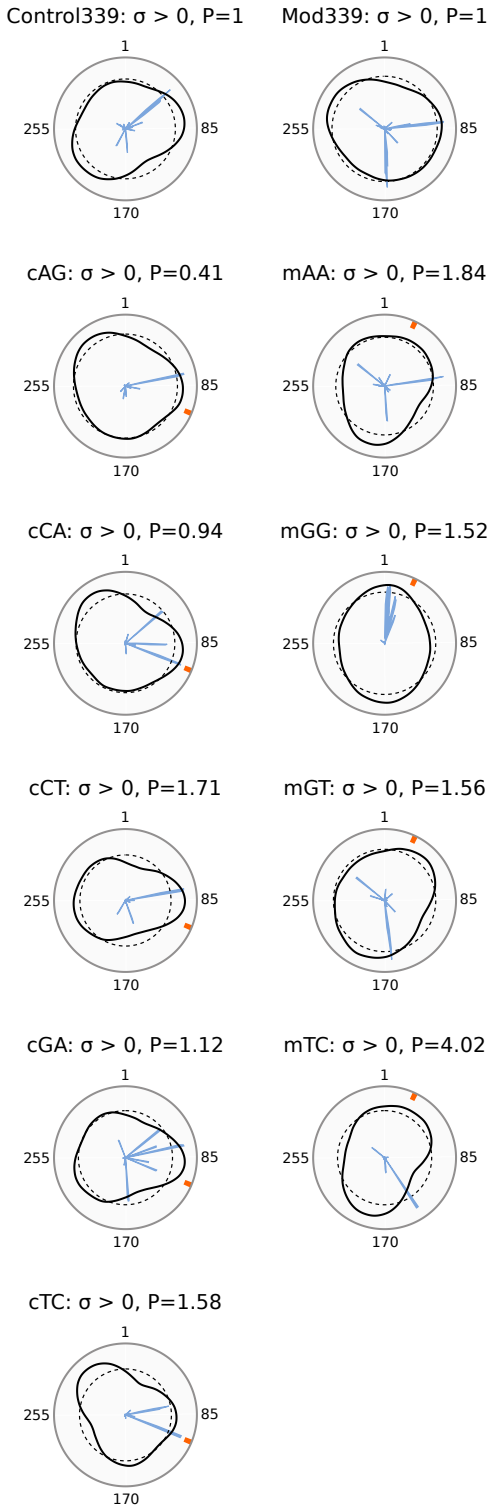

Figure 4.11: Under positive supercoiling, the mismatches pin more strongly in the Mod339 sequence—reflected by the larger P values—despite more mismatches being able to pin in Control339.

Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mismatches in the Control339 **(left)** and Mod339 **(right)** sequences under positive supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches are represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct under positive supercoiling is equal to 1.

Despite Control339 having a higher P value under positive supercoiling, five mismatches are able to pin the plectoneme here compared to only four in Mod339. This could be explained by the double-pinning effect. Due to the position of the mismatch within the Control339 sequence, when it is located at the plectoneme tip, the inherent pinning of the opposite side is in the same approximate region as that preferred in the undamaged control. Pinning by the mismatch required only a small shift in the plectoneme propensity landscape. This was necessary to ensure the same nearest neighbours as the mismatch in Mod339, and to be in a position of low pinning propensity across both supercoiling regimes. Conversely, in Mod339, pinning of the mismatch requires the opposite apex to be in a region of low plectoneme propensity in the undamaged control. This may make mismatch-driven pinning more unfavourable despite the P value of the undamaged Mod339 being lower than Control339 in the positive supercoiling regime. This might also account for the more deterministic pinning observed in Mod339, as only mismatches with the highest pinning propensity may be capable of overcoming the inherent unfavorability of this effect.

Predicting Positive Pinning

To investigate the drivers of mismatch-directed plectoneme pinning, we simulated the mismatch and the immediate surrounding sequence as a 55-mer in explicit solvent. These are referred to as Control55 and Mod55. The mismatch was introduced by mutating the 31st base pair—positioned off-centre—to disentangle the effects of global bending around the DNA midpoint from those specifically induced by the mismatch. Under positive supercoiling, intrinsic curvature is the main influence of sequence-driven plectoneme pinning [41]. Here, we investigate trends in the structural properties of the mismatch-containing 55-mers to elucidate the factors contributing to mismatch-induced plectoneme pinning.

Intrinsic Curvature

In all MutS-DNA crystal structures, the DNA is sharply bent or kinked with the hinge point at the mismatch [154, 477]. Therefore, we first investigated intrinsic curvature as the angle between the mismatch and the two DNA ends (fig. 4.12). In general, there was an increased bend centred at the mismatch; however, this was not true in all cases and, overall, angles were in line with the variation seen in the undamaged constructs. This was not unexpected as the surrounding DNA is known to be able to buffer against significant global distortions despite local deformations [154]. Furthermore, the bends seen in bare DNA have a global energy minimum at around 160° and do not reach the degree of sharp kinking seen in protein-bound DNA [493]. However, this could offer an explanation as to why the propensity to kink versus the stability of the unbent state does not completely explain variations in mismatch repair efficiencies.

Secondly, we sought to investigate bending in the context of smooth curvature, analogous to the intrinsic curvature measured in the previous chapter, rather than sharp kinking hinged at the

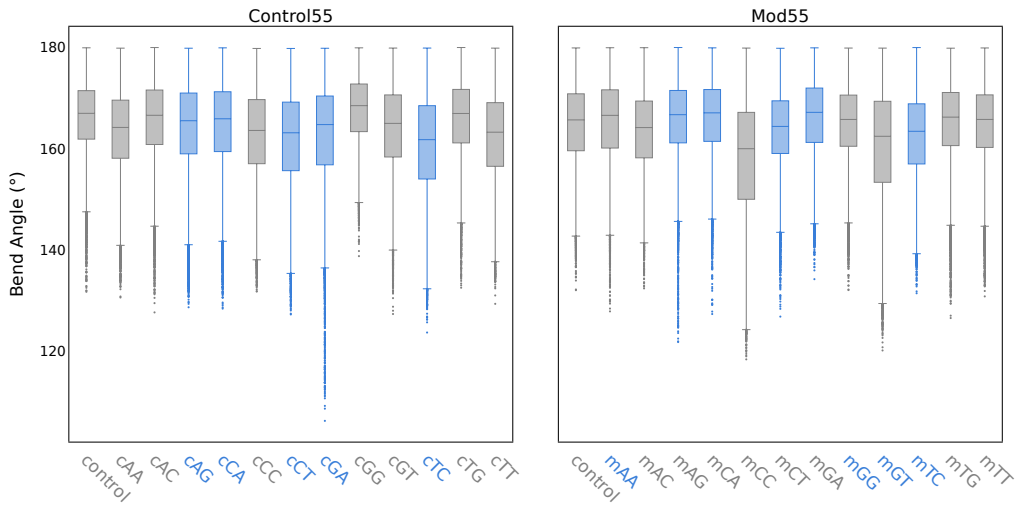

Figure 4.12: The bend angle centred at the mismatch does not correlate to plectoneme pinning under positive supercoiling.
Box plots of the bending angles centred at the mismatch between the nucleotides 10 nt away in either direction, across the three repeats of each simulation. The mismatches that pin under positive supercoiling are depicted with blue box plots.

mismatch. When measured over tangents separated by 16 nucleotides, the intrinsic curvature was both mismatch- and sequence-dependent. Surprisingly, not all mismatches increased the intrinsic curvature of the duplex relative to the controls. This was more obvious in Mod55, potentially owing to a higher intrinsic curvature in the undamaged region relative to Control55. Those mismatches that pinned the plectoneme under positive supercoiling in Control339 had very clearly the largest intrinsic curvatures in the corresponding 55-mer (fig. 4.13). However,

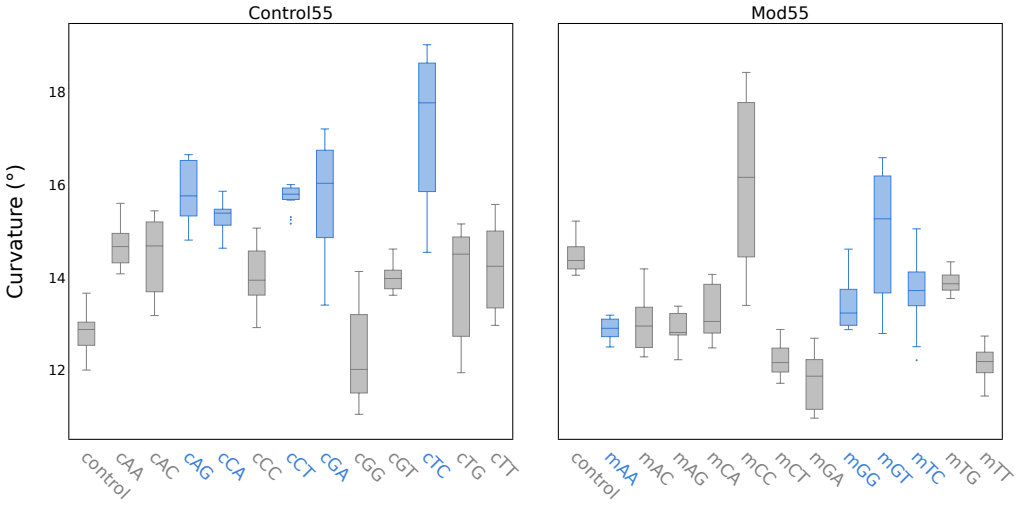

Figure 4.13: The intrinsic curvature of the 55-mer sequences correlates to pinning, but not the strength of pinning, under positive supercoiling in the Control339 sequence. The trend does not hold for Mod339.
The curvature of the 55-mer sequences measured over tangents separated by 16 nucleotides. As the mismatch is not always centred at the mid-point of the curvature, the curvature at nucleotides 15-32 over the three simulation replicas was plotted as a box plot. Due to the separation of the tangents, the plotted curvature spans the 15th and 48th nucleotides with the mismatch at position 31. The mismatches that pin under positive supercoiling are depicted with blue box plots.

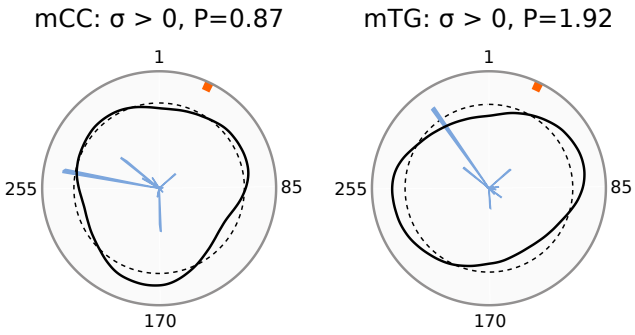

Figure 4.14: Two of the mismatches with the highest intrinsic curvatures, mCC and mTG, fail to offer any competition towards the sequence for the plectoneme tip.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mCC **(left)** and mTG **(right)** mismatches under positive supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The P values are normalised such that the undamaged Mod339 under positive supercoiling is equal to 1. The location of the mismatch is represented by an orange band.

the magnitude of the intrinsic curvature does not correlate to the strength of pinning in the minicircles.

Surprisingly, the trend is not as clear in Mod55 despite our hypothesis that it must overcome unfavourable pinning at the opposite apex. If intrinsic curvature were the sole driver of pinning propensity, it would be expected that mTG and, especially, mCC would pin strongly. We have previously seen how the mismatches must compete with the sequence for the plectoneme tip; however, there is no detectable peak—not even a minor peak—at the mismatch site in these sequences (fig. 4.14). This highlights an unexpected complete absence of pinning preference for these mismatches. Furthermore, all of the mismatches apart from mGT that pin the plectoneme under positive supercoiling have a lower intrinsic curvature than in the undamaged 55-mer.

Local Flexibility

Whilst intrinsic curvature is the main driver of plectoneme pinning under positive supercoiling according to the Dekker model, the local flexibility also plays a role [41]. We, therefore, sought to determine whether the flexibility of the mismatch could account for the discrepancies observed between intrinsic curvature and the plectoneme pinning strength of the mismatches. Although curvature and flexibility are related, we could not rule out the possibility that a highly flexible mismatch might enhance pinning even in the case of only moderate curvature. Conversely, a mismatch with low flexibility may hinder plectoneme pinning despite exhibiting high intrinsic curvature.

In the Dekker model, flexibility is quantified by the variances of the tilt and roll base step parameters [41]. These variances are inversely proportional to the local persistence length [494], allowing for the estimation of local flexibility along the sequence by summing the variances over a given stretch of DNA [41]. We have measured the flexibility at the mismatch by summing the

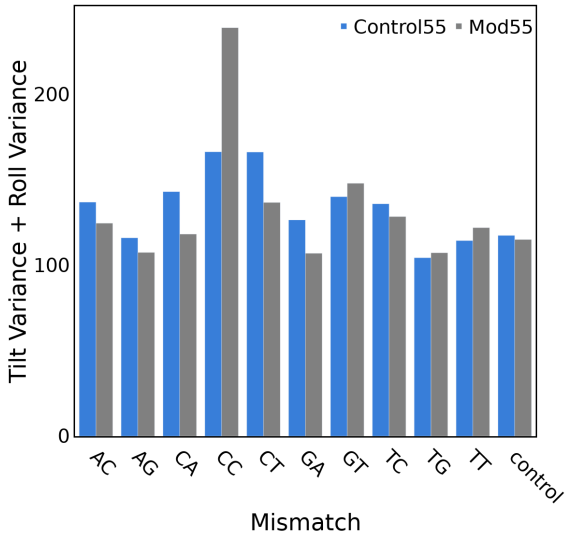

Figure 4.15: The flexibility of the mismatches cannot explain the discrepancies between the predicted versus simulated pinning strength of the mismatches.
The summed variances of the tilt and roll over the three replicas of each simulation at the two base steps involving the mismatch. The variances are a quantitative measure of flexibility and are inversely proportional to the local persistence length.

tilt and roll variances over the three replica simulations across the two base steps involving the mismatch.

In general, the flexibility of the mismatch is greater than in the control and it is similar for a given mismatch across both sequences (fig. 4.15). Furthermore, it is largely unaffected by strand direction as equivalent, non-symmetrical mismatches, such as AC and CA, had similar flexibilities. In agreement with the literature [454, 455, 481], the greatest flexibility was found in the CC mismatch. Thus, along with the high intrinsic curvature of mCC, the failure to show any plectoneme pinning propensity remains unexplained. The cCC and cCT mismatches both exhibit similarly high flexibility. However, only cCT—which also exhibits high intrinsic curvature—successfully pins the plectoneme. This contrasts with cCC, where high flexibility fails to compensate for its lower curvature. Conversely, the more highly curved cAG is able to pin the plectoneme despite its relatively low flexibility. Therefore, we conclude that the flexibility of the mismatch does not significantly impact on its ability to pin a plectoneme. To verify that flexibility was not significantly affected beyond the nearest neighbours, the base step parameters were calculated for the next nearest neighbours (see fig. A.21 for example in 5' direction). The flexibilities of these base steps are comparable across simulations of both the control and the mismatched DNA.

Persistence Length

The *nastruct* command was unable to calculate base step parameters for the *anti/syn* conformations. Further, the change in flexibility may propagate further along the duplex than just at the mismatch site. Thus, we also calculated global persistence length by measuring

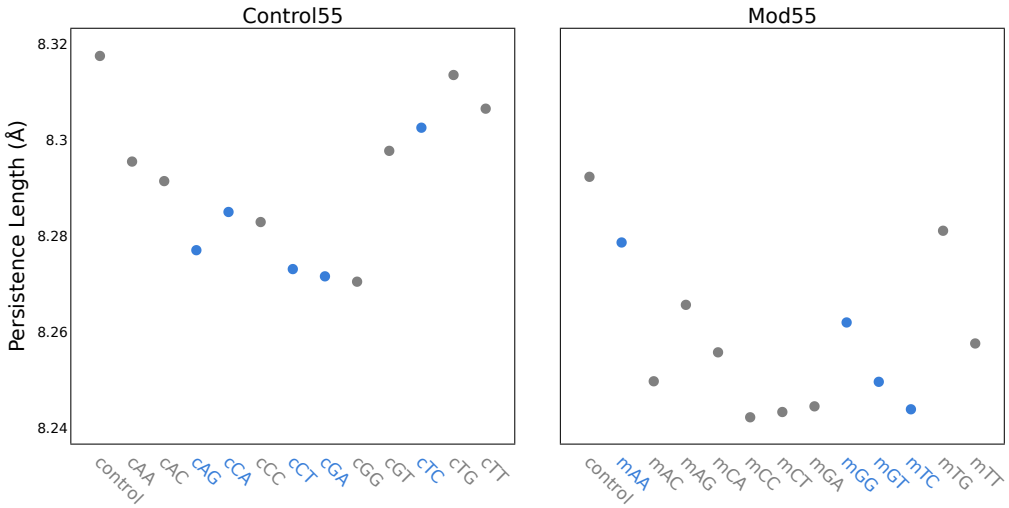

Figure 4.16: The relative persistence lengths of the mismatch-containing 55-mers do not, in general, account for flexibility-driven discrepancies between the expected and actual pinning strengths of the mismatches.
The persistence length (Å) of the mismatch-containing Control55 and Mod55 constructs was calculated using correlation-decay curves of the bond vectors. The persistence length was calculated across all three replicas for each simulation and averaged. Mismatches that pin the plectoneme under positive supercoiling are shown in blue.

the autocorrelation of the bond vectors across the entire 55-mers (fig. 4.16). In all cases, the calculated value on the order of 8 Å is a significant deviation from the commonly accepted value of approximately 50 nm [120]. However, using correlation-decay methods are less applicable on length scales shorter than the persistence length. Therefore, these calculations serve only to compare relative trends and not absolute values.

Across the 55-mer, Mod55 had a shorter persistence length than Control55 and, thus, should have aided plectoneme positioning. However, this effect is likely negligible in the context of the whole minicircle. Moreover, since a greater number of mismatches pin in Control339, it is evident that this local flexibility is not able to enhance pinning by a mismatch to a discernible degree. As expected, across both sequences, the mismatches decreased the global persistence length of the 55-mer; however, the relative magnitude of the effect was sequence-dependent. For example, cTC has one of the largest persistence lengths across the Control55 sequences, but mTC has one of the lowest across Mod55 despite pinning the plectoneme in both.

Whilst local flexibility at the mismatch site does not necessarily correlate to a shorter persistence length—exemplified by cGT having a long persistence length and a relatively flexible mismatch—this second measure of flexibility also fails to account for discrepancies between the expected and actual pinning strengths of the mismatches. In addition to inducing the greatest curvature into the 55-mer, mCC has amongst the shortest persistence lengths of all the mismatched constructs and still fails to pin the plectoneme. Furthermore, mAA, which does pin the plectoneme, has an intrinsic curvature less than that of mTG and comparable to three other mismatches, none of which have a strong pinning propensity; however, of the mismatched Mod55 constructs, mAA has one of the longest persistence lengths. This is especially surprising

as the bubble density in mAA is relatively low compared to the pinning landscapes of the other mismatches (fig. 4.11) and so curvature and flexibility, and not bubble formation, is likely driving the pinning.

However, in the Control55 constructs, where the degree of intrinsic curvature predicts plectoneme pinning but not pinning strength, the persistence length may offer an explanation. For example, the intrinsic curvature of cTC is much higher than its reciprocal, cCT, but does not pin the plectoneme as deterministically. It has a lower P value and a less pronounced peak at the mismatch. This could be explained by cTC having one of the longest persistence lengths across all of the mismatch-containing sequences. However, this may merely be coincidental as both the average intrinsic curvature and the persistence length of cCT are approximately equal to that of cGA despite cCT having a larger P value.

Negative Supercoiling

With Mod339 having a P value nearly six times greater than that of Control339 under negative supercoiling, it is unsurprising that only one mismatch (mGG) pins the plectoneme in Mod339 compared to five in Control339. Consistent with the double-pinning effect seen in Mod339 under positive supercoiling, the mismatches pin Control339 more deterministically in the negative supercoiling regime compared to the positive Control339. This is due to the positioning of the mismatch under negative supercoiling having a more favourable inherent pinning of the opposite apex which lessens the competition for the tip.

As expected, the strength of pinning generally correlated to bubble localisation, at least amongst the mismatches that did pin the plectoneme, as determined by visual inspection of the polar plots (fig. 4.17). The weakest pinning mismatch, cAC, which had the lowest P value, showed bubble formation broadly spread across the plectoneme propensity peak as well as elsewhere in the minicircle. In contrast, the strongest pinning mismatch, cCC, localises bubble formation around the mismatch-containing peak. Notably, mGG—when accounting for the relatively higher P value of Mod339 compared to Control339 under negative supercoiling—has the highest P value seen in any of the single mismatch simulations. It has an especially strong bubble localisation at the mismatch-containing peak which strongly positions the opposite side of the plectoneme. Interestingly, this strong localisation was also seen for the same mismatch under positive supercoiling. However, in the positively supercoiled regime, mGG had a smaller P value than mTC, which had comparatively little bubble formation. Without further mismatches that pin the plectoneme in Mod339 against which to compare mGG in the negatively supercoiled regime, it is unclear if the strong bubble localisation is required for plectoneme positioning under negative supercoiling in this sequence, or if the bubble formation is an artefact of certain properties of the GG mismatch that are independent of the driving forces of plectoneme localisation.

Once again, we see a direction-dependent pinning propensity with cAC and, especially, cAG

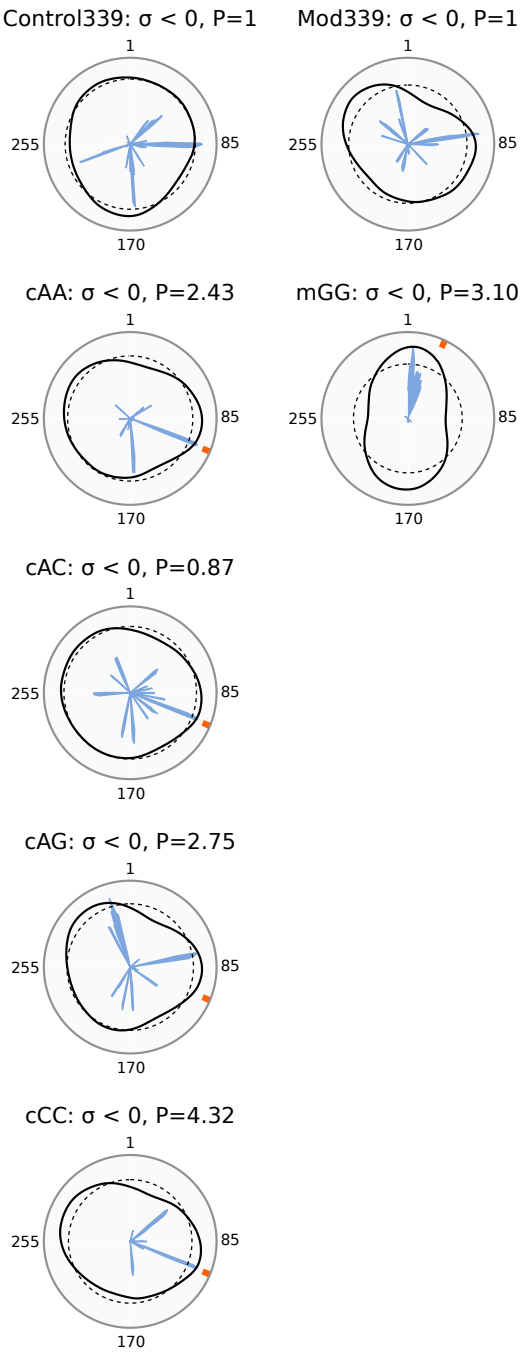

Figure 4.17: Under negative supercoiling, only mGG pins the plectoneme in Mod339 compared to four mismatches in Control339.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mismatches in the Control339 **(left)** and Mod339 **(right)** sequences under negative supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches are represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct under positive supercoiling is equal to 1.

pinning the plectoneme relatively strongly. However, whilst cGA increases local pinning propensity, it is not the dominant region localising to the plectoneme tip. Conversely, despite a large amount of bubble formation at the mismatch, cCA does not show any local pinning propensity (fig. 4.18).

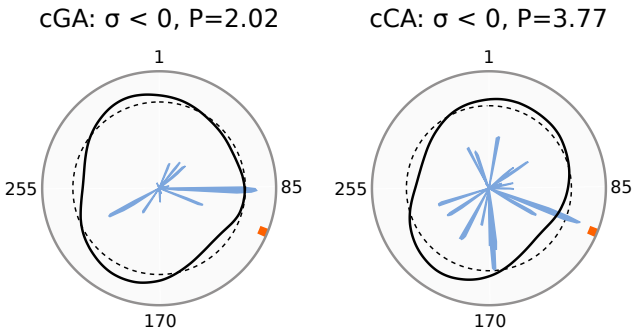

Figure 4.18: Whilst cAG and cAC pin under negative supercoiling, cGA does not increase local pinning propensity sufficiently to overcome the sequence-driven pinning, and cCA shows significant bubble formation at the mismatch but no local pinning propensity. Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cGA **(left)** and cCA **(right)** mismatches under positive supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The P values are normalised such that the undamaged Mod339 under positive supercoiling is equal to 1. The location of the mismatch is represented by an orange band. Neither mismatch pin the plectoneme despite relatively strong plectoneme pinning by their reciprocals (cAG and cAC).

Predicting Negative Pinning

It is more difficult to predict plectoneme positioning under negative supercoiling where, for the undamaged constructs, statistical mechanics calculations are required to find the correct balance of the various possible denatured states and accurately predict bubble positioning. The local denaturation is described by a combination of the energies of denaturation and torsional deformation, as well as the interactions between the denaturation and torsional deformations. Without a full SIDD model capable of modelling mismatches, we are limited to investigating only the main parameter in the algorithm: the energies of denaturation [109].

Denaturation Energies

The melting temperature of all single and double mismatches has recently been measured across all nearest neighbour sequence contexts and, from this, the contributions from the hydrogen bonding and stacking interactions calculated [472]. In this model, the Morse potential, D, describes the base-pair-dependent interactions, predominantly reflecting hydrogen bonding, while the elastic potential, k, mimics the stacking interactions. We first investigate whether trends in these parameters can predict pinning under negative supercoiling (table 4.19). The choice of these parameters is further supported by prior work: the thermodynamic destabilization has previously been attributed to a contribution of nearest neighbour and hydrogen bonding effects [495], whilst, in the SIDD model, the initiation energy required to first nucleate the bubble is largely dependent upon the stacking interactions [109].

Firstly, the melting temperature for the trimer centred on the mismatch showed no correlation with the pinning propensity of the mismatch as both the mismatch with the lowest (CC: 55.3 °C) and highest (GG: 61.2 °C; AG: 61.7 °C) melting temperatures pinned the plectoneme. In

Table 4.19: ***The melting temperature and constituent contributions from the hydrogen bonding and stacking interactions show no trend with the pinning propensity of the mismatch under negative supercoiling.***
The measured melting temperature, T_i, Morse potential, D, and elastic constants, k, of the single mismatches in the TXC/AXG sequence context. The mismatches are in the same trimer sequence context in both the Control339 and Mod339 sequences. Values calculated from data from Oliveira et al. [472]. The rows where the mismatch pins under negative supercoiling in either sequence are shaded grey.

			k / eV nm^{-2}		
Mismatch	T_i / °C	D / meV	k AT-MM	k MM-CG	Sum of k values
AA	58.3	7.73	2.33	2.08	4.41
AC	58.1	2.3	1.11	1.86	2.97
AG	61.7	18.78	1.77	5.45	7.22
CA	57.4	2.3	3.29	1.94	5.23
CC	55.3	2.29	1.13	0.6	1.73
CT	56.3	1.15	1.99	1.32	3.31
GA	59.8	18.78	6.09	1.5	7.59
GG	61.2	27.42	2.28	2.97	5.25
GT	60.4	17.46	4.42	2.33	6.75
TC	56.9	1.15	1.45	0.76	2.21
TG	60	17.46	5.64	2.42	8.06
TT	57	6.47	2.64	1.17	3.81

direct disagreement with the expected trend, GG had the second-highest melting temperature and the strongest pinning propensity in mGG across all sequences, driven by significant bubble formation. The GG mismatch further showed the strongest hydrogen bonding of any mismatch with a Morse potential nearly 10 meV higher than the second strongest hydrogen bonded: AG. It is comparable in magnitude to that of canonical A-T and yet induces significant denaturation at the mismatch site under both supercoiling regimes. In line with expectations, three of the five mismatches that pin the plectoneme have a Morse potential of below 10 meV; however, there are numerous other mismatches with comparable or smaller Morse potentials that do not pin in either sequence.

With the stacking interaction being a significant contributing factor to the initial nucleation of the bubble, we then considered the possibility that a low elastic constant was more predictive than the hydrogen bonding interactions. However, in addition to a large Morse potential, AG has one of the strongest stacking interactions of all the mismatches, whilst AA and AC have two of the weakest. Furthermore, the reciprocal mismatches TC and CT have the lowest combination of Morse potential and elastic constant and do not pin the plectoneme in either sequence under

negative supercoiling. Both of these pin under positive supercoiling and, therefore, will have an adequate intrinsic curvature to complement this destabilisation. Thus, we conclude that the melting temperature, Morse potential, and elastic constant are not predictive of pinning the plectoneme under negative supercoiling.

Secondly, we sought to take advantage of the increased granularity of the molecular dynamics data and calculate the actual percentage of simulation time each mismatched base pair has a given number of hydrogen bonds. Furthermore, this would allow us to elucidate any trends in the hydrogen bonding patterns beyond the nearest neighbour. Thus, we considered all, not just the Watson-Crick, hydrogen bonds throughout the simulations of the 55-mers. In the case of C/G, this identified up to five hydrogen bonds, compared to the normal three Watson-Crick hydrogen bonds (fig. 4.20).

In general, the observed hydrogen bonding patterns were consistent with the previously described Morse potentials as those with the lowest values were largely dominated by a single hydrogen bond conformation (table 4.21). Furthermore, the hydrogen bonding behaviour is generally similar for the same mismatch across both sequences. The only notable exception is the AA mismatch where cAA exhibited approximately 20% more simulation time with no hydrogen bonds between the mismatch than mAA. Similarly, mGG has a 10% lower relative population of the three-hydrogen-bond state compared to cGG. Whilst this may explain why mGG has a strong pinning propensity and cGG does not, mGG remains with two hydrogen bonds for nearly 60% of the simulation time. Furthermore, it remains significantly more stable than many other non-pinning mismatches. Surprisingly, mCC is one of only three mismatches to spend approximately one-quarter of the simulation time with no hydrogen bonding, yet is the only one of the three to not pin the plectoneme. Whilst this could be attributed to the stronger competition in the Mod339 construct for the plectoneme tip, it is significantly less stable than mGG—that did pin the plectoneme—and had the lowest elastic potential in the melting temperature study (table 4.19). Finally, several mismatches—namely CA, CT, and their

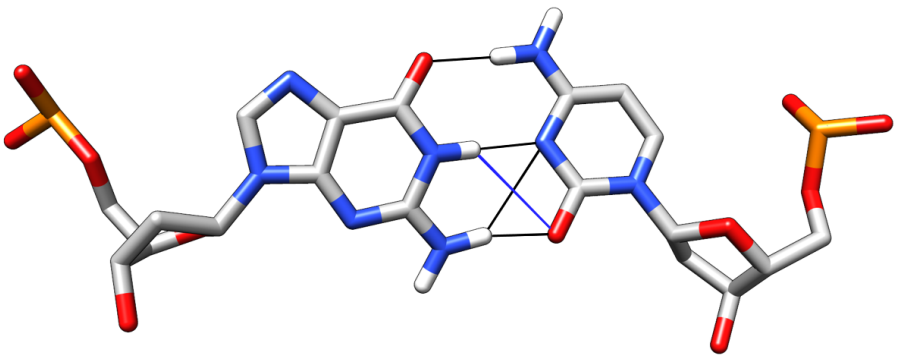

Figure 4.20: Due to calculating all hydrogen bonding, and not just Watson-Crick hydrogen bonding, the C/G base pair can have up to five hydrogen bonds.
The C/G base pair has four hydrogen bonds (black) for approximately 46% of simulation time with a transient fifth hydrogen bond (blue) for 0.05% of simulation time. The dominant hydrogen-bonding scheme, however, is that of the Watson-Crick three hydrogen bond conformation.

Table 4.21: ***The relative populations of the number of hydrogen bonds between the mismatches are not indicative of pinning propensity under negative supercoiling.***
A table showing the percentage of simulation time across all three repeats where a given mismatch has the given number of hydrogen bonds. All hydrogen bonds, and not just Watson-Crick hydrogen bonds, were calculated. The mismatches that pin the plectoneme in that sequence under negative supercoiling are highlighted in grey. The control is the C/G base pair.

Mismatch		Number of Hydrogen Bonds					
		0	1	2	3	4	5
Control	Control55	0.00%	0.01%	0.26%	55.15%	44.57%	0.00%
	Mod55	0.00%	0.01%	0.19%	54.13%	45.62%	0.05%
AA	Control55	24.13%	75.87%	0.00%	0.00%	0.00%	0.00%
	Mod55	3.99%	96.01%	0.00%	0.00%	0.00%	0.00%
AC	Control55	2.28%	97.53%	0.18%	0.00%	0.00%	0.00%
	Mod55	0.93%	99.07%	0.00%	0.00%	0.00%	0.00%
AG	Control55	0.11%	0.54%	57.04%	42.30%	0.00%	0.00%
	Mod55	0.01%	0.22%	56.05%	43.72%	0.00%	0.00%
CA	Control55	2.66%	96.90%	0.39%	0.05%	0.00%	0.00%
	Mod55	0.39%	99.61%	0.00%	0.00%	0.00%	0.00%
CC	Control55	28.27%	64.76%	6.97%	0.00%	0.00%	0.00%
	Mod55	26.16%	65.93%	7.91%	0.00%	0.00%	0.00%
CT	Control55	7.02%	70.62%	19.83%	2.54%	0.00%	0.00%
	Mod55	4.29%	87.22%	7.47%	1.02%	0.00%	0.00%
GA	Control55	0.04%	0.48%	60.22%	39.25%	0.00%	0.00%
	Mod55	0.24%	1.82%	60.76%	37.18%	0.00%	0.00%
GG	Control55	0.03%	3.09%	69.37%	27.51%	0.00%	0.00%
	Mod55	0.03%	3.01%	59.64%	37.33%	0.00%	0.00%
GT	Control55	0.13%	0.91%	63.67%	35.29%	0.00%	0.00%
	Mod55	0.04%	1.76%	70.20%	27.99%	0.00%	0.00%
TC	Control55	7.48%	70.31%	18.42%	3.79%	0.00%	0.00%
	Mod55	5.66%	77.44%	14.78%	2.13%	0.00%	0.00%
TG	Control55	0.04%	1.11%	58.23%	40.62%	0.00%	0.00%
	Mod55	0.59%	1.27%	61.19%	36.95%	0.00%	0.00%
TT	Control55	0.39%	3.49%	96.13%	0.00%	0.00%	0.00%
	Mod55	0.80%	10.96%	88.23%	0.00%	0.00%	0.00%

respective reciprocal mismatches—spend a significant proportion of simulation time in a single-hydrogen-bond state, and yet they do not exhibit a strong pinning propensity. Conversely, AG

strongly pins the plectoneme in Control339 and is also the most stable mismatch, maintaining the highest proportion of the three-hydrogen-bond state. These contrasting cases suggest that the relative populations of the hydrogen-bonded states alone are not sufficient to predict plectoneme pinning under negative supercoiling.

Base Pair Parameters

Analogous to the tilt and roll parameters being a potential predictor of positive pinning, we sought to understand if the parameters describing bubble formation (bubble, propeller, and opening) were predictive of pinning in the negative supercoiling regime. We hypothesised that flexibility in these parameters in the absence of torsional strain may mean that bubble formation is favoured here under negative supercoiling. However, in general, the average and variance of the parameters for the mismatched base pairs were within that expected from the control simulations (fig. 4.22). The buckle was highly stable for most mismatches and, whilst there were larger deviations in the average propeller, they largely remained within the range of the canonical base pairs. As previously described in the literature [154], the opening parameter was the most significantly affected.

The GG mismatch was the only mismatch to have a significant, negative-only opening, signifying opening of the base pair towards the minor groove. Conversely, mAA was the only mismatch to

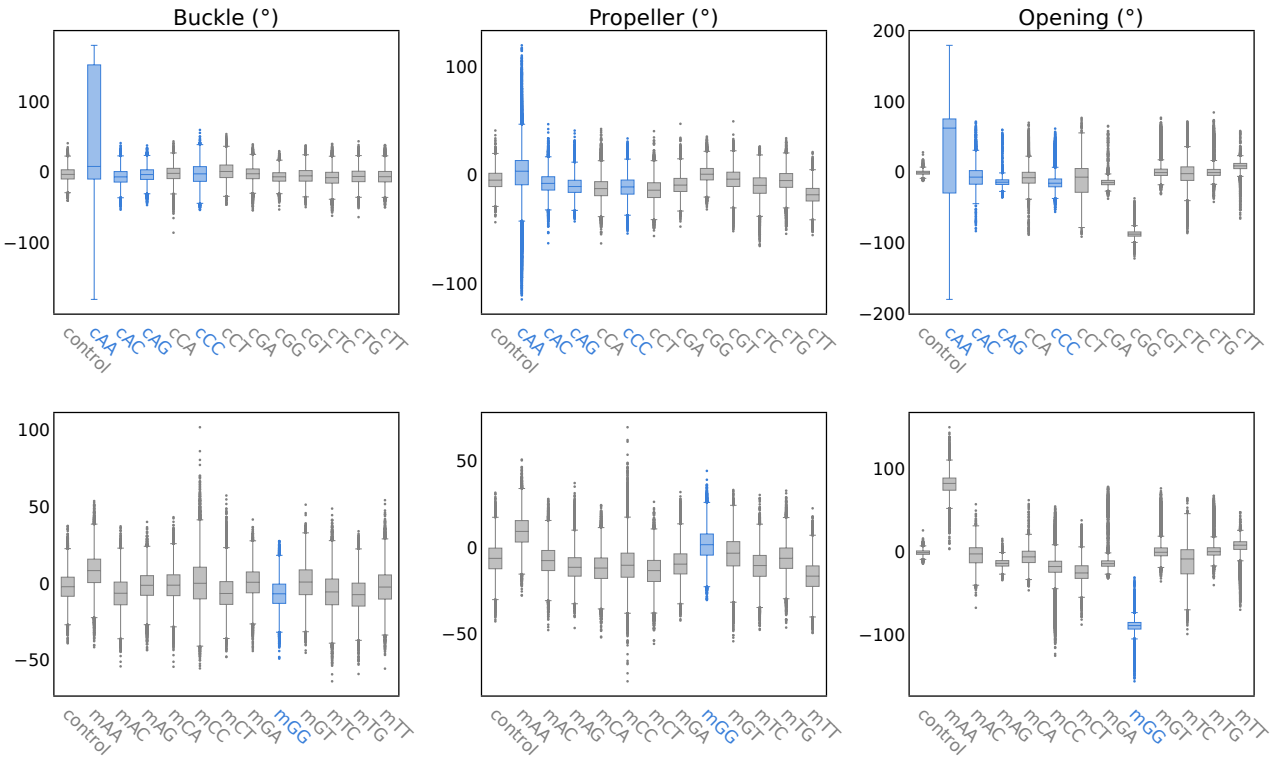

Figure 4.22: None of the average value, direction, or variance of the buckle, propeller, and opening parameters are predictive of plectoneme pinning under negative supercoiling. Box plots of the buckle, propeller, and opening base pair parameters calculated over the three replicas of each simulation in Control55 **(top)** and Mod55 **(bottom)**. Simulations where the mismatch pins under negative supercoiling are represented in blue.

have a significant, positive-only opening towards the major groove. However, from the trends seen in the Control55 sequence, none of the direction, average, or variance of the opening parameter appears to be predictive of strong pinning propensity. It is likely coincidental that the only mismatch to pin in Mod339, GG, is the only mismatch to have a large negative opening as the same trend did not hold in the control sequence.

As bubble formation requires the opening of at least three mismatches, we also calculated the same parameters (fig. A.22) as well as the hydrogen bonding patterns (table A.23) for the nearest neighbours. Here, all average parameters were within the range seen in the control. Consistent with that seen for the mismatched base pair, the largest deviations were seen for the Opening parameter where variances tended to be larger than in the control. However, no trend was found between the magnitude of the variance and the propensity to pin the plectoneme under negative supercoiling. The hydrogen bonding of the base pair in the 5' direction, T/A, is not significantly altered by the presence of mismatches. In the undamaged sequences, there are two hydrogen bonds present for 96.9-98.7% of simulation time across the two sequences and the majority of mismatches retain over 90% of simulation time with two hydrogen bonds. Only mAA has less than this threshold with 10% of simulation time in a one-hydrogen-bond state. However, this was not a significant enough deviation to drive plectoneme pinning. In the 3' direction, as non-Watson-Crick hydrogen bonds are also considered, C/G has a population corresponding to 43.2-47.3% simulation time with four hydrogen bonds, and 51.4-56.5% with three in the control sequences. Unexpectedly, the mismatches actually tend to increase the time spent with four hydrogen bonds, with the largest deviation seen for mTT with 49% of simulation time. The only mismatch to significantly—in comparison to the variation seen across the two undamaged controls—reduce the population of the four-hydrogen-bond state is mGA with a reduction of 3.5%, However, overall, these variations are negligible as the three-hydrogen-bond state dominates in all cases.

Large deviations in the opening of the mismatch describe the 'breathing' of the base pair and it has previously been well-described in the context of DNA mismatches [154, 496]. Consistent with conventions in the literature [154], we quantified breathing percentage as the percentage of simulation time in which the opening deviated significantly from the value observed in canonical B-form DNA (table 4.23).

Unsurprisingly, the large structural deviations seen in the AA and GG mismatches resulted in 94-100% breathing. However, beyond this, breathing is not predictive of plectoneme pinning. In Control339, AG undergoes breathing for as little as as 2.6% simulation time and it is able to pin the plectoneme. In contrast, the relatively large proportions of breathing of 33% in TC, coupled with a low stacking energy and Morse potential, fail to pin. This could be attributed to the strongly pinning surrounding sequence in Mod339 as mTC has a small, local peak in plectoneme propensity (fig. 4.24). Notably, the similarly large breathing frequency in mAA compared to mGG, coupled with having only one hydrogen bond for 96% of the simulation

Table 4.23: ***The breathing frequencies of a mismatch does not correlate to pinning strength under negative supercoiling.***
The percentage of the three repeats of the simulation that a mismatch is breathing in the given sequence. The mismatches that pin under the corresponding sequence are highlighted in grey.

	Breathing	
Mismatch	Control55	Mod55
Control	0.03%	0.003%
AA	93.7%	100.0%
AC	13.0%	9.5%
AG	2.6%	1.7%
CA	10.2%	4.4%
CC	18.0%	29.4%
CT	36.9%	56.3%
GA	2.3%	3.4%
GG	100.0%	100.0%
GT	6.9%	10.0%
TC	19.4%	33.4%
TG	5.1%	6.3%
TT	8.7%	10.7%

time, suggests that this mismatch should pin the plectoneme if breathing were to be predictive of pinning strength. Further, mAA pins the plectoneme under positive supercoiling and, as such, has the required intrinsic curvature to preferentially locate to the plectoneme tip.

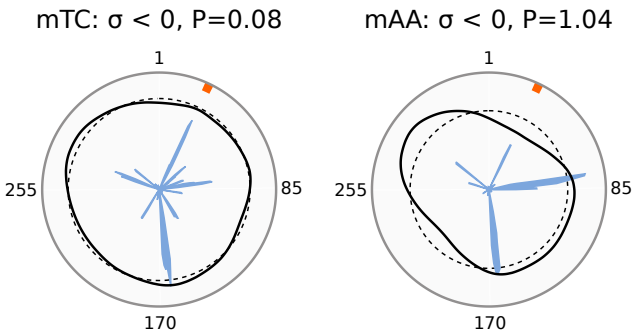

Figure 4.24: Whilst mTC shows significant breathing in the 55-mer but only a small local peak in plectoneme propensity, mAA exhibits 100% breathing and no competition for the plectoneme tip.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mTC **(left)** and mAA **(right)** mismatches under negative supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The P values are normalised such that the undamaged Mod339 under positive supercoiling is equal to 1. The location of the mismatch is represented by an orange band.

4.5 Single mismatch opening does not drive plectoneme pinning strength

Thus far, in the discussion of predicting plectoneme pinning under negative supercoiling, we have actually been predicting bubble formation assuming that it drives plectoneme formation. However, as previously discussed, and, as evident in numerous plectoneme propensity plots, bubble formation is seen at the mismatch site even in cases that do not pin the plectoneme (fig. 4.25).

Although plectoneme pinning is influenced by the interplay of multiple regions of bubble formation, given that the surrounding sequence is the same, it is surprising that a significant bubble formation can successfully pin the plectoneme in one mismatch, while a different mismatch—despite also generating considerable bubble formation—does not result in any local increase in pinning propensity. This suggests that absolute bubble formation alone is insufficient to explain pinning behaviour under negative supercoiling. Therefore, we investigated the role of relative bubble densities in determining plectoneme pinning. Specifically, we investigate if bubble formation in areas distant to the mismatch out-compete the defect formation at the mismatch site.

In the previous chapter, we showed that the relative bubble density correlates to the degree of twist absorption in that region (fig. 3.13). This relationship is a consequence of bubbles being inherently completely and locally untwisted. Thus, regions containing fewer but more dominant bubbles exhibit a greater twist absorption—defined by a lower local twist—compared to regions with a lower relative bubble density (fig. 4.26). This is not the first time that mismatches have been shown to absorb twist [288]; however, the current literature has not yet investigated how the twist absorption influences plectoneme pinning.

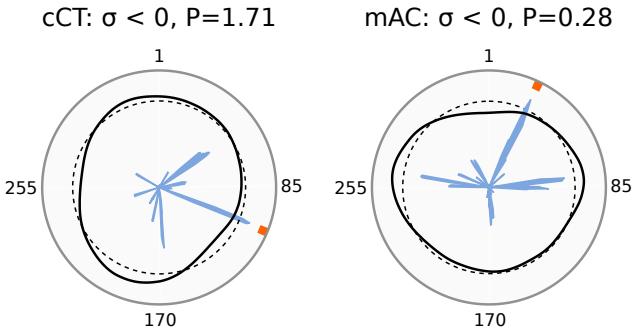

Figure 4.25: Significant bubble formation at the mismatch site does not necessarily cause pinning under negative supercoiling
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cCT **(left)** and mAC **(right)** mismatches under negative supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The P values are normalised such that the undamaged Mod339 under positive supercoiling is equal to 1. The location of the mismatch is represented by an orange band.

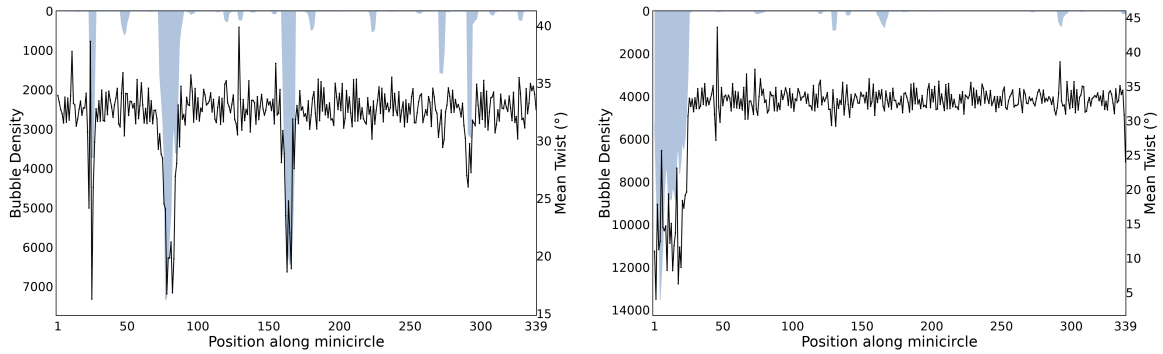

Figure 4.26: The twist absorption correlates directly to the relative bubble density. The bubble densities (blue) and twist (black) averaged over all simulations of the gAA **(left)** and gGG **(right)** mismatches under negative supercoiling. The mismatches are inserted at position 25 within the minicircle.

The ability of DNA to buffer against significant changes in conformation, including overall twist, means that the twist over all simulations remains constant for a given linking number. Thus, we can use local twist absorption as a quantitative proxy for the relative bubble density and bubble localisation (fig. 4.27). The only mismatch with a significantly different overall twist is mGG at the most negative superhelical densities (fig. A.24). However, our twist absorption proxy is averaged over all superhelical densities, where there is no discernible difference (fig. A.25).

We have previously discussed—qualitatively, through visual inspection of the pinning propensity plots—the significant bubble formation at the mismatch site for both cCA (fig. 4.18) and cCT (fig. 4.25) despite the lack of local pinning propensity. Here, quantitative analysis in the form of twist absorption confirms that bubble localisation does not drive plectoneme pinning under

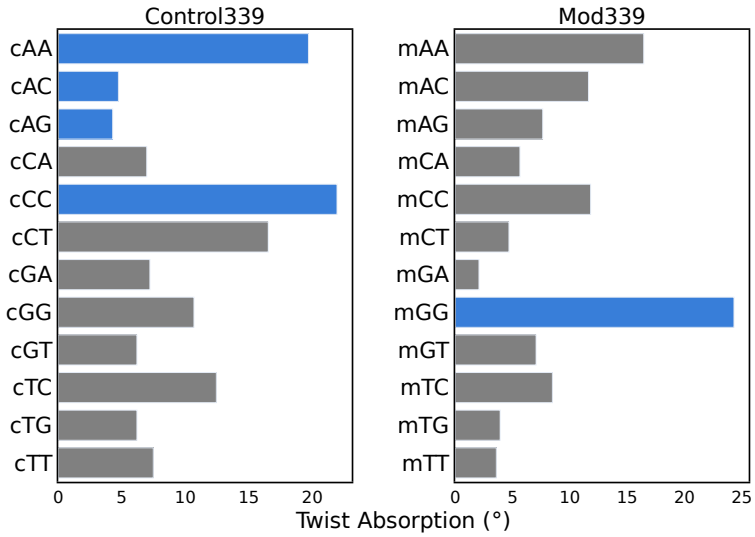

Figure 4.27: The twist absorption—and, therefore, degree of bubble localisation—does not correlate to the plectoneme pinning strength of the mismatch under negative supercoiling. The twist absorption, defined as the twist at the mismatch subtracted from the baseline twist of the duplex, of the twelve mismatches in Control339 **(left)** and Mod339 **(right)** under negative supercoiling. The mismatches that pin the plectoneme under negative supercoiling are shown in blue.

negative supercoiling. This is mostly clearly demonstrated in the Control339 construct, where the cAG mismatch has the second highest mismatch-driven pinning propensity (fig. 4.17) but the lowest degree of bubble localisation by any of the mismatches (fig. 4.27). The trend is less pronounced in Mod339, where increased competition by the sequence for the plectoneme tip allows only one mismatch to pin the plectoneme. However, consider the mAA and mTC mismatches (fig. 4.24). mAA has the second largest twist absorption, after mGG, and offers little competition for the plectoneme tip. This is compared to mTC which has a relatively small bubble localisation compared to mAA but a small local peak in plectoneme propensity (fig. 4.24). Even in the absence of direct comparisons between pinning and non-pinning mismatches in Mod339, it is clear that bubble localisation does not aid pinning propensity in either sequence.

The Effect of Salt

Previously, in the undamaged constructs, the effect of increasing salt on pinning propensity was variable depending on the pinning landscape. Increasing the ionic concentration resulted a shift of the pinning landscape to be more in line with that under positive supercoiling. Furthermore, we found that this was not driven by changes in bubble size or lifetime. However, we have seen in the mismatched constructs that a relatively high bubble density is not necessarily required or indicative of plectoneme positioning. There have been significant bubble density peaks around many of the mismatches, including in cases where the mismatch does not pin the plectoneme. Conversely, some mismatches with strong pinning propensities have low bubble densities, especially under the positive supercoiling regime. Thus, we investigated the effect of salt concentration to better understand the factors driving plectoneme positioning. The cAC and cGG mismatches were chosen to investigate the effects of salt on pinning propensity. They were chosen due to their differing pinning propensities across the two sequences and directions of supercoiling: cAC pins under the negative supercoiling regime, and cGG does not pin under either; but mGG pins under both regimes with strong bubble localisation. Consistent with the results of the undamaged constructs in the previous chapter, as salt concentration is increased, the twist becomes more positive along with the complementary decrease in writhe (fig. A.26).

Due to the pinning landscapes under positive supercoiling, we would expect neither mismatch to pin at increased salt concentration in Control339. However, the local pinning propensity of both mismatches generally increases with increasing salt concentration. Notably, cAC overall pins the plectoneme at multiple salt concentrations (fig. A.27), whereas cGG increases local pinning propensity but fails to dominate the plectoneme pinning landscape overall (fig. A.28). These results are consistent with the pinning propensity of the mismatches at physiological conditions as cAC pins the plectoneme at 0.2 M, whereas npGG has no local pinning propensity and, therefore, has more to overcome than cAC. Further, it is clear that increasing salt concentration increases the local pinning propensity of the mismatch irrespective of the pinning propensity of the same mismatch under positive supercoiling. The local pinning propensity is also improved with increasing salt concentration under the positive supercoiling regime. This agrees with

previous literature that shows that a 2 nt bubble pins the plectoneme under positive supercoiling more deterministically in a high salt-high force regime than at low salt-low force [288].

However, based on the results from the previous chapter (fig. A.2), we would expect an increase in plectoneme pinning to correlate with greater bubble localisation. However, in cAC, the largest twist absorption is observed at an ionic concentration of 0.3 M (fig. 4.28). This corresponds to the ionic concentration with the lowest local pinning propensity, as the mismatch lies within a local minimum of the plectoneme propensity profile (fig. A.27). Strikingly, the sequence-driven pinning dominates under these conditions and results in a P value of nearly 11.5 relative to the same construct at 0.2 M ionic concentration. The significant bubble localisation at the 0.3 M concentration not only fails to pin the plectoneme but instead drives a substantial sequence-driven pinning signal. This behaviour contrasts with the simulation at 1 M ionic concentration, where the mismatch pins over four times more strongly than at 0.2 M. Yet, at physiological conditions, the twist absorption is comparable to that at 1 M and at 0.4 M: a condition in which the mismatch is within the region of high plectoneme density, although not positioned at the apex.

In the cGG mismatch, the twist absorption is comparable across the 0.2 M, 0.5 M, and 1 M ionic strengths; however, all three resulting plectoneme pinning landscapes differ markedly. At 0.2 M ionic concentration, as previously discussed, significant bubble formation fails to induce any local pinning propensity in cGG. Conversely, at 0.5 M salt concentration, there is a small but distinct peak in pinning propensity centred directly on the mismatch. Finally, at 1 M ionic strength, there is a pronounced local increase in pinning propensity, especially compared to that at 0.2 M salt concentration. However, the mismatch is not centred at the apex of the peak.

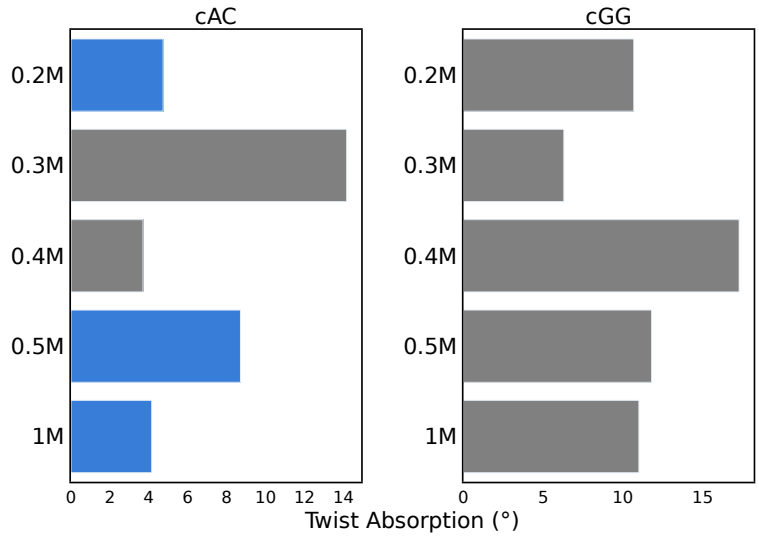

Figure 4.28: The twist absorption—and, by extension, the relative bubble density—by a mismatch does not correlate to the increased pinning propensity at increased salt.
The twist absorption, defined as the twist at the mismatch subtracted from the baseline twist of the duplex, of cAC **(left)** and cGG **(right)** under negative supercoiling. The data is calculated over all simulation replicas under negative supercoiling. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

These differing pinning responses arise despite similar degrees of bubble localisation and an identical surrounding sequence context across all three salt conditions. Thus, we conclude that bubble formation—and, in particular, bubble localisation—is not the primary factor governing mismatch-driven plectoneme pinning. Nonetheless, increasing salt concentration undoubtedly increases the local plectoneme density of the mismatch.

We propose that, in agreement with the previous chapter, the change in plectoneme propensity at increased salt concentrations is writhe-mediated. As the salt concentration increases and the DNA adopts more highly strained conformations with multiple sharp bends, placing the mismatch at these tips is favourable. The advent of multiple regions of high curvature reduces the competition between mismatch and sequence for the plectoneme tip as now, multiple apices are present allowing both the lesion and the mismatch to be positioned there, thus increasing local pinning propensity irrespective of the pinning propensity at physiological conditions.

4.6 Mismatch Length Can Override Sequence Effects Beyond a Threshold

Previously, it has been shown that increasing the number of consecutive mismatches increases the pinning propensity under positive supercoiling [287, 288]. At three consecutive mismatches, this length of hydrogen bonding loss and structural deformation is consistent with the definition of a bubble [106]. Thus, for triple mismatches, the pinning strength of the mismatch may be increasingly driven by the inherent bubble formation, rather than via sequence-mediated effects. Here, we investigate the pinning propensities of two double and one triple mismatch within the Control339 and Mod339 constructs in both directions of supercoiling. For simplicity, symmetrical mismatches were chosen: GG/GG (referred to as 2GG), TG/TG (referred to as 2TG), and CTG/CTG (referred to as 3nt). They were mutated at the same positions of low plectoneme density as the single mismatches. Previously, we suggested that Mod339 may have fewer mismatches able to pin the plectoneme under positive supercoiling—despite the lower P value and, therefore, lesser sequence competition compared to Control339—due to the double pinning effect of supercoiled minicircles. To test this theory, the double and triple mismatches were also inserted directly into the region of high plectoneme density in the Mod339 sequence. They were mutated within the region of the smaller plectoneme propensity peak so that the inherent pinning of the opposite side is in the more favourable area. We further simulated the G/G single mismatch in the same region to determine if it would have the same bubble localisation effect less than ∼60 nucleotides away from its original location. Due to the complete removal of all sequence competition, we expect the mismatches to pin the plectoneme in all simulations of this sequence. Herein, this construct is referred to as inserted339, and the mismatched constructs are labelled with the letter "i" preceding the mismatch species.

The patterns of mismatch-driven pinning clearly demonstrate that both double and the triple

mismatches continue to exhibit sequence-dependent effects (table 4.29). For example, the m2GG double mismatch elicited the same bubble localisation effects as seen in the equivalent mGG single mismatch (fig. A.29). This is consistent with the observations of Control339 where the bubble localisation was not seen in either the single, cGG, or double, c2GG, mismatch (fig. A.30). Furthermore, neither the single T/T mismatch nor the TG/TG double mismatches pinned the plectoneme in either sequence under either direction of supercoiling. Thus, from this highly limited sample, double mismatches appear to retain the same, or similar, sequence-dependent properties characteristic of the corresponding single mismatches (table 4.29). Specifically, the weak pinning strength of the single T/T mismatch may impede the pinning potential of the TG/TG double mismatch. This agrees with results from the literature that have shown that multiple mismatches retain local structural elements of the constituent single mismatches [497]. However, further validation against the TG/TG double mismatch and further multiple mismatches is required. Importantly, the effects of multiple mismatches cannot be assumed to result from a simple combination of the effects of the individual mismatches. Sequence-dependent interactions between the tandem mismatches may give rise to non-additive behaviours. Indeed, some double and triple mismatches have demonstrated surprisingly strong stacking and hydrogen bonding interactions through structural rearrangements that increase local stability [472].

Surprisingly, the triple CTG/CTG mismatch did not pin the plectoneme in the positive supercoiling regime in any sequence, even when inserted directly into the region of high sequence-driven pinning propensity in i3nt. This suggests that, even with up to three consecutive mismatches, the pinning propensity is still subject to sequence effects and is not entirely driven by bubble formation. The underlying mechanism of this remains unclear; however, it may be linked to the poor pinning strength of the C/C single mismatch in the positive supercoiling regime. As previously discussed, the mCC mismatch has a high intrinsic curvature yet low local pinning

Table 4.29: ***The pinning strengths of multiple mismatches are also sequence- and lesion-dependent.*** *A table to show which mismatches do (1) and don't (0) pin the plectoneme under each direction of supercoiling within each sequence. inserted339 refers to the position of the mismatch within the region of high plectoneme propensity in Mod339. For clarity, the mismatches that pin are also shaded in grey.*

	Control339		Mod339		inserted339	
Mismatch	Positive	Negative	Positive	Negative	Positive	Negative
1 nt - GG	0	0	1	1	1	1
1 nt - TT	0	0	0	0		
1 nt - CC	0	1	0	0		
2 nt - GG/GG	1	0	1	1	1	1
2 nt - TG/TG	0	0	0	0	1	1
3 nt - CTG/CTG	0	1	0	1	0	1

strength, which could similarly limit the CTG/CTG triple mismatch's ability to pin the plectoneme. Conversely, the introduction of three consecutive mismatches was sufficient to pin the plectoneme relatively deterministically under negative supercoiling in Control339. The triple mismatch also pinned the plectoneme in Mod339 and inserted339; however, it did so with a low P value. These pinning trends may be serendipitous, and any conclusions drawn here are purely speculative given the small sample of multiple mismatches under investigation. To better understand the mismatch-driven plectoneme pinning across multiple mismatches, further simulations are necessary—particularly those involving different multiple mismatches as well as the corresponding linear constructs to allow for structural characterization of the mismatches in their respective sequence contexts in the absence of supercoiling. However, it is clear that simply increasing the length of the tandem mismatches, at least to 3 nucleotides, is not sufficient to deterministically pin the plectoneme and neglect any sequence-dependent effects of the mismatches. This is evident in the twist absorption data (fig. 4.30), where the triple mismatch m3nt absorbs significantly less twist than the smaller m2GG mismatch. Furthermore, despite the comparable bubble densities of the i2GG and i3nt mismatches, i2GG pins significantly more strongly under negative supercoiling. Together, these results underscore the persistent and dominant role of both the mismatch sequence and the surrounding sequence context, even in the presence of extended mismatched regions.

In the absence of simulations of the linear constructs with multiple mismatches, we use the same stacking and Morse potentials to investigate the factors driving bubble formation in the double and triple mismatches within the Control339 and Mod339 sequences (table 4.31). While these models are not expected to accurately reflect the inserted339 constructs—given that the

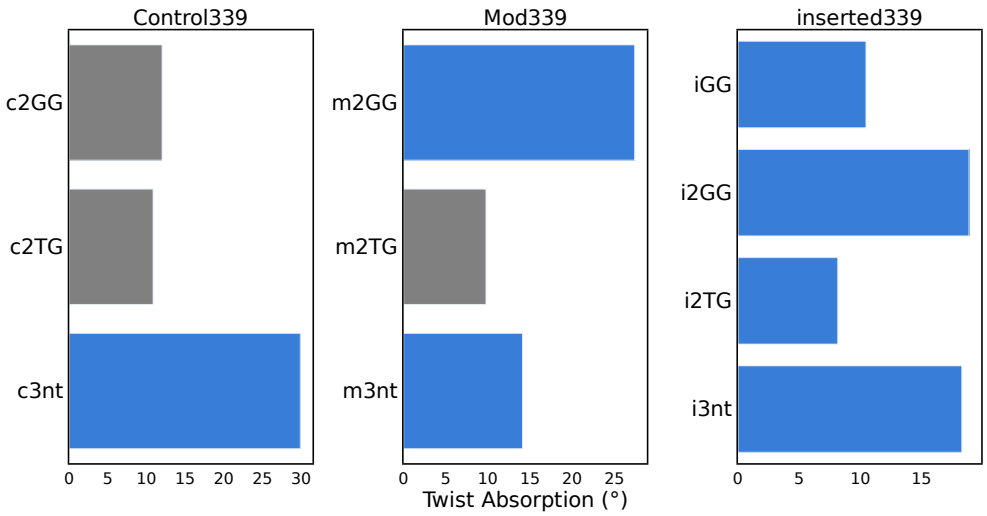

Figure 4.30: Increasing the length of the consecutive mismatches does not directly correlate to bubble localisation as sequence-dependent effects are still seen for up to the triple mismatch.
The twist at the mismatch subtracted from the baseline twist for the double and triple mismatches in Control339 **(left)** and Mod339 **(middle)**, as well as the GG single mismatch in inserted339 **(right)** under negative supercoiling. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

Table 4.31: ***The hydrogen bonding and stacking interactions of the double and triple mismatches do not correlate to plectoneme pinning strength under negative supercoiling in their respective sequences.***
The measured melting temperature, T_i, Morse potential, D, and elastic constants, k, of the double and triple mismatches summed across the nearest neighbour sequence context. The D and k values are summed over each base step in the given sequence context. The T_i is only measured for trimers. The raw data is taken from data from Oliveira et al. [472]. The individual parameters over which the sum is given below are in table A.32. The rows where the mismatch pins under negative supercoiling in either sequence are shaded grey.

Mismatch		Sequence Context	T_i / °C	$\sum D$ / meV	$\sum k$ / eV nm^{-2}
GG/GG	Control339	TGGA/AGGT	53.5[a]	107.36	7.57
	Mod339	TGGG/AGGC	56.1[a]	144.17	6.79
TG/TG	Control339	CTGG/GTGC	54.9[b]	163.06	4.26
	Mod339	TTGC/ATGG	51.7[b]	125.28	6.46
CTG/CTG	Control339	TCTGC/ACTGG	48.0[c]	129.82	6.20
	Mod339	ACTGC/TCTGG	48.0[c]	129.82	5.22

[a]Both are T/A in the 5' direction. Melting temperature is of the trimer with the 3' nearest neighbour.
[b]Both are C/G in the 3' direction. Melting temperature is of the trimer with the 5' nearest neighbour.
[c]Independent of nearest neighbours as T_i is only given in trimer context.

mismatches were introduced directly into regions of strong plectoneme localisation—they also fail to predict bubble localisation within the Control339 and Mod339 sequences. Within the Control339 sequence, both double mismatches exhibited similar degrees of twist absorption despite the TG/TG mismatch forming significantly stronger hydrogen bond interactions. This could be argued that the stronger hydrogen bonding interactions may have been offset by weaker stacking interactions. However, in Mod339, the m2GG mismatch showed a significantly greater degree of bubble localisation compared to m2TG, despite a comparable elastic constant between the two, and the hydrogen bonding interactions in m2GG being significantly stronger. Unlike the single mismatches that were in the same nearest neighbour sequence context, here there are significant differences between the hydrogen bonding interactions for the same mismatch across the two sequences. Furthermore, the stacking of the TG/TG mismatch is greatly influenced by sequence context; however, despite these significant differences, the 2TG mismatch elicits a comparable degree of bubble formation across both sequences. Even in cases where the interactions are similar across the two sequences, such as is the case for 3nt, there are varied bubble localisation effects—as measured by the twist absorption—depending on the sequence context. Specifically, c3nt absorbs significantly more twist than m3nt. These results indicate that bubble formation is governed by a more complex interplay of factors than hydrogen bonding strength and stacking potential alone—even in the case of multiple mismatches.

Finally, to further investigate the hypothesis that plectoneme pinning is not solely driven by

bubble localisation, the same double and triple mismatches were inserted into the Strong339 sequence, as well as the G/G single mismatch. As demonstrated in the previous chapter, Strong339 has a strong, bubble-driven plectoneme pinning region against which the mismatches must compete. In keeping with the previous naming conventions, these mismatched sequences are denoted with an "s" prefix. Surprisingly, under positive supercoiling, the s2TG mismatch pinned the plectoneme strongly under positive supercoiling despite exhibiting effectively no bubble formation (fig. A.33). This is notable as the TG/TG mismatch does not pin in either the Control339 or the Mod339 sequence under either direction of supercoiling. Conversely, the s2GG and s3nt mismatches both exhibited a large degree of local bubble density despite neither having a local peak in pinning propensity. Notably, the CTG/CTG mismatch failed to pin the plectoneme under positive supercoiling across all three sequences tested. As expected, the strong sequence competition for the plectoneme tip—driven by significant bubble formation—prevented any of the mismatches from pinning under negative supercoiling.

To further investigate the relationship between bubble formation by mismatches and pinning strength, the length of the tandem mismatches was extended to four, six, seven, and eight nucleotides (fig. A.34). Here, for simplicity, one strand of the duplex was mutated in each case to generate symmetrical mismatches of increasing length. For the four, six, and seven consecutive mismatches, there was a local peak in pinning propensity that increased with increasing mismatch length under negative supercoiling. For eight consecutive mismatches, this was sufficient to overall pin the plectoneme. This trend is mirrored by a generally linear increase in twist absorption as the length of the tandem mismatches increases (fig. 4.32). This

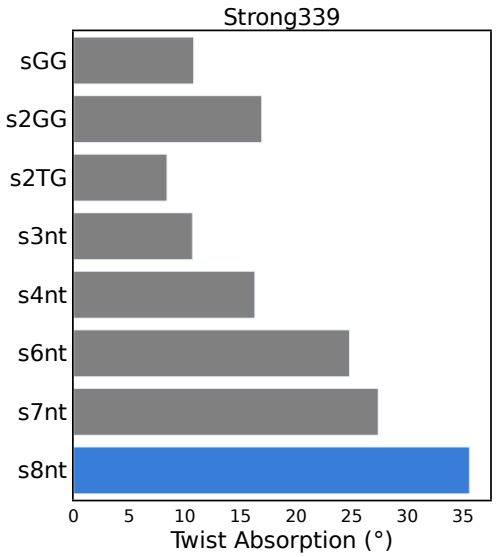

Figure 4.32: Increasing the length of the consecutive mismatches increases the twist absorption by the mismatch.
The twist absorption, defined as the twist at the mismatch subtracted from the baseline twist of the duplex, of the multiple mismatches in the Strong339 sequence under negative supercoiling. The blue bars represent the simulations where the mismatch pins the plectoneme in the negative supercoiling regime.

may indicate that, at this length scale, the sequence-dependent effects of the mismatch may be outweighed by the influence of mismatch length and, thus, bubble formation. However, further sequences and mismatches would need to be investigated at this length scale to verify this.

The twist absorption by s2GG under negative supercoiling is comparable to that of s4nt; however, s2GG does not show the same local peak in pinning propensity. This suggests the existence of two distinct mechanisms contributing to plectoneme pinning across two different length scales. Firstly, in short-length tandem mismatches, sequence-dependent effects may dominate and elicit pinning propensities driven by something other than bubble formation. As evidenced through the results in this chapter, these factors are as of yet unknown, but they are separate to bubble localisation. Secondly, in longer mismatches, the length of the mismatch and thus the inherent length of the bubble formed drives the mismatch-dominant pinning. It is probable that in this case, the threshold length at which bubble-driven pinning becomes dominant may vary depending on the local sequence environment and the strength of competing pinning elements. Notably, mismatched-induced bubbles, even in short mismatches, are often highly stable, as evidenced by pronounced peaks in the bubble density plots. However, their ability to pin the plectoneme may be limited by sequence-specific effects, especially when competing with other, strong pinning regions within the same construct.

The stable, shorter bubbles induced by mismatches may simply be insufficient in length to reliably outcompete the less stable, yet longer regions of instability elsewhere in the duplex. Thus, the extent to which the mismatch-induced disruption propagates into the surrounding sequence may play a critical role in determining plectoneme pinning strength. However, this idea is not fully supported by the trends in bubble localisation. As evidenced by the increasing twist absorption by the longer mismatches in Strong339, the twist absorption metric is able to capture both the stability of the bubble (i.e. the height of the bubble density peak) and the degree of conformational disruption (i.e.. the width of the bubble density peak). However, as previously shown, twist absorption does not consistently correlate with pinning strength. This further reinforces the notion that plectoneme pinning is governed by a multifactorial interplay between mismatch characteristics, sequence context, and broader mechanical factors.

Under positive supercoiling, eight consecutive mismatches were also sufficient to elicit strong plectoneme pinning. However, this may be a result of the specific mismatched sequence rather than a factor of mismatch length. This hypothesis is supported by seven successive mismatches not producing even a small local peak in plectoneme propensity, while six tandem mismatches did (fig. 4.33). This suggests that, even at these extended mismatch lengths, sequence context continues to influence pinning behaviour under positive supercoiling.

Finally, we aimed to disentangle the effects of bubble size and stability on the twist absorption. In the previous chapter, we found that bubble size drives plectoneme pinning strength in the undamaged constructs; and, in the literature, bubbles of increasing size had an increasing likelihood of pinning the plectoneme [288]. Furthermore, we hypothesise a point at which

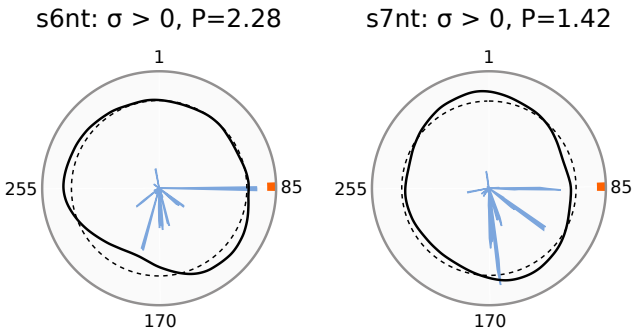

Figure 4.33: At long lengths of tandem mismatches, sequence-effects may still dominate the plectoneme positioning landscape.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the s6nt mismatch **(left)** and s7nt mismatch **(right)** under positive supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Strong339 construct under positive supercoiling is equal to 1.

the size of the bubble drives plectoneme positioning more than sequence effects. Thus, we calculated the width of the bubble from the bubble density plot under negative supercoiling. Firstly, within the single mismatches, only the mTG mismatch did not form a bubble under negative supercoiling (fig. 4.34). Here, a 3nt bubble formed adjacent to the mismatch, occupying the base pairs in the 5' direction. Interestingly, in the linear constructs, the TG mismatch retained the highest probability of three hydrogen bonds (table 4.21); however, it did so more in Control55 than Mod55. And whilst the TG mismatch also exhibits low breathing propensity (table 4.23), it is not the lowest. Even more notable is the cTG mismatch that has a significantly

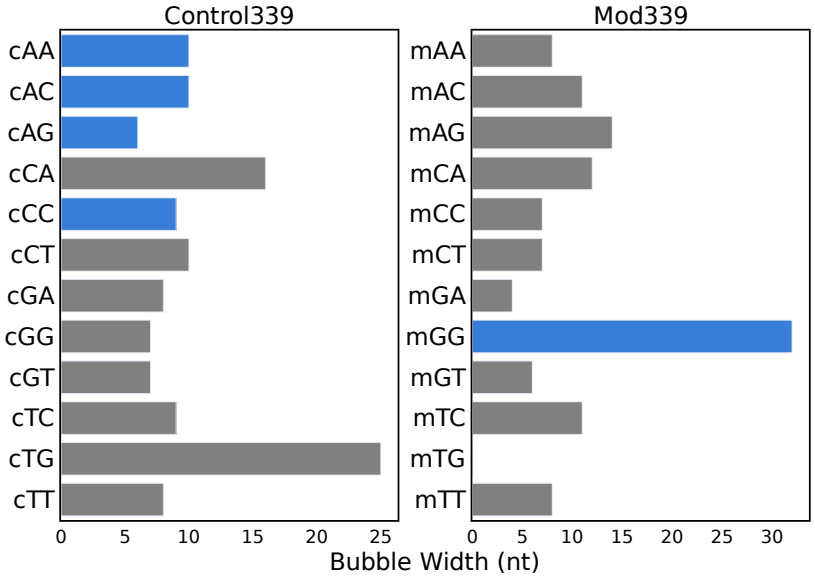

Figure 4.34: The size of the bubble at the mismatch does not correlate to pinning propensity.
The width of the bubble density peak at the mismatch site under negative supercoiling for the single mismatches. The width is given in terms of the number of nucleotides (nt) that the bubble spans. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

longer bubble width than any other mismatch in the Control339 sequence. Overall, the bubble width does not correlate with the pinning propensity of the single mismatches. In the previous chapter, the undamaged constructs showed no clear trend, in terms of bubble properties, with salt concentration. We concluded that bubble size was specifically unaffected by ionic strength. Here, the formation of the bubble at the mismatch is further unaffected by salt concentration (fig. A.35), despite the mismatches having stronger pinning propensities at higher ionic strengths. This reinforces our hypothesis merely forming a bubble at the site of a single mismatch is not sufficient to drive the plectoneme pinning propensity.

It is expected that increasing the size of the mismatch would increase the width of the resulting bubble; however, in agreement with the results seen for twist absorption, mismatches up to three nucleotides in length are still subject to sequence effects (fig. A.36). Here, the width of the bubbles is all beyond the length of the mismatch; thus, the degree to which the lesion disturbs the surrounding sequence remains sequence- and mismatch-dependent. Beyond this, even for mismatches of up to eight nucleotides in length, the disruption of the surrounding sequence remains more mismatch-dependent than the trend in bubble absorption (fig. 4.35). Surprisingly, a 4-nucleotide mismatch caused greater disruption—evidenced by the bubble width—than the mismatch twice its length. Furthermore, this 16 nt bubble was the largest in the Strong339 sequences, whereas the mGG mismatch was able to cause disruption along 30 consecutive base pairs. The mGG mismatch has been shown to cause significantly more disruption than any other mismatch; nevertheless, the mTG single mismatch is still able to create a larger bubble than the

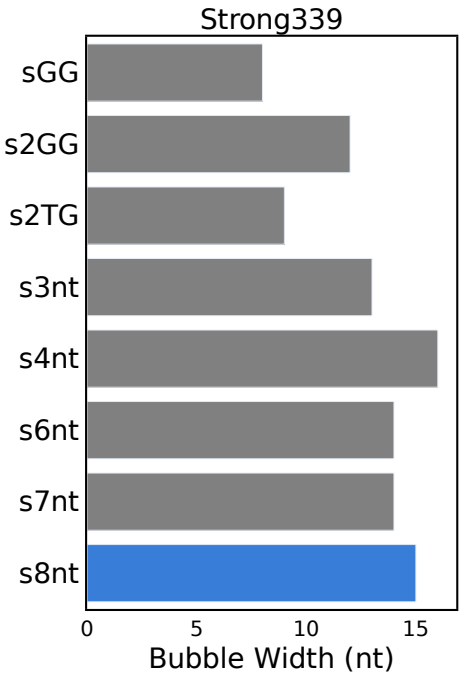

Figure 4.35: The size of the bubble shows no correlation to the size of the mismatch within it. The width of the bubble density peak at the mismatch site under negative supercoiling for the mismatches in the Strong339 sequence. The width is given in terms of the number of nucleotides (nt) that the bubble spans. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

4 consecutive mismatches, and the mCA and c2GT create a bubble of comparable size. Thus, even for large runs of consecutive mismatches, the degree of disruption that this causes seems to remain dependent on the nature of the mismatch itself. Further investigation of multiple mismatches would be required to confirm this. One explanation to the lower-than-expected bubble widths in the Strong339 sequence is the competition with the bubble-driven pinning propensity of the surrounding sequence. Here, a wide bubble density peak in the sequence may limit the disruption at the site of the mismatch as further bubble formation is not required. The results in Mod339 agree with this hypothesis as, except for the strong deformation-inducing mGG, the 2TG and 3nt mismatches have smaller bubble widths than the respective mismatches in Control339. Furthermore, the P value of Mod339 is nearly 6 times that of Control339 under negative supercoiling.

Finally, for all twist absorption and bubble width results presented here, the trends do not exactly agree with one another. Specifically, similar bubble widths can afford very different degrees of twist absorption, and this is further uncorrelated to mismatch pinning propensity—even more so than the twist absorption.

4.7 Discussion

By studying the structural perturbation of linear DNA as a result of single mismatches, we have attempted to understand the varying pinning strengths of the different mismatches across different sequences. Previous studies on mismatches have focused on nearest- or next-nearest-neighbour models [442, 472]; however, our results provide evidence to support the hypothesis that subtle mechanical changes must be propagated along longer length scales and can cause significantly different effects even when being contained within the same surrounding sequence on a scale of hundreds of base pairs. Despite the ability of the duplex to well-absorb variations in local structure due to mismatched DNA [154], our results highlight that a subtle change can elicit large global effects when subject to torsional strain. However, we have highlighted a fundamental disconnect in that a strong local distortion in the form of bubble formation is not, on its own, predictive of strong pinning propensity. Notably, mismatches of increasing length show a near-linear increase in twist absorption beyond a certain threshold. This suggests that longer mismatches may eventually overcome sequence effects through a broader structural disruption. We suggest that, at a threshold defined by the strength of the competing surrounding sequence, the plectoneme pinning propensity of shorter mismatches is dominated by a sequence-driven effect independent of bubble density, whereas longer mismatches induce sufficient disruption to drive pinning through bubble formation alone. This is a novel idea that has not previously been considered in the literature. To date, the working hypothesis generally accepted by the field is that increasing the length of the tandem mismatches increases their pinning strength. However, our results have revealed a complicated interplay between the sequence of the mismatch and that of the surrounding DNA in governing plectoneme pinning strength.

This is the most in-depth study of mismatch-driven plectoneme pinning to date, and the first to consider both the nature of the mismatch and the surrounding sequence context. Dittmore *et al.* are the only group to study the pinning of a specific single mismatch [287] rather than a coarse-grained bubble. They showed, using magnetic tweezers, that a single AA mismatch is able to pin the plectoneme under high force-high salt conditions in positively supercoiled DNA. Our results are in support of this with the AA mismatch pinning Mod339 under positive supercoiling, and Control339 under negative supercoiling. However, we have shown that modelling a mismatch as a bubble is not completely representative or sufficient to investigate mismatch-driven plectoneme pinning as we have seen plectoneme pinning in instances where there was little bubble formation, and we have seen occasions where strong bubble formation at the mismatch did not result in the mismatch pinning the plectoneme. Furthermore, the pinning of a plectoneme by a mismatch has also been attributed to their ability to kink and lower the required bending energy of plectoneme tip formation [104, 288]. However, owing to our atomic resolution and study of sequence-effects, we have shown that the process is not that simple.

We propose that the strength of the pinning propensity of a mismatch may partially dictate its repair efficiency by influencing its recognition by MutS. To our knowledge, there has not been an assay developed to evaluate the relative mismatch repair rates or efficiencies that considers the role of DNA supercoiling. However, the TT and CC mismatches have been identified as especially poorly repaired across a range of studies [441–443, 465]. Interestingly, the TT mismatch was one of only two mismatches that did not pin the plectoneme under either sequence or direction of supercoiling. Even more notable is the case of the CC mismatch where there were strong structural deformations in the linear constructs, including high intrinsic curvature and flexibility, moderate amounts of breathing, and a significant population with no hydrogen bonding within the mismatch. Despite these factors that, in theory, would promote plectoneme pinning, it failed to pin the plectoneme unless there was effectively no sequence-based competition. This was especially notable under positive supercoiling where the CC mismatch elicited the highest degree of curvature and flexibility. This suggests that certain mismatches may be intrinsically poor competitors within supercoiled systems and that current metrics alone may be insufficient to predict this behaviour. Our focus thus far has been on the structural parameters that drive plectoneme pinning; however, it may be the case that there is some structural or mechanical factor that is impeding the pinning that we have not considered. This is further supported by the triple mismatch involving C/C not pinning the plectoneme under positive supercoiling in any of the sequences investigated, including when inserted directly into a region of high local pinning propensity. Thus, it is plausible that its inability to compete for the plectoneme tip in supercoiled DNA could partially explain its reduced repair efficiency. Finally, the GG mismatch has been noted to be especially efficiently repaired in multiple studies [441, 442] and has the strongest pinning efficiency seen in any of our constructs. Overall, our results support the hypothesis that supercoiling may explain part of the variation in mismatch repair efficiencies due to the combination of mismatch-, sequence- and supercoiling direction-dependent natures

of mismatch pinning strengths. A definitive test would require an assay capable of measuring mismatch repair efficiencies in supercoiled DNA—our results are in support of this now being a logical next step for future studies.

The position of insertion of the mismatch within Strong339 was non-optimal as, here, the sequence was specifically designed to have two significantly different plectoneme positioning landscapes across the two directions of supercoiling. Therefore, the position for the mismatch was chosen to be orthogonal to the strongly-pinning region under negative supercoiling. This was chosen as the competition between the bubble-driven sequence pinning and the mismatch for the plectoneme tip was the key aim in designing this sequence. Thus, under positive supercoiling, the mismatch is placed within a region of high local plectoneme propensity. Therefore, the pinning propensities of the mismatches under positive supercoiling in Strong339 may be affected by the underlying sequence. However, both the results of Strong339 under positive supercoiling, and inserted339 under both directions of supercoiling, show that merely inserting a mismatch into a high plectoneme density region is not necessarily sufficient to induce mismatch-driven plectoneme pinning.

We have demonstrated, for the first time, that the directionality and sequence polarity are also important factors when investigating plectoneme pinning propensities. We observed asymmetric effects where the same mismatch showed different pinning propensities depending on whether it was mutated as, for example, a TC or a CT mismatch. This indicates the importance of the orientation of the mismatch within the helical context, possibly due to the effect of this on the base stacking interactions. This is often overlooked in the literature, with reciprocal mismatches being considered equal [442, 451, 455], but we suggest that it is an important consideration for future studies investigating both supercoiling and mismatch repair.

Finally, we acknowledge the limitations within this analysis that arise from using DNA minicircles as a model. This is primarily the double-pinning effect, where the interpretation of pinning propensity is complicated by needing to consider the favourability of pinning the opposite apex. Furthermore, it is difficult to quantitatively discuss situations in which the mismatch increases local pinning propensity but, as we are simulating just a single plectoneme, it does not pin overall. These situations may arise due to the three-apex conformations, which were more common in the mismatched constructs than in the undamaged duplexes. However, this has allowed for a discussion of sequence-driven competition for the plectoneme tip which will be relevant to biological questions. For example, mismatch repair has been shown to be more efficient in areas of the genome that have a high GC content [451, 498]. Our results suggest that—should supercoiling have the hypothesised role in DNA damage detection—this may be attributed to overall less competition with the sequence for the plectoneme tip. This interpretation is further supported by the observation that, in general, when a single, or even a double, mismatch must compete with the native sequence for the tip position, it rarely succeeds in pinning the plectoneme. Overall, our results can offer insights into the structure and dynamics of supercoiled

DNA, mismatch-containing DNA, and, potentially, help to explain the varying rates of mismatch repair across the genome.

We have proposed that the changes in pinning propensity observed at higher salt concentrations are mediated by writhe, owing to the formation of multiple, sharply bent apices. As we are simulating a single plectoneme, we cannot study the balance of single versus multiple plectoneme formation, which is known to depend on the ionic strength [104, 409]. Specifically, at low salt concentrations, multiple plectonemes can coexist, while at high salt, writhe is stored as a single plectoneme. The presence of multiple high-curvature regions in our more writhed conformations at high ionic strength may, therefore, be specific to the minicircle model. Nevertheless, the observed increase in mismatch pinning propensity at higher salt concentrations is consistent with findings from linear DNA systems under torsional strain [288], suggesting that our conclusions may still extend to biologically relevant DNA topologies.

However, our pinning propensity (P) metric does not quantitatively capture these local increases at the mismatch site. The P metric reflects the regularity of the conformation and its sequence context, rather than the plectoneme density. This is a further limitation of the minicircle model: in cases where both the sequence and the mismatch site promote plectoneme formation, the P value may decrease, despite an actual increase in local pinning at the mismatch. While this quantitatively captures that the original sequence context no longer dominates pinning, it obscures the contribution of the mismatch. As a result, quantifying increases in local pinning at mismatches remains challenging in the minicircle system.

Surprisingly, there were inconsistencies between mismatch pinning strength and the local structural parameters expected to drive plectoneme pinning, namely curvature, flexibility, hydrogen bonding, and stacking. Whilst this does not account for the lack of trend in all cases, there may be a bias in the calculation of intrinsic curvature. Due to the need to exclude bases subject to high curvatures at the ends of the duplex, intrinsic curvature was calculated over 30 nucleotides. This value is less than the ∼40 nt peak that we consider when determining if a mismatch pins or not. However, the largest inconsistency regarding intrinsic curvature was that of mCC, where a large curvature did not induce plectoneme pinning. Thus, it is likely that this bias does not heavily impede our results. Furthermore, the choice of two *syn*/*anti* conformations may be the reason behind their high calculated breathing frequencies; however, we have shown that the trends in the linear constructs in the absence of torsional strain are not predictive overall of the pinning propensity of the mismatch. Although rare, the mismatches within our simulations have undergone spontaneous transitions between the *syn* and *anti* conformations. This was not seen in previous simulations of DNA mismatches [154]; however, Rossetti *et al.* used the predecessor force field, *bsc0*, that was known to have a non-optimal *syn-anti* equilibrium [344]. Furthermore, they were concerned with the *anti*-to-*syn* transition which may have a different barrier to the *syn*-to-*anti* transitions presented here. *Parmbsc1* can now produce stable Hoogsteen conformations on the timescales relevant to this thesis, as well as well-reproduce

high-level QM data in terms of both geometries and energies [344]. Additionally, the reduced stability and friction of implicit solvent—and thus, the wider conformational space sampled by the DNA—coupled with the ability of the mismatches to form bubbles and flipped-out bases, reinforces the ability of the mismatches to undergo individual conformational changes. It is unlikely that the choice of stereochemistry greatly impacts the results. To verify this, it may be necessary for the mismatches to be compared from both *syn*/*anti* and *anti*/*anti* starting conformations within the supercoiled minicircles. However, we can be confident that the high degree of bubble localisation in mGG is not an artifact of the *syn-anti* conformation, as the same effect was not seen in cGG, iGG, or sGG nor either of the AA-containing sequences.

Overall, our results point to a nuanced interplay of local structure disruption, mismatch length and directionality, and competition with the surrounding sequence. While we can identify general trends—for example, Mod339 impedes mismatch-driven pinning under negative supercoiling via strong competition with sequence—we lack a comprehensive framework for predicting when a mismatch will pin. Extending the Dekker or SIDD models to explicitly incorporate mismatches might help to clarify these long-range influences and allow us to further investigate the balance between the mismatch- and sequence-driven effects. One such explanation may be that the extremes of the expected structural parameters, such as the high intrinsic curvature and flexibility in mCC, and the high breathing in mAA, that do not cause plectoneme pinning, may prefer structural deformation at the mismatch to dissipate torsional strain in addition to writhing and plectoneme positioning. Specifically, torsional strain is preferentially dissipated via bubble formation over writhe [40]. Thus, more favourable deformations at these mismatches in addition to the sequence-driven pinning may lower the energy of the conformation. Ultimately, we propose that mismatch-driven plectoneme pinning on shorter length scales is not a function of the local structural disruption of the mismatch itself, but of how the mismatch interacts with and competes against the broader sequence context under supercoiling.

4.8 Summary and Future Work

We have demonstrated, for the first time, that the plectoneme pinning strength of single mismatches depends on both the identity of the mismatch and on the surrounding sequence. However, somewhat unexpectedly, we found no consistent correlation between pinning strength and the expected factors driving pinning propensity: intrinsic curvature, flexibility, and bubble formation. In Chapter 3, we established that the salt response of undamaged DNA is sequence-dependent—in line with the pinning propensity of the same sequence under positive supercoiling—but found no clear influence on bubble formation. Interestingly, mismatched DNA shows a sequence-independent increase in local pinning propensity with increased salt concentration, affecting two distinct mismatches with differing bubble localisation effects. Together, our results suggest that the underlying mechanisms driving single mismatch plectoneme pinning are not solely governed by bubble formation. However, our current data set is limited,

and further simulations would be needed to determine how salt modifies the conformation or stability of mismatches and their surrounding duplex. Specifically, simulations are needed of more mismatches under increased salt concentrations, along with the mismatches in the linear DNA at increased ionic concentration in explicit solvent. The change in properties upon salt formation may offer an explanation as to what interplay of factors is driving plectoneme pinning under negative supercoiling at physiological conditions.

Secondly, we propose that a threshold mismatch length—specific to the surrounding sequence context—determines whether mismatch pinning is sequence-driven or driven by bubble length. Below this threshold, pinning appears to rely on the unexplained sequence effects induced by mismatches; above it, the length of the continuous mismatch dominates. Testing this hypothesis will require systematic studies of different double, triple, and multiple tandem mismatches across the three sequence contexts and in the linear duplexes.

Finally, to corroborate our findings and truly determine their relevance to mismatch repair efficiency, an assay must be developed that explicitly studies mismatch repair in the presence of DNA supercoiling.

5

Discussion and Conclusions

This thesis has used all-atom implicit solvent molecular dynamics simulations to explore how DNA sequence and damage influence the formation and positioning of plectonemes under torsional strain. We have demonstrated that implicit solvent simulations successfully reproduce known sequence-dependent behaviours of supercoiled DNA and can distinguish between subtle magnitudes of plectoneme pinning propensity in line with experimental results. To experimentally validate our findings, we developed a novel smFRET assay and demonstrated that smFRET and molecular dynamics in tandem provide new insights into bubble formation, positioning, and the conformational dynamics of supercoiled DNA minicircles. We further examined the various properties of bubbles under both directions of supercoiling and across a range of sequences to elucidate a novel role of bubble size in influencing plectoneme positioning, which further depended on whether the main driver of that pinning was the denaturation itself, such as was the case with the Strong339 construct. Here, we have restricted our analysis specifically to bubble formation and their related properties. It should be noted that, in agreement with previous literature [40, 299, 330], type I and II kinks—categorised by sharp bends over two intact base pairs, or over three base pairs with only the hydrogen bonding between the central base pair broken, respectively—are still present.

We next examined the effect of single mismatches on supercoiled DNA and found that plectoneme pinning strength depends on mismatch identity, the direction of supercoiling, and the surrounding sequence context. Surprisingly, this behaviour could not be fully explained by traditional

predictors such as intrinsic curvature, flexibility, or bubble formation. This was the first study to date that evaluated and compared the plectoneme pinning response of the different mismatches, and the first to evaluate the effect of the surrounding sequence. While undamaged DNA exhibited a sequence-dependent response to salt, mismatched DNA exhibited a consistent, sequence-independent increase in local plectoneme propensity under high-salt conditions, indicating an as-of-yet unknown underlying mechanism. We proposed a novel hypothesis in that a threshold mismatch length—which itself could be dependent on the surrounding sequence—may determine whether pinning is driven by mismatch-specific effects or by the bubble formation inherently induced by the mismatch itself. Overall, this work provides a comprehensive simulation-based framework for understanding the influences of sequence and damage effects on supercoiled DNA and provides novel insights into the interplay between intrinsic sequence properties and bubble formation in driving plectoneme positioning. Furthermore, it highlights the importance of considering the role of DNA supercoiling in future studies of mismatch recognition and repair.

5.1 Discussion

The recurring theme throughout this thesis was that, based on current literature, we still do not know, with certainty, what is driving plectoneme positioning under negative supercoiling. Whilst the plectoneme positioning effect of denaturation bubbles has been reported both in this work and previously in the literature [106, 118], the cooperative action of plectoneme and bubble positioning is under-researched. In fact, the only study to date to explicitly consider bubble-plectoneme coupling evaluates only one sequence and investigates the role of bubble nature, size and stability on plectoneme mobility, rather than plectoneme positioning [118]. However, the accuracy of a SIDD prediction in a DNA system capable of writhe has now been confirmed, previously, both experimentally [118] and through simulations [106]. These results suggested that the bubbles formed act as nucleation points for plectonemes. Our simulations are in agreement, as here, the plectoneme positioning landscape has shifted to high bubble density areas, as predicted by the SIDD model, compared to the curvature-driven landscapes under positive supercoiling. However, on the kilo-base length scale applicable to the combined optical and magnetic tweezer experiments, bubbles and plectonemes can co-exist [118]. As plectonemes and bubbles may form separately, further work must be done to understand which bubbles are driving plectoneme positioning and which are not, if there is such a distinction. As this is not possible in our minicircle model, we are restricted to solely studying the plectoneme positioning effect of bubbles as we neglect their ability to be positioned both within and separate from the plectoneme. However, this comes with the trade-off of atomic-level resolution. Furthermore, currently, it is only possible to investigate the coupling atomistically by using minicircles as a model. On the DNA length scales currently accessible to even implicit solvent molecular dynamics, the plectoneme is constrained to a position in the centre of the DNA in a linear construct [106]. Therefore, despite hypothesising the role of bubbles in nucleating and positioning plectonemes, this could not be confirmed previously. However, the structural

map of representative conformations accessible to linear DNA under torsional strain [106] is in agreement with the conformations seen in our minicircles, including the possibility of multiple, not just one, sharply bent apices, especially at high superhelical densities. Thus, minicircles are an appropriate model for a continued investigation in this area.

Whilst the role of intrinsic curvature in positive supercoiling has been determined through the Dekker work [41], their algorithm is only able to predict where the plectoneme will localise and fails to distinguish between relative differences in pinning strength. This was the case for the two Curve75 constructs, where it correctly predicted the position of the plectoneme in these regions, but predicted Curve75-1 to have a higher plectoneme density than Curve75-2. Both in their experimental validation using ISD and our simulations, this was not the case. One possibility is that these discrepancies may arise due to the incorrect omission of bubble formation, even under positive supercoiling. Our results support this hypothesis due to the stronger bubble localisation Curve75-2 has in this region in our simulations under positive supercoiling. This suggests that bubble formation can modulate pinning behaviour, even in regimes where bubbles are traditionally considered less relevant.

However, to date, previous literature has solely focused on where plectonemes form. This thesis is the first to consider the relative pinning strength of a sequence, rather than plectoneme positioning alone. This distinction is crucial when interpreting subtle conformational changes, such as in mismatch-induced pinning, where the P values associated with the pinning strength were relatively low. Therefore, understanding the interplay of intrinsic curvature and bubble formation on plectoneme positioning is crucial to understanding these results, especially in the negatively supercoiled regime. In the absence of a predictive model capable of combining both influences, the implicit solvent simulations presented here bridge that gap.

The primary aim of this thesis was to investigate plectoneme pinning by mismatches. Previous literature had studied this in the form of coarse-grained bubbles [288, 290, 470] and kinks [104]; however, the pinning of a specific single mismatch has only been investigated in the case of the AA mismatch [287]. This was done solely under positive supercoiling and in one sequence context; however, the same experiment also investigated the AC/AC double mismatch. Thus, we present the first investigation to combine both the sequence-dependence of the surrounding DNA and the mismatch itself. The field has previously leaned towards the belief that single mismatches have a stochastic probability of pinning the plectoneme that can be increased with the length of the mismatch [111, 288], or by high salt or high force conditions [287, 288]. We have shown that this is not the case, and the sequence must be considered, both in terms of the sequence context surrounding the mismatch and that of the lesion itself. In fact, a single mismatch can pin the plectoneme rather deterministically, as is the case for mGG. This was previously predicted to be possible by statistical mechanics calculations [287] but was as-of-yet unproven. The fact that mismatches do not have an inherent pinning propensity typical of that mismatch type, but actually that it depends on the surrounding sequence context, the strand

direction, and the direction of supercoiling is something that has not been considered in the literature before, likely due to never before considering that the different mismatch types could exert different behaviours. Finally, the correlation between the pinning propensity of mismatches such as CC, TT, and GG and their repair efficiencies in the literature, whilst they may be serendipitous, well-places DNA supercoiling as an important consideration in future studies of MMR efficiency. Even in the absence of a conclusive link between the pinning propensities and mismatch repair rates, our results confidently prove that it is no longer sufficient to treat all mismatches as equal or be modelled as bubbles in future investigations of plectoneme pinning.

A bubble role hypothesis

Combining the results from both chapters offers new insights into the conformational dynamics of supercoiled DNA. Most notably, it highlighted that bubble formation alone does not drive plectoneme positioning, surprisingly, even under negative supercoiling. This was seen clearly in the plectoneme positioning landscapes of the mismatches, but also in the undamaged constructs where bubbles formed both at and away from the plectoneme tips. Under positive supercoiling, our results agree with the Dekker predictions, showing that intrinsic curvature is driving the positioning. Here, bubble formation at the plectoneme tip may have aided pinning in some cases, but cases of relatively strong formation of bubbles elsewhere in the minicircle as well as off-centre bubble formation at the tip showed the bubbles not to be the driving force behind the pinning. It has previously been reported that the bubbles formed during positive supercoiling remain closer to canonical structure compared to the "catastrophic" structural disruption seen in the bubbles formed during negative supercoiling [118]. Our simulations agreed with this, which reinforces the idea that bubble formation is secondary to intrinsic curvature in driving plectoneme positioning under positive supercoiling. We have seen this also in our simulations under negative supercoiling, where bubbles formed at the plectoneme tip were shorter, with significantly more structural disruption compared to the longer, locally unwound segments of DNA formed elsewhere in the minicircle (fig. 3.12).

The bubbles considered to be a "significant structural disruption" would present a larger energy barrier to plectoneme repositioning due to the greater energetic cost of re-nucleating at a second site. Thus, it may be the case that these bubbles exert the plectoneme positioning effect, and the less-disturbed bubbles are merely artefacts of torsional strain, born due to high stress or curvature. The explicit solvent simulations of Strong339 support this idea, where there was reduced bubble formation, overall, compared to implicit solvent. The most significant bubbles that remained in explicit solvent were at the plectoneme tip, in the region strongly predicted by the SIDD algorithm, whilst the large peak around the $\sim 300^{\text{th}}$ nucleotide was lost. This suggests that the plectoneme positioning bubbles remain, whilst the more stable nature of the explicit solvent causes fewer bubbles born out of torsional strain and instability to form.

Should this be the case, under negative supercoiling, the barrier to plectoneme repositioning

would increase with the size of the bubble, explaining why stronger pinning mismatches and sequences had statistically larger bubble sizes. This would further explain the lack of trend between mismatch twist absorption and plectoneme pinning, as, to confirm this hypothesis, we must consider the nature of the bubble rather than solely its formation, size or lifetime. The hypothesis holds under positive supercoiling as, in this regime, only the lesser-disrupted bubbles form, consistent with the lack of bubble-driven plectoneme positioning in this regime. However, this would then not explain the unexpected higher plectoneme pinning strength of Curve75-2, compared to Curve75-1, which was not predicted by the Dekker algorithm. In this case, it may be simply that the algorithm does not find the correct balance of the relative contributions of flexibility and curvature; or, potentially, other factors may be at play. There is no indication that the Dekker algorithm actually aimed to accurately distinguish between varying degrees of plectoneme density, as, to our knowledge, this thesis is the first to consider this idea. However, the existence of plectoneme-positioning and non-plectoneme-positioning bubbles does offer an explanation as to the increase in pinning propensity of the 126-mer under increased salt concentrations. Here, as the writhe is forced at higher ionic strength, the conformational perturbation of the bubbles would become more severe, thus increasing the barrier to plectoneme repositioning. Finally, this may offer an explanation as to the increase in local plectoneme propensity of the mismatch, regardless of sequence context, if the increased ionic strength is able to stabilise a more disrupted bubble conformation at the mismatch. It may be the case that all mismatches form bubbles, but the nature of the bubble formed is dependent on the lesion and the surrounding sequence. This is a reasonable assumption as the degree of disruption will depend largely on the nearest neighbours, as well as potential stabilising or destabilising factors within the immediate local sequence context. Should the nature of the bubbles change depending on salt conditions, this also may be one such mechanism that the DNA employs to buffer against large conformational changes under more severe conditions in the positive supercoiling regime. Visual inspection of the most populated clusters under positive supercoiling at high temperatures suggests this to be possible. Overall, more work would need to be done to understand this further. First, a consistent, definitive criterion capable of distinguishing between the two possible degrees of disruption would be needed. Then, it would be interesting to plot the different bubble types separately on our polar plots to visualise their sequence dependence and influence on plectoneme positioning. Overall, this work has proved the usefulness, validity, and accessibility of implicit solvent simulations as a model capable of further investigating this.

Entropy

For a complete discussion of plectoneme pinning and the conformational dynamics of supercoiled DNA, one cannot neglect a consideration of entropy. Pinning a plectoneme in our simulations refers to the preferential positioning of a certain region of the DNA across multiple simulation replicas. This is largely due to the timescale of our simulations not allowing for significant repositioning of the plectoneme tip. On the kilobase length scale, and at lengths relevant to

genomic DNA, pinning instead refers specifically to the barrier to plectoneme diffusion and, similarly, the plectoneme being consistently positioned in one region of the DNA over time. In either case, the pinning of a plectoneme represents a balance between the entropic cost of restricting plectoneme mobility and the energetic gain from occupying a favourable structural context, such as regions of high intrinsic curvature or local base-pair disruption. The effect of entropy becomes more complex at longer DNA lengths, where multiple plectonemes may coexist. In these cases, the positional entropic contribution—that is, the contribution from the variety of positions within the DNA that the plectoneme can be positioned at—is combined with length-exchange entropy: the partitioning of the total writhe, or plectoneme length, between the multiple plectonemes [499]. Our minicircle simulations and other comparably short DNA models are only able to capture positional entropy. Thus, the barrier to plectoneme diffusion by a lesser-disrupted bubble is not only energetically lower but also does not afford the same decrease in entropy as a bubble capable of driving plectoneme pinning. Similarly, the bubble formed by a single-base mismatch may lower the energy locally but may not be sufficient to overcome the entropic cost of pinning, which may explain the low pinning propensities we observe. Therefore, entropy may not simply be a background factor to overcome, but an active determinant of whether pinning occurs at all. The absence of length-exchange entropy in our simulations likely underestimates the true complexity of this balance and may force our minicircles to favour lower-energy conformations than is truly required. As a result, bubble formation may be over-represented in our simulations compared to longer, more entropically diverse DNA models.

5.2 Reflection

It was not our original intention to investigate the drivers of plectoneme positioning, nor was it to investigate both directions of supercoiling. In fact, the true aim of this project changed numerous times throughout my PhD. As such, in its final form, there are changes to this work that I would have made had I known the importance that the sequence would have. Firstly, the original minicircle sequence and size were chosen as it was able to be made experimentally. Unfortunately, this turned out not to be the case. Therefore, the size of 339 bp is not optimal due to not having an integer number of turns - 336 base pairs would have been more appropriate. Furthermore, none of the sequences were designed with consideration of the circular nature of the DNA. In reality, the double pinning effect did not come to light until the mismatches were investigated. Thus, better-designed sequences would have considered the favourability of the inherent pinning of the region opposite the desired plectoneme. We also, primarily, only intended to investigate DNA under negative supercoiling. Thus, the positioning of the mismatches in regions of low plectoneme propensity across both regimes was serendipitous rather than good planning. However, this also reflects the similar underlying features that drive plectoneme positioning under both directions of supercoiling.

Simulations were run where a structure was taken from simulations with the mismatch at

the plectoneme tip, and a different mismatch was added to the crossing point. The second mismatch had a higher pinning propensity than the one at the tip in Control339 in order to see if it could drive plectoneme repositioning. This would have allowed us to further study the competition element for the plectoneme tip and, potentially, the varying hypothesised role of bubbles dependent on their degree of disruption. However, no simulations were done of the mismatches at this second point in the Control339 sequence, and so, due to the effect that sequence had, there is no way of actually knowing the pinning propensity of the mismatch at that crossing point location. Therefore, I made the decision that we did not have the appropriate controls to include those experiments in this thesis. The use of a high-performance computer (HPC) was invaluable, as it enabled us to run a large number of simulations in parallel. However, given the time required to analyse the extensive dataset generated, this also meant that there were occasions when decisions about subsequent simulations had to be made before the results of prior ones were available. Nevertheless, the findings presented here point to several promising directions for further investigation into DNA supercoiling and mismatch repair.

5.3 Future Work

The primary aim of this project was to investigate plectoneme pinning by mismatches using atomistic simulations. We hypothesised that different mismatches would exhibit varying propensities for localising to the plectoneme tip, and that this behaviour may be further influenced by the surrounding sequence. Our findings support this hypothesis and suggest that the process is influenced by subtle, sequence-specific factors that, as of yet, are not fully captured by the existing predictive models.

Short-term

A logical and immediate next step is to systematically investigate the effects of different double and triple mismatches, in the same manner as has been done here for single mismatches. As both hardware and software continue to improve, it will become feasible to use our simulation framework to systematically investigate all combinations of even the triple mismatches. This will provide deeper insight into the factors driving the competition for plectoneme positioning. We should also explore the longer tandem mismatches in various sequence contexts to test our hypothesis regarding the threshold mismatch length below which sequence effects dominate the competition offered by the mismatch for plectoneme pinning. The control simulations for the excluded experiments should also be done by simulating the mismatches at the crossing point location in Control339. In parallel, the smFRET assay should be optimised and extended using the 339 bp minicircle constructs to enable direct experimental validation of the pinning behaviours observed in our simulations.

It was not our aim to fully understand the drivers of plectoneme positioning; our interest lies in uncovering a potential biological role for mismatch-driven plectoneme pinning. However,

such a large dataset has allowed us to investigate this further. With more time, having such a large dataset may allow us to further probe the bubble-plectoneme relationship. For example, investigating the relative timepoints of bubble and plectoneme formation within our simulations may uncover differences in which bubbles are driving plectoneme positioning, versus which are artefacts of torsional strain. This may help to explain the unexpected bubble formation in mismatches that did not pin the plectoneme, or the positioning of the plectoneme at mismatches with little bubble formation. Furthermore, categorising bubbles into degrees of disruption may also help to understand this and further take advantage of the atomistic resolution offered by molecular dynamics. Doing so will offer insight into the possibility of plectoneme positioning and non-plectoneme positioning bubble formation as hypothesised above. Finally, repeating the simulations described in this thesis under both positive and negative supercoiling—particularly focusing on the negative supercoiling regime—whilst maintaining the restraints that preserve base pair hydrogen bonding and suppress bubble formation throughout the simulation, may offer further insight into the role of bubble formation in driving plectoneme pinning. Additionally, selectively permitting base pair denaturation at specific sites, while maintaining restraints elsewhere, could provide a more nuanced understanding of how local destabilisation contributes to the overall pinning mechanism.

Medium-term

In the medium term, the field should work towards developing a predictive model of plectoneme pinning that can distinguish between the two directions of supercoiling. While existing models, such as Dekker's, perform well in many respects, our results demonstrate that they cannot reliably capture subtle sequence or structural variations, as exemplified by the Curve75 constructs. The mismatched constructs studied here have relatively small P values, suggesting that we are investigating very subtle effects. Therefore, accurate and, especially, relative quantitative prediction becomes critical. At present, implicit solvent simulations remain the only method capable of predicting plectoneme pinning with this level of accuracy in both directions of supercoiling. We should also begin exploring other forms of DNA damage, such as abasic sites or oxidative lesions, to assess whether they, too, pin the plectoneme. Comparing the conformational dynamics of different lesion types under torsional strain may help to uncover the broader mechanisms driving plectoneme positioning.

Long-term

In the longer term, to fully understand the biological relevance of these findings, it will be necessary to determine whether the differences in pinning propensity influence DNA damage detection and repair. This will require a truly interdisciplinary and multiscale approach. One possibility is to extend our smFRET assay to study MutS binding kinetics in both supercoiled and relaxed topoisomers. For example, by labelling MutS with one dye and placing the complementary dye near the mismatch in a minicircle, we could directly investigate mismatch

recognition under torsional stress. It would also be interesting to extend the MMR-seq assay for use with supercoiled DNA to evaluate the sequence dependence of repair efficiency on a larger length scale. Together, these experiments would unambiguously provide direct insight into the role of DNA supercoiling in mismatch recognition. Furthermore, disentangling the contributions of damage accumulation and mismatch repair efficiency across the genome would allow the correlation of repair rates against known plectonemic hotspots.

In conclusion, we have demonstrated that our simulation framework is an accurate and accessible tool for investigating plectoneme pinning. Taken together, these future directions outline an exciting research framework spanning the atomistic to genome scale, with the potential to uncover how supercoiling and local DNA structure influence damage recognition and repair.

References

Back-references to the pages where the publication was cited are given by [•].

[1] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". *Nature*, 1953.
DOI: 10.1038/171737a0 [1]

[2] R. E. Franklin and R. G. Gosling. "The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content". *Acta Crystallographica*, 1953.
DOI: 10.1107/S0365110X53001939 [1]

[3] Allison Piovesan, Maria Chiara Pelleri, Francesca Antonaros, Pierluigi Strippoli, Maria Caracausi, and Lorenza Vitale. "On the Length, Weight and GC Content of the Human Genome". *BMC Research Notes*, 2019.
DOI: 10.1186/s13104-019-4137-z [2 , 18]

[4] The Free Dictionary. *Deoxyribonucleic Acid*.
https://en.thefreedictionary.com/deoxyribonucleic+acid.
(Visited on 07/22/2021) [2]

[5] M. Levitt. "How Many Base-Pairs per Turn Does DNA Have in Solution and in Chromatin? Some Theoretical Calculations." *Proceedings of the National Academy of Sciences*, 1978.
DOI: 10.1073/pnas.75.2.640 [2]

[6] J. Mazur and R.L. Jernigan. "Comparison of Rotation Models for Describing DNA Conformations: Application to Static and Polymorphic Forms". *Biophysical Journal*, 1995.
DOI: 10.1016/S0006-3495(95)80320-6 [2]

[7] J. Griffith. "DNA Structure: Evidence from Electron Microscopy". *Science*, 1978.
DOI: 10.1126/science.663672 [2]

[8] Stephen Neidle (editor). *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, 1999.
DOI: 10.1093/oso/9780198500384.001.0001 [3]

[9] Xiang-Jun Lu and Wilma K. Olson. "3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-dimensional Nucleic Acid Structures". *Nucleic Acids Research*, 2003.
DOI: 10.1093/nar/gkg680 [3 , 112]

[10] R. E. Dickerson. “Definitions and Nomenclature of Nucleic Acid Structure Components.” *Nucleic Acids Research*, 1989.
DOI: 10.1093/nar/17.5.1797 [4, 5]

[11] Stephan Diekmann. “Definitions and Nomenclature of Nucleic Acid Structure Parameters”. *Journal of Molecular Biology*, 1989.
DOI: 10.1016/0022-2836(89)90324-0 [4]

[12] Wilma K Olson, Manju Bansal, Stephen K Burley, Richard E Dickerson, Mark Gerstein, Stephen C Harvey, Udo Heinemann, Xiang-Jun Lu, Stephen Neidle, Zippora Shakked, Heinz Sklenar, Masashi Suzuki, Chang-Shung Tung, Eric Westhof, Cynthia Wolberger, and Helen M Berman. “A Standard Reference Frame for the Description of Nucleic Acid Base-Pair Geometry”. *Journal of Molecular Biology*, 2001.
DOI: 10.1006/jmbi.2001.4987 [4, 6, 113]

[13] Xiang-Jun Lu, M.A. El Hassan, and C.A. Hunter. “Structure and Conformation of Helical Nucleic Acids: Rebuilding Program (SCHNArP)”. *Journal of Molecular Biology*, 1997.
DOI: 10.1006/jmbi.1997.1345 [4]

[14] Richard E. Dickerson and Horace R. Drew. “Structure of a B-DNA Dodecamer: II. Influence of Base Sequence on Helix Structure”. *Journal of Molecular Biology*, 1981.
DOI: 10.1016/0022-2836(81)90357-0 [4]

[15] Andrey A. Gorin, Victor B. Zhurkin, and Wilma K. “B-DNA Twisting Correlates with Base-pair Morphology”. *Journal of Molecular Biology*, 1995.
DOI: 10.1006/jmbi.1994.0120 [4, 5]

[16] W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. “DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes.” *Proceedings of the National Academy of Sciences of the United States of America*, 1998.
DOI: 10.1073/pnas.95.19.11163 [5, 19]

[17] V. B. Zhurkin, Yu. P. Lysov, and V. I. Ivanov. “Different Families of Double-Stranded Conformations of DNA as Revealed by Computer Calculations”. *Biopolymers*, 1978.
DOI: 10.1002/bip.1978.360170209 [6]

[18] Marla S. Babcock and Wilma K. Olson. “The Effect of Mathematics and Coordinate System on Comparability and "Dependencies" of Nucleic Acid Structure Parameters”. *Journal of Molecular Biology*, 1994.
DOI: 10.1006/jmbi.1994.1212 [6]

[19] D. Bhattacharyya and M. Bansal. “Local Variability and Base Sequence Effects in DNA Crystal Structures.” *Journal of Biomolecular Structure & Dynamics*, 1990.
DOI: 10.1080/07391102.1990.10507828 [6]

[20] Kim López-Güell, Federica Battistini, and Modesto Orozco. “Correlated Motions in DNA: Beyond Base-Pair Step Models of DNA Flexibility”. *Nucleic Acids Research*, 2023.
DOI: 10.1093/nar/gkad136 [6]

[21] Korbinian Liebl and Martin Zacharias. “Accurate Modeling of DNA Conformational

Flexibility by a Multivariate Ising Model". *Proceedings of the National Academy of Sciences*, 2021.
DOI: 10.1073/pnas.2021263118 [6]

[22] F. H. Crick. "On Protein Synthesis." *Symposia of the Society for Experimental Biology*, 1958.
[6, 7]

[23] Matthew Cobb. "60 Years Ago, Francis Crick Changed the Logic of Biology." *PLOS Biology*, 2017.
DOI: 10.1371/journal.pbio.2003243 [6]

[24] Francis Crick. *"Ideas on Protein Synthesis (Oct. 1956)"*. https://wellcomecollection.org/works/xmscu3g4. 1956.
(Visited on 01/31/2025) [7]

[25] Paulo Amaral et al. "The Status of the Human Gene Catalogue". *Nature*, 2023.
DOI: 10.1038/s41586-023-06490-x [6]

[26] Klas Hatje, Stefanie Mühlhausen, Dominic Simm, and Martin Kollmar. "The Protein-Coding Human Genome: Annotating High-Hanging Fruits". *BioEssays*, 2019.
DOI: 10.1002/bies.201900066 [6]

[27] National Cancer Institute. *Transcription - NCI Dictionary of Genetics Terms*. https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/transcription. 2012.
(Visited on 01/31/2025) [8]

[28] Khan Academy. *DNA Replication*. https://www.khanacademy.org/science/hs-bio/x230b3ff252126bb6:gene-expression-and-regulation/x230b3ff252126bb6:dna-structure-and-replication/a/dna-replication.
(Visited on 01/31/2025) [9]

[29] S. S. Patel and K. M. Picha. "Structure and Function of Hexameric Helicases." *Annual Review of Biochemistry*, 2000.
DOI: 10.1146/annurev.biochem.69.1.651 [7]

[30] Martin R. Singleton, Mark S. Dillingham, and Dale B. Wigley. "Structure and Mechanism of Helicases and Nucleic Acid Translocases." *Annual Review of Biochemistry*, 2007.
DOI: 10.1146/annurev.biochem.76.052305.115300 [7]

[31] Timothy M. Lohman, Eric J. Tomko, and Colin G. Wu. "Non-Hexameric DNA Helicases and Translocases: Mechanisms and Regulation." *Nature Reviews Molecular Cell Biology*, 2008.
DOI: 10.1038/nrm2394 [7]

[32] Anna Marie Pyle. "Translocation and Unwinding Mechanisms of RNA and DNA Helicases." *Annual Review of Biophysics*, 2008.
DOI: 10.1146/annurev.biophys.37.032807.125908 [7]

[33] Andrew H. Laszlo, Jonathan M. Craig, Momčilo Gavrilov, Ramreddy Tippana, Ian C. Nova, Jesse R. Huang, Hwanhee C. Kim, Sarah J. Abell,

Mallory deCampos-Stairiker, Jonathan W. Mount, Jasmine L. Bowman, Katherine S. Baker, Hugh Higinbotham, Dmitriy Bobrovnikov, Taekjip Ha, and Jens H. Gundlach. "Sequence-Dependent Mechanochemical Coupling of Helicase Translocation and Unwinding at Single-Nucleotide Resolution". *Proceedings of the National Academy of Sciences*, 2022.
DOI: 10.1073/pnas.2202489119 [8]

[34] Richard Wheeler. *Circular DNA Supercoiling*. https://commons.wikimedia.org/w/index.php?title=File:Circular_DNA_Supercoiling.png&oldid=8 2007.
(Visited on 06/26/2021) [9, 10]

[35] David Argudo and Prashant K. Purohit. "Competition between Supercoils and Toroids in Single Molecule DNA Condensation". *Biophysical Journal*, 2012.
DOI: 10.1016/j.bpj.2012.05.033 [9]

[36] Philipp U. Walker, Willem Vanderlinden, and Jan Lipfert. "Dynamics and Energy Landscape of DNA Plectoneme Nucleation". *Physical Review E*, 2018.
DOI: 10.1103/physreve.98.042412 [9, 16]

[37] N. Patrick Higgins and Alexander Vologodskii. "Topological Behavior of Plasmid DNA". *Microbiology Spectrum*, 2015.
DOI: 10.1128/microbiolspec.plas-0036-2014 [9]

[38] Giuseppe Perugino, Anna Valenti, Anna D'Amaro, Mosè Rossi, and Maria Ciaramella. "Reverse Gyrase and Genome Stability in Hyperthermophilic Organisms". *Biochemical Society Transactions*, 2009.
DOI: 10.1042/BST0370069 [10, 89]

[39] P. Forterre, A. Bergerat, and P. Lopez-Garcia. "The Unique DNA Topology and DNA Topoisomerases of Hyperthermophilic Archaea." *FEMS Microbiology Reviews*, 1996.
DOI: 10.1111/j.1574-6976.1996.tb00240.x [10, 89]

[40] Alice L. B. Pyne, Agnes Noy, Kavit H. S. Main, Victor Velasco-Berrelleza, Michael M. Piperakis, Lesley A. Mitchenall, Fiorella M. Cugliandolo, Joseph G. Beton, Clare E. M. Stevenson, Bart W. Hoogenboom, Andrew D. Bates, Anthony Maxwell, and Sarah A. Harris. "Base-Pair Resolution Analysis of the Effect of Supercoiling on DNA Flexibility and Major Groove Recognition by Triplex-Forming Oligonucleotides". *Nature Communications*, 2021.
DOI: 10.1038/s41467-021-21243-y [10, 14, 29, 30, 60, 61, 65, 69, 81, 105, 148, 150]

[41] Sung Hyun Kim, Mahipal Ganji, Eugene Kim, Jaco van der Torre, Elio A. Abbondanzieri, and Cees Dekker. "DNA Sequence Encodes the Position of DNA Supercoils". *eLife*, 2018.
DOI: 10.7554/elife.36557 [11, 14, 15, 16, 17, 26, 33, 54, 56, 61, 66, 69, 72, 73, 99, 118, 120, 152, 206]

[42] George D. Watson, Elliot W. Chan, Mark C. Leake, and Agnes Noy. "Structural Interplay between DNA-shape Protein Recognition and Supercoiling: The Case of IHF." *Computational and Stuctural Biotchnology Journal*, 2022.
DOI: 10.1016/j.csbj.2022.09.020 [11, 26]

[43] Shmoop Editorial Team. *Biology DNA Packaging - Shmoop Biology.* https://www.shmoop.com/study-guides/dna/dna-packaging.html. 2008. (Visited on 02/05/2025) [12]

[44] Kazuhiro Maeshima, Satoru Ide, and Michael Babokhov. "Dynamic Chromatin Organization without the 30-Nm Fiber". *Cell Nucleus*, 2019. DOI: 10.1016/j.ceb.2019.02.003 [11]

[45] Nancy J. Crisona, Robert L. Weinberg, Brian J. Peter, De Witt Sumners, and Nicholas R. Cozzarelli. "The Topological Mechanism of Phage λ Integrase". *Journal of Molecular Biology*, 1999. DOI: 10.1006/jmbi.1999.2771 [11, 12]

[46] Rifka Vlijm, Sung Hyun Kim, Paul L. De Zwart, Yamini Dalal, and Cees Dekker. "The Supercoiling State of DNA Determines the Handedness of Both H3 and CENP-A Nucleosomes". *Nanoscale*, 2017. DOI: 10.1039/C6NR06245H [11]

[47] International Human Genome Sequencing Consortium. "Finishing the Euchromatic Sequence of the Human Genome". *Nature*, 2004. DOI: 10.1038/nature03001 [12]

[48] Jie Ma and Michelle D. Wang. "DNA Supercoiling during Transcription". *Biophysical Reviews*, 2016. DOI: 10.1007/s12551-016-0215-9 [13]

[49] Leroy F. Liu and James C. Wang. "Supercoiling of the DNA Template during Transcription." *Proceedings of the National Academy of Sciences of the United States of America*, 1987. DOI: 10.1073/pnas.84.20.7024 [12]

[50] Andrew Travers and Georgi Muskhelishvili. "DNA Supercoiling — a Global Transcriptional Regulator for Enterobacterial Growth?" *Nature Reviews Microbiology*, 2005. DOI: 10.1038/nrmicro1088 [12, 14]

[51] B ten Heggeler-Bordier, W Wahli, M Adrian, A Stasiak, and J Dubochet. "The Apical Localization of Transcribing RNA Polymerases on Supercoiled DNA Prevents Their Rotation around the Template." *The EMBO Journal*, 1992. DOI: 10.1002/j.1460-2075 [12]

[52] Fedor Kouzine, Juhong Liu, Suzanne Sanford, Hye-Jung Chung, and David Levens. "The Dynamic Response of Upstream DNA to Transcription-Generated Torsional Stress". *Nature Structural & Molecular Biology*, 2004. DOI: 10.1038/nsmb848 [13]

[53] Fedor Kouzine, Suzanne Sanford, Zichrini Elisha-Feil, and David Levens. "The Functional Response of Upstream DNA to Dynamic Supercoiling in Vivo". *Nature Structural & Molecular Biology*, 2008. DOI: 10.1038/nsmb.1372 [13]

[54] Fenfei Leng, Luciana Amado, and Roger McMacken. "Coupling DNA Supercoiling to Transcription in Defined Protein Systems". *Journal of Biological Chemistry*, 2004.

DOI: 10.1074/jbc.M403798200 [13]

[55] Elena A. Oussatcheva, Jeffrey Pavlicek, Otto F. Sankey, Richard R. Sinden, Yuri L. Lyubchenko, and Vladimir N. Potaman. "Influence of Global DNA Topology on Cruciform Formation in Supercoiled DNA". *Journal of Molecular Biology*, 2004.
DOI: 10.1016/j.jmb.2004.02.075 [13]

[56] Kiyoshi Mizuuchi, Michiyo Mizuuchi, and Martin Gellert. "Cruciform Structures in Palindromic DNA Are Favored by DNA Supercoiling". *Journal of Molecular Biology*, 1982.
DOI: 10.1016/0022-2836(82)90325-4 [13]

[57] Alan Herbert and Alexander Rich. "The Biology of Left-handed Z-DNA". *Journal of Biological Chemistry*, 1996.
DOI: 10.1074/jbc.271.20.11595 [13]

[58] Florian C. Oberstrass, Louis E. Fernandes, Paul Lebel, and Zev Bryant. "Torque Spectroscopy of DNA: Base-Pair Stability, Boundary Effects, Backbending, and Breathing Dynamics". *Physical Review Letters*, 2013.
DOI: 10.1103/PhysRevLett.110.178103 [13]

[59] Christopher Deufel, Scott Forth, Chad R Simmons, Siavash Dejgosha, and Michelle D Wang. "Nanofabricated Quartz Cylinders for Angular Trapping: DNA Supercoiling Torque Detection". *Nature Methods*, 2007.
DOI: 10.1038/nmeth1013 [13]

[60] Zev Bryant, Michael D. Stone, Jeff Gore, Steven B. Smith, Nicholas R. Cozzarelli, and Carlos Bustamante. "Structural Transitions and Elasticity from Torque Measurements on DNA". *Nature*, 2003.
DOI: 10.1038/nature01810 [13]

[61] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. "Stretched and Overwound DNA Forms a Pauling-like Structure with Exposed Bases". *Proceedings of the National Academy of Sciences*, 1998.
DOI: 10.1073/pnas.95.24.14152 [13]

[62] Alicia D. D'Souza, Boris P. Belotserkovskii, and Philip C. Hanawalt. "A Novel Mode for Transcription Inhibition Mediated by PNA-induced R-loops with a Model in Vitro System." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 2018.
DOI: 10.1016/j.bbagrm.2017.12.008 [13]

[63] Maikel Castellano-Pozo, José M. Santos-Pereira, Ana G. Rondón, Sonia Barroso, Eloisa Andújar, Mónica Pérez-Alegre, Tatiana García-Muse, and Andrés Aguilera. "R Loops Are Linked to Histone H3 S10 Phosphorylation and Chromatin Condensation". *Molecular Cell*, 2013.
DOI: 10.1016/j.molcel.2013.10.006 [13]

[64] Paul A. Ginno, Paul L. Lott, Holly C. Christensen, Ian Korf, and Frédéric Chédin. "R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters." *Molecular cell*, 2012.
DOI: 10.1016/j.molcel.2012.01.017 [13]

[65] Boris P. Belotserkovskii, Silvia Tornaletti, Alicia D. D'Souza, and Philip C. Hanawalt.

"R-Loop Generation during Transcription: Formation, Processing and Cellular Outcomes." *DNA Repair*, 2018.
DOI: 10.1016/j.dnarep.2018.08.009 [13]

[66] Julie Sollier and Karlene A. Cimprich. "Breaking Bad: R-loops and Genome Integrity." *Trends in Cell Biology*, 2015.
DOI: 10.1016/j.tcb.2015.05.003 [13]

[67] Lorenzo Costantino and Douglas Koshland. "The Yin and Yang of R-loop Biology." *Current Opinion in Cell Biology*, 2015.
DOI: 10.1016/j.ceb.2015.04.008 [13]

[68] Robert Stolz, Shaheen Sulthana, Stella R. Hartono, Maika Malig, Craig J. Benham, and Frédéric Chédin. "Interplay between DNA Sequence and Negative Superhelicity Drives R-loop Structures." *Proceedings of the National Academy of Sciences of the United States of America*, 2019.
DOI: 10.1073/pnas.1819476116 [13]

[69] H. Tabuchi, H. Handa, and S. Hirose. "Underwinding of DNA on Binding of Yeast TFIID to the TATA Element". *Biochemical and Biophysical Research Communications*, 1993.
DOI: 10.1006/bbrc.1993.1576 [13, 27]

[70] Driss El Hanafi and Lionello Bossi. "Activation and Silencing of Leu-500 Promoter by Transcription-Induced DNA Supercoiling in the Salmonella Chromosome". *Molecular Microbiology*, 2000.
DOI: 10.1046/j.1365-2958.2000.02015.x [13]

[71] Lok-hang So, Anandamohan Ghosh, Chenghang Zong, Leonardo A Sepúlveda, Ronen Segev, and Ido Golding. "General Properties of Transcriptional Time Series in Escherichia Coli". *Nature Genetics*, 2011.
DOI: 10.1038/ng.821 [13]

[72] Christophe Bécavin, Maria Barbi, Jean-Marc Victor, and Annick Lesne. "Transcription within Condensed Chromatin: Steric Hindrance Facilitates Elongation". *Biophysical Journal*, 2010.
DOI: 10.1016/j.bpj.2009.10.054 [13]

[73] Maxim Y. Sheinin, Ming Li, Mohammad Soltani, Karolin Luger, and Michelle D. Wang. "Torque Modulates Nucleosome Stability and Facilitates H2A/H2B Dimer Loss". *Nature Communications*, 2013.
DOI: 10.1038/ncomms3579 [13]

[74] Heta P. Patel, Stefano Coppola, Wim Pomp, Umberto Aiello, Ineke Brouwer, Domenico Libri, and Tineke L. Lenstra. "DNA Supercoiling Restricts the Transcriptional Bursting of Neighboring Eukaryotic Genes". *Molecular Cell*, 2023.
DOI: 10.1016/j.molcel.2023.04.015 [13]

[75] M R Gartenberg and J C Wang. "Positive Supercoiling of DNA Greatly Diminishes mRNA Synthesis in Yeast." *Proceedings of the National Academy of Sciences*, 1992.
DOI: 10.1073/pnas.89.23.11461 [13]

[76] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. "Real-Time Kinetics of Gene Activity in Individual Bacteria". *Cell*, 2005.
DOI: 10.1016/j.cell.2005.09.031 [13]

[77] Yuichi Taniguchi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie. "Quantifying E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells". *Science*, 2010.
DOI: 10.1126/science.1188308 [13]

[78] Chenghang Zong, Lok-hang So, Leonardo A Sepúlveda, Samuel O Skinner, and Ido Golding. "Lysogen Stability Is Determined by the Frequency of Activity Bursts from the Fate-determining Gene". *Molecular Systems Biology*, 2010.
DOI: 10.1038/msb.2010.96 [13]

[79] Judith Montag et al. "Burst-Like Transcription of Mutant and Wildtype MYH7-Alleles as Possible Origin of Cell-to-Cell Contractile Imbalance in Hypertrophic Cardiomyopathy". *Frontiers in Physiology*, 2018.
DOI: 10.3389/fphys.2018.00359 [13]

[80] David M. Suter, Nacho Molina, David Gatfield, Kim Schneider, Ueli Schibler, and Felix Naef. "Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics". *Science*, 2011.
DOI: 10.1126/science.1198817 [13]

[81] Katie L. Pennington, Sharon K. Marr, Gung-Wei Chirn, and Michael T. Marr. "Holo-TFIID Controls the Magnitude of a Transcription Burst and Fine-Tuning of Transcription". *Proceedings of the National Academy of Sciences - PNAS*, 2013.
DOI: 10.1073/pnas.1221712110 [13]

[82] Shasha Chong, Chongyi Chen, Hao Ge, and X. Sunney Xie. "Mechanism of Transcriptional Bursting in Bacteria". *Cell*, 2014.
DOI: 10.1016/j.cell.2014.05.038 [14]

[83] Yathish Jagadheesh Achar, Mohamood Adhil, Ramveer Choudhary, Nick Gilbert, and Marco Foiani. "Negative Supercoil at Gene Boundaries Modulates Gene Topology". *Nature*, 2020.
DOI: 10.1038/s41586-020-1934-4 [14]

[84] Mukund Thattai and Alexander van Oudenaarden. "Stochastic Gene Expression in Fluctuating Environments". *Genetics*, 2004.
DOI: 10.1534/genetics.167.1.523 [14]

[85] E. Kussell. "Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments". *Science*, 2005.
DOI: 10.1126/science.1114383 [14]

[86] Denise M. Wolf, Vijay V. Vazirani, and Adam P. Arkin. "Diversity in Times of Adversity: Probabilistic Strategies in Microbial Survival Games". *Journal of Theoretical Biology*, 2005.
DOI: 10.1016/j.jtbi.2004.11.020 [14]

[87] Catherine Naughton, Nicolaos Avlonitis, Samuel Corless, James G Prendergast, Ioulia K Mati, Paul P Eijk, Scott L Cockroft, Mark Bradley, Bauke Ylstra, and

Nick Gilbert. “Transcription Forms and Remodels Supercoiling Domains Unfolding Large-Scale Chromatin Structures”. *Nature Structural & Molecular Biology*, 2013.
DOI: 10.1038/nsmb.2509 [14, 29]

[88] Fabrizio Benedetti, Julien Dorier, and Andrzej Stasiak. “Effects of Supercoiling on Enhancer–Promoter Contacts”. *Nucleic Acids Research*, 2014.
DOI: 10.1093/nar/gku759 [14]

[89] Mary-Ann Bjornsti and Scott H. Kaufmann. “Topoisomerases and Cancer Chemotherapy: Recent Advances and Unanswered Questions.” *F1000Research*, 2019.
DOI: 10.12688/f1000research.20201.1 [14]

[90] Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker, and Leonid A Mirny. “Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization”. *Nature Methods*, 2012.
DOI: 10.1038/nmeth.2148 [14]

[91] Jorge B. Schvartzman, Pablo Hernández, Dora B. Krimer, Julien Dorier, and Andrzej Stasiak. “Closing the DNA Replication Cycle: From Simple Circular Molecules to Supercoiled and Knotted DNA Catenanes.” *Nucleic Acids Research*, 2019.
DOI: 10.1093/nar/gkz586 [14]

[92] Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions”. *Nature*, 2012.
DOI: 10.1038/nature11082 [14]

[93] Elphège P. Nora, Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L. van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. “Spatial Partitioning of the Regulatory Landscape of the X-inactivation Centre”. *Nature*, 2012.
DOI: 10.1038/nature11049 [14]

[94] Dusan Racko, Fabrizio Benedetti, Julien Dorier, and Andrzej Stasiak. “Are TADs Supercoiled?” *Nucleic Acids Research*, 2019.
DOI: 10.1093/nar/gky1091 [14, 29]

[95] Nick Gilbert and James Allan. “Supercoiling in DNA and Chromatin”. *Genome Architecture and Expression*, 2014.
DOI: 10.1016/j.gde.2013.10.013 [14, 15]

[96] Fedor Kouzine, Ashutosh Gupta, Laura Baranello, Damian Wojtowicz, Khadija Ben-Aissa, Juhong Liu, Teresa M Przytycka, and David Levens. “Transcription-Dependent Dynamic Supercoiling Is a Short-Range Genomic Force”. *Nature Structural & Molecular Biology*, 2013.
DOI: 10.1038/nsmb.2517 [15, 29]

[97] Thana Sutthibutpong, Christian Matek, Craig J. Benham, Gabriel G. Slade, Agnes Noy, Charles A. Laughton, Jonathan P. K. Doye, Ard A. Louis, and Sarah A. Harris.

"Long-Range Correlations in the Mechanics of Small DNA Circles under Topological Stress Revealed by Multi-Scale Simulation." *Nucleic Acids Research*, 2016.
DOI: 10.1093/nar/gkw815 [15, 32, 33, 60, 105]

[98] Anna Reymer, Krystyna Zakrzewska, and Richard Lavery. "Sequence-Dependent Response of DNA to Torsional Stress: A Potential Biological Regulation Mechanism". *Nucleic Acids Research*, 2018.
DOI: 10.1093/nar/gkx1270 [15]

[99] Rifka Vlijm, Jaco v.d. Torre, and Cees Dekker. "Counterintuitive DNA Sequence Dependence in Supercoiling-Induced DNA Melting". *PLOS ONE*, 2015. ed. by Fenfei Leng.
DOI: 10.1371/journal.pone.0141576 [15, 62]

[100] Agnes Noy, Anthony Maxwell, and Sarah A. Harris. "Interference between Triplex and Protein Binding to Distal Sites on Supercoiled DNA". *Biophysical Journal*, 2017.
DOI: 10.1016/j.bpj.2016.12.034 [15, 65, 69]

[101] Rossitza N. Irobalieva, Jonathan M. Fogg, Daniel J. Catanese, Thana Sutthibutpong, Muyuan Chen, Anna K. Barker, Steven J. Ludtke, Sarah A. Harris, Michael F. Schmid, Wah Chiu, and Lynn Zechiedrich. "Structural Diversity of Supercoiled DNA". *Nature Communications*, 2015.
DOI: 10.1038/ncomms9440 [15, 60, 65, 69]

[102] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker. "Dynamics of DNA Supercoils". *Science*, 2012.
DOI: 10.1126/science.1225810 [16, 30]

[103] Job Dekker and Leonid Mirny. "The 3D Genome as Moderator of Chromosomal Communication". *Cell*, 2016.
DOI: 10.1016/j.cell.2016.02.007 [16]

[104] Hergen Brutzer, Nicholas Luzzietti, Daniel Klaue, and Ralf Seidel. "Energetics at the DNA Supercoiling Transition". *Biophysical Journal*, 2010.
DOI: 10.1016/j.bpj.2009.12.4292 [16, 63, 106, 145, 147, 152]

[105] Scott Forth, Christopher Deufel, Maxim Y. Sheinin, Bryan Daniels, James P. Sethna, and Michelle D. Wang. "Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules". *Physical Review Letters*, 2008.
DOI: 10.1103/PhysRevLett.100.148301 [16]

[106] Matthew Burman and Agnes Noy. "Atomic Description of the Reciprocal Action between Supercoils and Melting Bubbles on Linear DNA". *Physical Review Letters*, 2025.
DOI: 10.1103/PhysRevLett.134.038403 [16, 61, 62, 66, 75, 76, 80, 82, 136, 151, 152]

[107] T. R. Strick, V. Croquette, and D. Bensimon. "Homologous Pairing in Stretched Supercoiled DNA". *Proceedings of the National Academy of Sciences*, 1998.
DOI: 10.1073/pnas.95.18.10579 [16]

[108] T.R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette. "Behavior of Supercoiled DNA". *Biophysical Journal*, 1998.
DOI: 10.1016/S0006-3495(98)77908-1 [16]

[109] Richard M. Fye and Craig J. Benham. "Exact Method for Numerically Analyzing a Model of Local Denaturation in Superhelically Stressed DNA". *Physical Review E*, 1999.
DOI: 10.1103/PhysRevE.59.3408 [17, 54, 125]

[110] He Meng, Johan Bosman, Thijn van der Heijden, and John van Noort. "Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains." *Biophysical Journal*, 2014.
DOI: 10.1016/j.bpj.2014.01.017 [17, 75]

[111] Mahipal Ganji, Sung Hyun Kim, Jaco van der Torre, Elio Abbondanzieri, and Cees Dekker. "Intercalation-Based Single-Molecule Fluorescence Assay To Study DNA Supercoil Dynamics". *Nano Letters*, 2016.
DOI: 10.1021/acs.nanolett.6b02213 [17, 28, 30, 75, 106, 152]

[112] Christian Matek, Thomas E. Ouldridge, Jonathan P. K. Doye, and Ard A. Louis. "Plectoneme Tip Bubbles: Coupled Denaturation and Writhing in Supercoiled DNA." *Scientific Reports*, 2015.
DOI: 10.1038/srep07655 [17, 29, 33, 62, 75, 105]

[113] Craig J. Benham and Chengpeng Bi. "The Analysis of Stress-Induced Duplex Destabilization in Long Genomic DNA Sequences". *Journal of Computational Biology*, 2004.
DOI: 10.1089/cmb.2004.11.519 [17, 54, 62, 67, 69]

[114] John SantaLucia. "A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics". *Proceedings of the National Academy of Sciences*, 1998.
DOI: 10.1073/pnas.95.4.1460 [17, 56]

[115] Andrew Krueger, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. "Sequence-Dependent Basepair Opening in DNA Double Helix". *Biophysical Journal*, 2006.
DOI: 10.1529/biophysj.105.078774 [17]

[116] Huiquan Wang, Michiel Noordewier, and Craig J. Benham. "Stress-Induced DNA Duplex Destabilization (SIDD) in the E. Coli Genome: SIDD Sites Are Closely Associated with Promoters." *Genome Research*, 2004.
DOI: 10.1101/gr.2080004 [17]

[117] Jozef Adamcik, Jae-Hyung Jeon, Konrad J. Karczewski, Ralf Metzler, and Giovanni Dietler. "Quantifying Supercoiling-Induced Denaturation Bubbles in DNA". *Soft Matter*, 2012.
DOI: 10.1039/C2SM26089A [17, 63, 75, 84, 86, 89]

[118] Jack W. Shepherd, Sebastien Guilbaud, Zhaokun Zhou, Jamieson A. L. Howard, Matthew Burman, Charley Schaefer, Adam Kerrigan, Clare Steele-King, Agnes Noy, and Mark C. Leake. "Correlating Fluorescence Microscopy, Optical and Magnetic Tweezers to Study Single Chiral Biopolymers Such as DNA". *Nature Communications*, 2024.
DOI: 10.1038/s41467-024-47126-6 [17, 29, 30, 33, 61, 151, 153]

[119] John SantaLucia Jr and Donald Hicks. "The Thermodynamics of DNA Structural Motifs". *Annual Review of Biophysics and Biomolecular Structure*, 2004.
DOI: 10.1146/annurev.biophys.32.110601.141800 [18]

[120] Paul J. Hagerman. "Investigation of the Flexibility of DNA Using Transient Electric Birefringence". *Biopolymers*, 1981.
DOI: 10.1002/bip.1981.360200710 [18, 122]

[121] Gerald S. Manning. "The Persistence Length of DNA Is Reached from the Persistence Length of Its Null Isomer through an Internal Electrostatic Stretching Force". *Biophysical Journal*, 2006.
DOI: 10.1529/biophysj.106.089029 [18]

[122] Justin P. Peters and L. James Maher. "DNA Curvature and Flexibility in Vitro and in Vivo". *Quarterly Reviews of Biophysics*, 2010.
DOI: 10.1017/S0033583510000077 [18, 31]

[123] Alexey Savelyev, Christopher K. Materese, and Garegin A. Papoian. "Is DNA's Rigidity Dominated by Electrostatic or Nonelectrostatic Interactions?" *Journal of the American Chemical Society*, 2011.
DOI: 10.1021/ja207984z [18]

[124] Ekaterina Protozanova, Peter Yakovchuk, and Maxim D. Frank-Kamenetskii. "Stacked–Unstacked Equilibrium at the Nick Site of DNA". *Journal of Molecular Biology*, 2004.
DOI: 10.1016/j.jmb.2004.07.075 [18]

[125] Paul J. Hagerman. "Flexibility of DNA". *Annual Review of Biophysics*, 1988.
DOI: 10.1146/annurev.bb.17.060188.001405 [19]

[126] Claudio Anselmi, Pasquale De Santis, Raffaella Paparcone, Maria Savino, and Anita Scipioni. "From the Sequence to the Superstructural Properties of DNAs". *Biophysical Chemistry*, 2002.
DOI: 10.1016/S0301-4622(01)00246-0 [19]

[127] Francesco Pedone, Filomena Mazzei, Mirella Matzeu, and Flavia Barone. "Torsional Constant of 27-Mer DNA Oligomers of Different Sequences". *Biophysical Chemistry*, 2001.
DOI: 10.1016/S0301-4622(01)00232-0 [19]

[128] Filip Lankas, Jirí Sponer, Jörg Langowski, and Thomas E. 3rd Cheatham. "DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations." *Biophysical Journal*, 2003.
DOI: 10.1016/S0006-3495(03)74710-9 [19]

[129] Jonathan S. Mitchell, Jaroslaw Glowacki, Alexandre E. Grandchamp, Robert S. Manning, and John H. Maddocks. "Sequence-Dependent Persistence Lengths of DNA". *Journal of Chemical Theory and Computation*, 2017.
DOI: 10.1021/acs.jctc.6b00904 [19]

[130] John A. Schellman and Stephen C. Harvey. "Static Contributions to the Persistence Length of DNA and Dynamic Contributions to DNA Curvature". *Recent Advances in Biophysical Chemistry*, 1995.

DOI: 10.1016/0301-4622(94)00144-9 [19]

[131] Stephanie Geggier and Alexander Vologodskii. "Sequence Dependence of DNA Bending Rigidity". *Proceedings of the National Academy of Sciences*, 2010.
DOI: 10.1073/pnas.1004809107 [19]

[132] Hui-Min Chuang, Jeffrey G. Reifenberger, Han Cao, and Kevin D. Dorfman. "Sequence-Dependent Persistence Length of Long DNA". *Physical Review Letters*, 2017.
DOI: 10.1103/PhysRevLett.119.227802 [19]

[133] Victor Velasco-Berrelleza, Matthew Burman, Jack W. Shepherd, Mark C. Leake, Ramin Golestanian, and Agnes Noy. "SerraNA: A Program to Determine Nucleic Acids Elasticity from Simulation Data". *Physical Chemistry Chemical Physics*, 2020.
DOI: 10.1039/D0CP02713H [19]

[134] Agnes Noy and Ramin Golestanian. "The Chirality of DNA: Elasticity Cross-Terms at Base-Pair Level Including A-Tracts and the Influence of Ionic Strength". *The Journal of Physical Chemistry B*, 2010.
DOI: 10.1021/jp104133j [19]

[135] Marco Pasi et al. "mABC: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA". *Nucleic Acids Research*, 2014.
DOI: 10.1093/nar/gku855 [19]

[136] Tania Gardasevic and Agnes Noy. "The Impact of Sequence Periodicity on DNA Mechanics: Investigating the Origin of A-tract's Curvature". *Nanoscale*, 2024.
DOI: 10.1039/D4NR02571G [20]

[137] Larissa Milano, Amit Gautam, and Keith W. Caldecott. "DNA Damage and Transcription Stress". *Molecular Cell*, 2024.
DOI: 10.1016/j.molcel.2023.11.014 [20]

[138] Douglas Brutlag and Arthur Kornberg. "Enzymatic Synthesis of Deoxyribonucleic Acid. 36. A Proofreading Function for the 3' Leads to 5' Exonuclease Activity in Deoxyribonucleic Acid Polymerases." *The Journal of Biological Chemistry*, 1972. [20]

[139] Heather M. O'Hagan, Helai P. Mohammad, and Stephen B. Baylin. "Double Strand Breaks Can Initiate Gene Silencing and SIRT1-dependent Onset of DNA Methylation in an Exogenous Promoter CpG Island." *PLoS Genetics*, 2008.
DOI: 10.1371/journal.pgen.1000155 [20]

[140] Damien Brégeon and Paul W. Doetsch. "Transcriptional Mutagenesis: Causes and Involvement in Tumour Development". *Nature Reviews Cancer*, 2011.
DOI: 10.1038/nrc3006 [20]

[141] Harvey F. Lodish. *Molecular Cell Biology*. 5th ed. W.H. Freeman and Co., 2004. chap. 1 volume (various pagings) : illustrations (some color) ; 29 cm. [20]

[142] Nives Pećina-Šlaus, Anja Kafka, Iva Salamon, and Anja Bukovac. "Mismatch Repair Pathway, Genome Stability and Cancer". *Frontiers in Molecular Biosciences*, 2020.
DOI: 10.3389/fmolb.2020.00122 [21]

[143] Ravi R. Iyer, Anna Pluciennik, Vickers Burdett, and Paul L. Modrich. "DNA Mismatch Repair: Functions and Mechanisms". *Chemical Reviews*, 2006.
DOI: 10.1021/cr0404794 [20, 21]

[144] Myron F. Goodman, Steven Creighton, Linda B. Bloom, John Petruska, and Thomas A. Kunkel. "Biochemical Basis of DNA Replication Fidelity". *Critical Reviews in Biochemistry and Molecular Biology*, 1993.
DOI: 10.3109/10409239309086792 [20]

[145] C. Kunz, Y. Saito, and P. Schär. "DNA Repair in Mammalian Cells". *Cellular and Molecular Life Sciences*, 2009.
DOI: 10.1007/s00018-009-8739-9 [20]

[146] Thomas A. Kunkel. "DNA Replication Fidelity*". *Journal of Biological Chemistry*, 2004.
DOI: 10.1074/jbc.R400006200 [21]

[147] Wei Yang. "Structure and Mechanism for DNA Lesion Recognition". *Cell Research*, 2008.
DOI: 10.1038/cr.2007.116 [21, 28, 29]

[148] Dekang Liu, Guido Keijzers, and Lene Juel Rasmussen. "DNA Mismatch Repair and Its Many Roles in Eukaryotic Cells". *Mutation Research/Reviews in Mutation Research*, 2017.
DOI: 10.1016/j.mrrev.2017.07.001 [21]

[149] Paul Modrich. "Mechanisms in E. Coli and Human Mismatch Repair (Nobel Lecture)". *Angewandte Chemie International Edition*, 2016.
DOI: 10.1002/anie.201601412 [21]

[150] Megan A Phillips, Jacob L Steenwyk, Xing-Xing Shen, and Antonis Rokas. "Examination of Gene Loss in the DNA Mismatch Repair Pathway and Its Mutational Consequences in a Fungal Phylum". *Genome Biology and Evolution*, 2021.
DOI: 10.1093/gbe/evab219 [21]

[151] Ujani Chakraborty and Eric Alani. "Understanding How Mismatch Repair Proteins Participate in the Repair/Anti-Recombination Decision". *FEMS Yeast Research*, 2016.
DOI: 10.1093/femsyr/fow071 [21]

[152] Peter D. Turnpenny. *Emery's Elements of Medical Genetics*. Edinburgh ; New York : Elsevier/Churchill Livingstone, 2005.
(Visited on 02/07/2025) [22]

[153] Junhong Guan and Guo-Min Li. "DNA Mismatch Repair in Cancer Immunotherapy". *NAR Cancer*, 2023.
DOI: 10.1093/narcan/zcad031 [22]

[154] Giulia Rossetti, Pablo D. Dans, Irene Gomez-Pinto, Ivan Ivani, Carlos Gonzalez, and Modesto Orozco. "The Structural Impact of DNA Mismatches." *Nucleic Acids Research*, 2015.
DOI: 10.1093/nar/gkv254 [22, 107, 108, 113, 118, 129, 130, 144, 147]

[155] Tomas Lindahl. "DNA Glycosylases in DNA Repair". IN: *Mechanisms of DNA Damage*

and Repair: Implications for Carcinogenesis and Risk Assessment. ed. by Michael G. Simic, Lawrence Grossman, Arthur C. Upton, and David S. Bergtold. Springer US, 1986.
DOI: 10.1007/978-1-4615-9462-8_36 [22]

[156] Paola Fortini and Eugenia Dogliotti. "Base Damage and Single-Strand Break Repair: Mechanisms and Functional Significance of Short- and Long-Patch Repair Subpathways". *DNA Repair*, 2007.
DOI: 10.1016/j.dnarep.2006.10.008 [22]

[157] Myriam Talpaert-Borlè. "Formation, Detection and Repair of AP Sites". *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1987.
DOI: 10.1016/0027-5107(87)90286-7 [22]

[158] Burton Tropp. *Molecular Biology: Genes to Proteins*. 4th ed. Jones & Bartlett, 2012. [22]

[159] Claudia P. Gonzalez-Hunt, Mandheer Wadhwa, and Laurie H. Sanders. "DNA Damage by Oxidative Stress: Measurement Strategies for Two Genomes". *Current Opinion in Toxicology*, 2018.
DOI: 10.1016/j.cotox.2017.11.001 [22]

[160] Sriram Kanvah, Joshy Joseph, Gary B. Schuster, Robert N. Barnett, Charles L. Cleveland, and Uzi Landman. "Oxidation of DNA: Damage to Nucleobases". *Accounts of Chemical Research*, 2010.
DOI: 10.1021/ar900175a [22]

[161] James A. Swenberg, Kun Lu, Benjamin C. Moeller, Lina Gao, Patricia B. Upton, Jun Nakamura, and Thomas B. Starr. "Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment". *Toxicological Sciences*, 2011.
DOI: 10.1093/toxsci/kfq371 [22]

[162] Michelle L. Hamilton, Holly Van Remmen, Jessica A. Drake, Hong Yang, Zhong Mao Guo, Kristen Kewitt, Christi A. Walter, and Arlan Richardson. "Does Oxidative Damage to DNA Increase with Age?" *Proceedings of the National Academy of Sciences*, 2001.
DOI: 10.1073/pnas.171202698 [22]

[163] Ben Nie, Wei Gan, Fei Shi, Guo-Xin Hu, Lian-Guo Chen, Hiroshi Hayakawa, Mutsuo Sekiguchi, and Jian-Ping Cai. "Age-Dependent Accumulation of 8-Oxoguanine in the DNA and RNA in Various Rat Tissues". *Oxidative Medicine and Cellular Longevity*, 2013.
DOI: 10.1155/2013/303181 [22]

[164] Barry Halliwell. "Oxidative Stress and Cancer: Have We Moved Forward?" *Biochemical Journal*, 2006.
DOI: 10.1042/BJ20061131 [22]

[165] Onyou Hwang. "Role of Oxidative Stress in Parkinson's Disease". *Experimental Neurobiology*, 2013.
DOI: 10.5607/en.2013.22.1.11 [22]

[166] Marian Valko, Dieter Leibfritz, Jan Moncol, Mark T.D. Cronin, Milan Mazur, and Joshua Telser. "Free Radicals and Antioxidants in Normal Physiological Functions and Human Disease". *The International Journal of Biochemistry & Cell Biology*, 2007.
DOI: 10.1016/j.biocel.2006.07.001 [22]

[167] Neelam Singh, Arvinder K. Dhalla, Charita Seneviratne, and Pawan K. Singal. "Oxidative Stress and Heart Failure". *Molecular and Cellular Biochemistry*, 1995.
DOI: 10.1007/BF00944786 [22]

[168] S Jill James, Paul Cutler, Stepan Melnyk, Stefanie Jernigan, Laurette Janak, David W Gaylor, and James A Neubrander. "Metabolic Biomarkers of Increased Oxidative Stress and Impaired Methylation Capacity in Children with Autism". *The American Journal of Clinical Nutrition*, 2004.
DOI: 10.1093/ajcn/80.6.1611 [22]

[169] Keith C. Cheng, D. S. Cahill, Hiroshi Kasai, Susumu Nishimura, and Lawrence A. Loeb. "8-Hydroxyguanine, an Abundant Form of Oxidative DNA Damage, Causes G-T and A-C Substitutions." *The Journal of Biological Chemistry*, 1992.
[22]

[170] Tae-Hee Lee and Tae-Hong Kang. "DNA Oxidation and Excision Repair Pathways". *International Journal of Molecular Sciences*, 2019.
DOI: 10.3390/ijms20236092 [23]

[171] R. B. Setlow. "Cyclobutane-Type Pyrimidine Dimers in Polynucleotides". *Science*, 1966.
DOI: 10.1126/science.153.3734.379 [23]

[172] Joyce T. Reardon and Aziz Sancar. "Purification and Characterization of Escherichia Coli and Human Nucleotide Excision Repair Enzyme Systems". IN: *Methods in Enzymology*. Academic Press, 2006.
DOI: 10.1016/S0076-6879(06)08012-8 [23]

[173] Michael G. Kemp and Aziz Sancar. "DNA Excision Repair: Where Do All the Dimers Go?" *Cell cycle (Georgetown, Tex.)*, 2012.
DOI: 10.4161/cc.21126 [23]

[174] Yingfu Li and Ronald R. Breaker. "Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group". *Journal of the American Chemical Society*, 1999.
DOI: 10.1021/ja990592p [24]

[175] Stephanie A Nick McElhinny, Dinesh Kumar, Alan B Clark, Danielle L Watt, Brian E Watts, Else-Britt Lundström, Erik Johansson, Andrei Chabes, and Thomas A Kunkel. "Genome Instability Due to Ribonucleotide Incorporation into DNA". *Nature Chemical Biology*, 2010.
DOI: 10.1038/nchembio.424 [24]

[176] Alan B. Clark, Scott A. Lujan, Grace E. Kissling, and Thomas A. Kunkel. "Mismatch Repair-Independent Tandem Repeat Sequence Instability Resulting from Ribonucleotide Incorporation by DNA Polymerase ϵ". *DNA Repair*, 2011.
DOI: 10.1016/j.dnarep.2011.02.001 [24]

[177] Nayun Kim, Shar-yin N. Huang, Jessica S. Williams, Yue C. Li, Alan B. Clark,

Jang-Eun Cho, Thomas A. Kunkel, Yves Pommier, and Sue Jinks-Robertson. "Mutagenic Processing of Ribonucleotides in DNA by Yeast Topoisomerase I". *Science*, 2011.
DOI: 10.1126/science.1205016 [24]

[178] Bret D. Wallace and R. Scott Williams. "Ribonucleotide Triggered DNA Damage and RNA-DNA Damage Responses." *RNA biology*, 2014.
DOI: 10.4161/15476286.2014.992283 [24]

[179] Y Machida, T Okazaki, and R Okazaki. "Discontinuous Replication of Replicative Form DNA from Bacteriophage phiX174." *Proceedings of the National Academy of Sciences*, 1977.
DOI: 10.1073/pnas.74.7.2776 [24]

[180] Pablo Huertas and Andrés Aguilera. "Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination". *Molecular Cell*, 2003.
DOI: 10.1016/j.molcel.2003.08.010 [24]

[181] Xialu Li and James L. Manley. "Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability". *Cell*, 2005.
DOI: 10.1016/j.cell.2005.06.008 [24]

[182] Klaus Förstemann and Joachim Lingner. "Telomerase Limits the Extent of Base Pairing between Template RNA and Telomeric DNA". *EMBO reports*, 2005.
DOI: 10.1038/sj.embor.7400374 [24]

[183] Catherine M. Joyce. "Choosing the Right Sugar: How Polymerases Select a Nucleotide Substrate". *Proceedings of the National Academy of Sciences*, 1997.
DOI: 10.1073/pnas.94.5.1619 [24]

[184] Sonya Vengrova and Jacob Z Dalgaard. "The Wild-type Schizosaccharomyces Pombe Mat1 Imprint Consists of Two Ribonucleotides". *EMBO reports*, 2006.
DOI: 10.1038/sj.embor.7400576 [24]

[185] Vanessa Kellner and Brian Luke. "Molecular and Physiological Consequences of Faulty Eukaryotic Ribonucleotide Excision Repair." *The EMBO Journal*, 2020.
DOI: 10.15252/embj.2019102309 [24]

[186] Martin A.M. Reijns et al. "Enzymatic Removal of Ribonucleotides from DNA Is Essential for Mammalian Genome Integrity and Development". *Cell*, 2012.
DOI: 10.1016/j.cell.2012.04.011 [24]

[187] Bjoern Hiller, Martin Achleitner, Silke Glage, Ronald Naumann, Rayk Behrendt, and Axel Roers. "Mammalian RNase H2 Removes Ribonucleotides from DNA to Maintain Genome Integrity". *Journal of Experimental Medicine*, 2012.
DOI: 10.1084/jem.20120876 [24]

[188] Justin L. Sparks, Hyongi Chon, Susana M. Cerritelli, Thomas A. Kunkel, Erik Johansson, Robert J. Crouch, and Peter M. Burgers. "RNase H2-Initiated Ribonucleotide Excision Repair". *Molecular Cell*, 2012.
DOI: 10.1016/j.molcel.2012.06.035 [24]

[189] Bjorn Rydberg and John Game. "Excision of Misincorporated Ribonucleotides in DNA by RNase H (Type 2) and FEN-1 in Cell-Free Extracts". *Proceedings of the National Academy of Sciences*, 2002.
DOI: 10.1073/pnas.262591699 [24]

[190] Marcin Nowotny, Sergei A. Gaidamakov, Rodolfo Ghirlando, Susana M. Cerritelli, Robert J. Crouch, and Wei Yang. "Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription". *Molecular Cell*, 2007.
DOI: 10.1016/j.molcel.2007.10.021 [24]

[191] Paul S. Eder, Roxanne Y. Walder, and Joseph A. Walder. "Substrate Specificity of Human RNase H1 and Its Role in Excision Repair of Ribose Residues Misincorporated in DNA". *Biochimie*, 1993.
DOI: 10.1016/0300-9084(93)90033-O [24]

[192] Sara Pizzi, Sarah Sertic, Simona Orcesi, Cristina Cereda, Marika Bianchi, Andrew P. Jackson, Federico Lazzaro, Paolo Plevani, and Marco Muzi-Falconi. "Reduction of hRNase H2 Activity in Aicardi–Goutières Syndrome Cells Leads to Replication Stress and Genome Instability". *Human Molecular Genetics*, 2015.
DOI: 10.1093/hmg/ddu485 [24]

[193] Yanick J Crow et al. "Mutations in Genes Encoding Ribonuclease H2 Subunits Cause Aicardi-Goutières Syndrome and Mimic Congenital Viral Brain Infection". *Nature Genetics*, 2006.
DOI: 10.1038/ng1842 [24]

[194] Jörg Fahrer and Bernd Kaina. "O6 -Methylguanine-DNA Methyltransferase in the Defense against N -Nitroso Compounds and Colorectal Cancer". *Carcinogenesis*, 2013.
DOI: 10.1093/carcin/bgt275 [24]

[195] Daniel B. Yarosh. "The Role of O6-methylguanine-DNA Methyltransferase in Cell Survival, Mutagenesis and Carcinogenesis". *Mutation Research/DNA Repair Reports*, 1985.
DOI: 10.1016/0167-8817(85)90034-3 [24]

[196] Aghdass Rasouli-Nia, Sigbhat-Ullah, Razmik Mirzayans, Malcolm C. Paterson, and Rufus S. Day. "On the Quantitative Relationship between O6-ethylguanine Residues, in Genomic DNA and Production of Sister-Chromatid Exchanges, Mutations and Lethal Events in a Mer- Human Tumor Cell Line". *Mutation Research/DNA Repair*, 1994.
DOI: 10.1016/0921-8777(94)90074-4 [24]

[197] Peter J. Abbott and Roy Saffhill. "DNA Synthesis with Methylated Poly(dC-dG) Templates. Evidence for a Competitive Nature to Miscoding by O6-methylguanine". *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 1979.
DOI: 10.1016/0005-2787(79)90125-4 [24]

[198] Panagiotis Georgiadis, Clive A. Smith, and Peter F. Swann. "Nitrosamine-Induced Cancer: Selective Repair and Conformational Differences between O6-methylguanine Residues in Different Positions in and around Codon 12 of Rat H-ras." *Cancer research*, 1991.
[24]

[199] Dimitrios Iliopoulos, Pagona Oikonomou, Ioannis Messinis, and Aspasia Tsezou. "Correlation of Promoter Hypermethylation in hTERT, DAPK and MGMT Genes with Cervical Oncogenesis Progression". *Oncology Reports*, 2009.
DOI: 10.3892/or_00000425 [24]

[200] Lanlan Shen, Yutaka Kondo, Gary L. Rosner, Lianchun Xiao, Natalie Supunpong Hernandez, Jill Vilaythong, P. Scott Houlihan, Robert S. Krouse, Anil R. Prasad, Janine G. Einspahr, Julie Buckmeier, David S. Alberts, Stanley R. Hamilton, and Jean-Pierre J. Issa. "MGMT Promoter Methylation and Field Defect in Sporadic Colorectal Cancer". *JNCI: Journal of the National Cancer Institute*, 2005.
DOI: 10.1093/jnci/dji275 [24]

[201] Kyung-Hwa Lee, Ji-Shin Lee, Jong-Hee Nam, Chan Choi, Min-Cheol Lee, Chang-Soo Park, Sang-Woo Juhng, and Jae-Hyuk Lee. "Promoter Methylation Status of hMLH1, hMSH2, and MGMT Genes in Colorectal Cancer Associated with Adenoma–Carcinoma Sequence". *Langenbeck's Archives of Surgery*, 2011.
DOI: 10.1007/s00423-011-0812-9 [24]

[202] Alessio Amatu, Andrea Sartore-Bianchi, Catia Moutinho, Alessandro Belotti, Katia Bencardino, Giuseppe Chirico, Andrea Cassingena, Francesca Rusconi, Anna Esposito, Michele Nichelatti, Manel Esteller, and Salvatore Siena. "Promoter CpG Island Hypermethylation of the DNA Repair Enzyme MGMT Predicts Clinical Response to Dacarbazine in a Phase II Study for Metastatic Colorectal Cancer". *Clinical Cancer Research*, 2013.
DOI: 10.1158/1078-0432.CCR-12-3518 [24]

[203] Pooneh Mokarram, M. Zamani, S. Kavousipour, F. Naghibalhossaini, C. Irajie, M. Moradi Sarabi, and S. V. Hosseini. "Different Patterns of DNA Methylation of the Two Distinct O6-methylguanine-DNA Methyltransferase (O6-MGMT) Promoter Regions in Colorectal Cancer". *Molecular Biology Reports*, 2013.
DOI: 10.1007/s11033-012-2465-3 [24]

[204] Mohamed Mokhtar, Kazuya Kondo, Toshiaki Namura, Abdellah H.K. Ali, Yui Fujita, Chikako Takai, Hiromitsu Takizawa, Yasushi Nakagawa, Hiroaki Toba, Koichiro Kajiura, Mitsuteru Yoshida, Gyokei Kawakami, Shoji Sakiyama, and Akira Tangoku. "Methylation and Expression Profiles of MGMT Gene in Thymic Epithelial Tumors". *Lung Cancer*, 2014.
DOI: 10.1016/j.lungcan.2013.12.004 [24]

[205] Xiao-Ping Zou, Bin Zhang, Xiao-Qi Zhang, Min Chen, Jun Cao, and Wen-Jia Liu. "Promoter Hypermethylation of Multiple Genes in Early Gastric Adenocarcinoma and Precancerous Lesions". *Human Pathology*, 2009.
DOI: 10.1016/j.humpath.2009.01.029 [24]

[206] J. Jin, L. Xie, C. H. Xie, and Y. F. Zhou. "Aberrant DNA Methylation of MGMT and hMLH1 Genes in Prediction of Gastric Cancer." *Genetics and molecular research : GMR*, 2014.
DOI: 10.4238/2014.May.30.9 [24]

[207] Suneet Dinglay, Sarah C. Trewick, Tomas Lindahl, and Barbara Sedgwick. "Defective

Processing of Methylated Single-Stranded DNA by E. Coli AlkB Mutants." *Genes & development*, 2000.

[24]

[208] Keith W. Caldecott. "Causes and Consequences of DNA Single-Strand Breaks". *Trends in Biochemical Sciences*, 2024.
DOI: 10.1016/j.tibs.2023.11.001 [25]

[209] Wendy Knapp Pogozelski and Thomas D. Tullius. "Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety". *Chemical Reviews*, 1998.
DOI: 10.1021/cr960437i [25]

[210] Jeffrey J. Pouliot, Carol A. Robertson, and Howard A. Nash. "Pathways for Repair of Topoisomerase I Covalent Complexes in Saccharomyces Cerevisiae". *Genes to Cells*, 2001.
DOI: 10.1046/j.1365-2443.2001.00452.x [25]

[211] Yves Pommier, Yilun Sun, Shar-yin N. Huang, and John L. Nitiss. "Roles of Eukaryotic Topoisomerases in Transcription, Replication and Genomic Stability". *Nature Reviews Molecular Cell Biology*, 2016.
DOI: 10.1038/nrm.2016.111 [25]

[212] Ivan Ahel, Ulrich Rass, Sherif F. El-Khamisy, Sachin Katyal, Paula M. Clements, Peter J. McKinnon, Keith W. Caldecott, and Stephen C. West. "The Neurodegenerative Disease Protein Aprataxin Resolves Abortive DNA Ligation Intermediates". *Nature*, 2006.
DOI: 10.1038/nature05164 [25]

[213] Percy Tumbale, Jessica S. Williams, Matthew J. Schellenberg, Thomas A. Kunkel, and R. Scott Williams. "Aprataxin Resolves Adenylated RNA–DNA Junctions to Maintain Genome Integrity". *Nature*, 2014.
DOI: 10.1038/nature12824 [25]

[214] Marie-France Langelier, Travis Eisemann, Amanda A Riccio, and John M Pascal. "PARP Family Enzymes: Regulation and Catalysis of the Poly(ADP-ribose) Posttranslational Modification". *Current Opinion in Structural Biology*, 2018.
DOI: 10.1016/j.sbi.2018.11.002 [25]

[215] Keith W. Caldecott. "Single-Strand Break Repair and Genetic Disease". *Nature Reviews Genetics*, 2008.
DOI: 10.1038/nrg2380 [25]

[216] Wendy J. Cannan and David S. Pederson. "Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin." *Journal of Cellular Physiology*, 2016.
DOI: 10.1002/jcp.25048 [25]

[217] J. Kent Moore and James E. Haber. "Cell Cycle and Genetic Requirements of Two Pathways of Nonhomologous End-Joining Repair of Double-Strand Breaks in Saccharomyces Cerevisiae." *Molecular and Cellular Biology*, 1996.
DOI: 10.1128/MCB.16.5.2164 [25]

[218] Joe Budman and Gilbert Chu. "Processing of DNA for Nonhomologous End-Joining by Cell-Free Extract." *The EMBO Journal*, 2005.
DOI: 10.1038/sj.emboj.7600563 [25]

[219] Silvia Espejel, Sonia Franco, Sandra Rodríguez-Perales, Simon D. Bouffler, Juan C. Cigudosa, and María A. Blasco. "Mammalian Ku86 Mediates Chromosomal Fusions and Apoptosis Caused by Critically Short Telomeres." *The EMBO Journal*, 2002.
DOI: 10.1093/emboj/21.9.2207 [25]

[220] Colin W. Garvie and Cynthia Wolberger. "Recognition of Specific DNA Sequences". *Molecular Cell*, 2001.
DOI: 10.1016/S1097-2765(01)00392-6 [25, 27]

[221] Nadrian C. Seeman, John M. Rosenberg, and Alexander Rich. "Sequence-Specific Recognition of Double Helical Nucleic Acids by Proteins." *Proceedings of the National Academy of Sciences*, 1976.
DOI: 10.1073/pnas.73.3.804 [25]

[222] Remo Rohs, Sean M. West, Alona Sosinsky, Peng Liu, Richard S. Mann, and Barry Honig. "The Role of DNA Shape in Protein–DNA Recognition". *Nature*, 2009.
DOI: 10.1038/nature08473 [25, 26]

[223] Remo Rohs, Xiangshu Jin, Sean M. West, Rohit Joshi, Barry Honig, and Richard S. Mann. "Origins of Specificity in Protein-DNA Recognition". *Annual Review of Biochemistry*, 2010.
DOI: 10.1146/annurev-biochem-060408-091030 [26]

[224] Eran Segal and Jonathan Widom. "Poly(dA:dT) Tracts: Major Determinants of Nucleosome Organization." *Current Opinion in Structural Biology*, 2009.
DOI: 10.1016/j.sbi.2009.01.004 [26]

[225] Noam Kaplan, Irene K. Moore, Yvonne Fondufe-Mittendorf, Andrea J. Gossett, Desiree Tillo, Yair Field, Emily M. LeProust, Timothy R. Hughes, Jason D. Lieb, Jonathan Widom, and Eran Segal. "The DNA-encoded Nucleosome Organization of a Eukaryotic Genome." *Nature*, 2009.
DOI: 10.1038/nature07667 [26]

[226] Tali Raveh-Sadka, Michal Levo, Uri Shabi, Boaz Shany, Leeat Keren, Maya Lotan-Pompan, Danny Zeevi, Eilon Sharon, Adina Weinberger, and Eran Segal. "Manipulating Nucleosome Disfavoring Sequences Allows Fine-Tune Regulation of Gene Expression in Yeast". *Nature Genetics*, 2012.
DOI: 10.1038/ng.2305 [26]

[227] Yair Field, Noam Kaplan, Yvonne Fondufe-Mittendorf, Irene K. Moore, Eilon Sharon, Yaniv Lubling, Jonathan Widom, and Eran Segal. "Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals". *PLOS Computational Biology*, 2008.
DOI: 10.1371/journal.pcbi.1000216 [26]

[228] Vishwanath Iyer and Kevin Struhl. "Poly(dA:dT), a Ubiquitous Promoter Element That Stimulates Transcription via Its Intrinsic DNA Structure." *The EMBO Journal*, 1995.

DOI: 10.1002/j.1460-2075.1995.tb07255.x [26]

[229] Paul J. Hagerman. "Sequence-Directed Curvature of DNA". *Annual Review of Biochemistry*, 1990.
DOI: 10.1146/annurev.bi.59.070190.003543 [26]

[230] Tali E. Haran and Udayan Mohanty. "The Unique Structure of A-tracts and Intrinsic DNA Bending". *Quarterly Reviews of Biophysics*, 2009.
DOI: 10.1017/S0033583509004752 [26]

[231] Anthony Tubbs et al. "Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse". *Cell*, 2018.
DOI: 10.1016/j.cell.2018.07.011 [26]

[232] Christian Frøkjær-Jensen, Nimit Jain, Loren Hansen, M. Wayne Davis, Yongbin Li, Di Zhao, Karine Rebora, Jonathan R.M. Millet, Xiao Liu, Stuart K. Kim, Denis Dupuy, Erik M. Jorgensen, and Andrew Z. Fire. "An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the C. Elegans Germline". *Cell*, 2016.
DOI: 10.1016/j.cell.2016.05.072 [26]

[233] Rashmi S. Hegde, Steven R. Grossman, Laimonis A. Laimins, and Paul B. Sigler. "Crystal Structure at 1.7 Å of the Bovine Papillomavirus-1 E2 DMA-binding Domain Bound to Its DNA Target". *Nature*, 1992.
DOI: 10.1038/359505a0 [26]

[234] Phoebe A Rice, Shu-wei Yang, Kiyoshi Mizuuchi, and Howard A Nash. "Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn". *Cell*, 1996.
DOI: 10.1016/S0092-8674(00)81824-3 [26, 31]

[235] Youngchang Kim, James. H. Geiger, Steven Hahn, and Paul B. Sigler. "Crystal Structure of a Yeast TBP/TATA-box Complex". *Nature*, 1993.
DOI: 10.1038/365512a0 [26]

[236] Joseph L. Kim, Dimitar B. Nikolov, and Stephen K. Burley. "Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element". *Nature*, 1993.
DOI: 10.1038/365520a0 [26]

[237] Stefjord Todolli, Robert T. Young, Abigail S. Watkins, Antonio Bu Sha, John Yager, and Wilma K. Olson. "Surprising Twists in Nucleosomal DNA with Implication for Higher-order Folding". *Journal of Molecular Biology*, 2021.
DOI: 10.1016/j.jmb.2021.167121 [26]

[238] Lu Bai and Alexandre V. Morozov. "Gene Regulation by Nucleosome Positioning". *Trends in Genetics*, 2010.
DOI: 10.1016/j.tig.2010.08.003 [26]

[239] Remus T. Dame, Fatema-Zahra M. Rashid, and David C. Grainger. "Chromosome Organization in Bacteria: Mechanistic Insights into Genome Structure and Function". *Nature Reviews Genetics*, 2020.
DOI: 10.1038/s41576-019-0185-4 [26]

[240] Rosana Collepardo-Guevara and Tamar Schlick. "Chromatin Fiber Polymorphism

Triggered by Variations of DNA Linker Lengths". *Proceedings of the National Academy of Sciences*, 2014.
DOI: 10.1073/pnas.1315872111 [26]

[241] Jenny Hizver, Haim Rozenberg, Felix Frolow, Dov Rabinovich, and Zippora Shakked. "DNA Bending by an Adenine-Thymine Tract and Its Role in Gene Regulation". *Proceedings of the National Academy of Sciences*, 2001.
DOI: 10.1073/pnas.151247298 [26]

[242] Remo Rohs, Heinz Sklenar, and Zippora Shakked. "Structural and Energetic Origins of Sequence-Specific DNA Bending: Monte Carlo Simulations of Papillomavirus E2-DNA Binding Sites". *Structure*, 2005.
DOI: 10.1016/j.str.2005.07.005 [26]

[243] Yesenia Rodriguez, John M. Hinz, and Michael J. Smerdon. "Accessing DNA Damage in Chromatin: Preparing the Chromatin Landscape for Base Excision Repair". *Cutting-edge Perspectives in Genomic Maintenance II*, 2015.
DOI: 10.1016/j.dnarep.2015.04.021 [26, 105]

[244] Yesenia Rodriguez and Michael J. Smerdon. "The Structural Location of DNA Lesions in Nucleosome Core Particles Determines Accessibility by Base Excision Repair Enzymes *". *Journal of Biological Chemistry*, 2013.
DOI: 10.1074/jbc.M112.441444 [26, 105]

[245] Katharina Bilotti, Erin E. Kennedy, Chuxuan Li, and Sarah Delaney. "Human OGG1 Activity in Nucleosomes Is Facilitated by Transient Unwrapping of DNA and Is Influenced by the Local Histone Environment." *DNA repair*, 2017.
DOI: 10.1016/j.dnarep.2017.08.010 [26]

[246] Amalthiya Prasad, Susan S. Wallace, and David S. Pederson. "Initiation of Base Excision Repair of Oxidative Lesions in Nucleosomes by the Human, Bifunctional DNA Glycosylase NTH1". *Molecular and Cellular Biology*, 2007.
DOI: 10.1128/MCB.00791-07 [26]

[247] Katharina Bilotti, Mary E. Tarantino, and Sarah Delaney. "Human Oxoguanine Glycosylase 1 Removes Solution Accessible 8-Oxo-7,8-Dihydroguanine Lesions from Globally Substituted Nucleosomes Except in the Dyad Region". *Biochemistry*, 2018.
DOI: 10.1021/acs.biochem.7b01125 [26]

[248] Erin E. Kennedy, Chuxuan Li, and Sarah Delaney. "Global Repair Profile of Human Alkyladenine DNA Glycosylase on Nucleosomes Reveals DNA Packaging Effects". *ACS Chemical Biology*, 2019.
DOI: 10.1021/acschembio.9b00263 [26]

[249] José Pérez-Martín, Fernando Rojo, and Victor de Lorenzo. "Promoters Responsive to DNA Bending: A Common Theme in Prokaryotic Gene Expression." *Microbiological Reviews*, 1994.
DOI: 10.1128/mr.58.2.268-290 [26]

[250] Aditi Kanhere and Manju Bansal. "Structural Properties of Promoters: Similarities and Differences between Prokaryotes and Eukaryotes". *Nucleic Acids Research*, 2005.
DOI: 10.1093/nar/gki627 [26]

[251] N. Olivares-Zavaleta, R. Jáuregui, and E. Merino. “Genome Analysis of Escherichia Coli Promoter Sequences Evidences That DNA Static Curvature Plays a More Important Role in Gene Transcription than Has Previously Been Anticipated”. *Genomics*, 2006.
DOI: 10.1016/j.ygeno.2005.11.023 [26]

[252] Beatrice ten Heggeler and Walter Wahli. “Visualization of RNA Polymerase II Ternary Transcription Complexes Formed in Vitro on a Xenopus Laevis Vitellogenin Gene.” *The EMBO Journal*, 1985.
DOI: 10.1002/j.1460-2075 [26]

[253] William Rees, Rebecca Keller, James P. Vesenka, Guoliang Yang, and Carlos Bustamante. “Evidence of DNA Bending in Transcription Complexes Imaged by Scanning Force Microscopy”. *Science*, 1993.
DOI: 10.1126/science.8503010 [26]

[254] Tahir H. Tahirov, Dmitry Temiakov, Michael Anikin, Vsevolod Patlan, William T. McAllister, Dmitry G. Vassylyev, and Shigeyuki Yokoyama. “Structure of a T7 RNA Polymerase Elongation Complex at 2.9 Å Resolution”. *Nature*, 2002.
DOI: 10.1038/nature01129 [26]

[255] Y. Whitney Yin and Thomas A. Steitz. “Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase”. *Science*, 2002.
DOI: 10.1126/science.1077464 [26]

[256] Richard E. Dickerson. “DNA Bending: The Prevalence of Kinkiness and the Virtues of Normality”. *Nucleic Acids Research*, 1998.
DOI: 10.1093/nar/26.8.1906 [26]

[257] Aneel K. Aggarwal, David W. Rodgers, Marie Drottar, Mark Ptashne, and Stephen C. Harrison. “Recognition of a DNA Operator by the Repressor of Phage 434: A View at High Resolution”. *Science*, 1988.
DOI: 10.1126/science.3187531 [26]

[258] Andrew Travers. “DNA–Protein Interactions: IHF - the Master Bender”. *Current Biology*, 1997.
DOI: 10.1016/S0960-9822(06)00114-X [26]

[259] Kazutoshi Kasho, Hiroyuki Tanaka, Ryuji Sakai, and Tsutomu Katayama. “Cooperative DnaA Binding to the Negatively Supercoiled datA Locus Stimulates DnaA-ATP Hydrolysis”. *Journal of Biological Chemistry*, 2017.
DOI: 10.1074/jbc.M116.762815 [26]

[260] N. P. Higgins, D. A. Collier, M. W. Kilpatrick, and H. M. Krause. “Supercoiling and Integration Host Factor Change the DNA Conformation and Alter the Flow of Convergent Transcription in Phage Mu.” *The Journal of Biological Chemistry*, 1989. [26]

[261] Ronald Chalmers, Anjan Guhathakurta, Howard Benjamin, and Nancy Kleckner. “IHF Modulation of Tn10 Transposition: Sensory Transduction of Supercoiling Status via a Proposed Protein/DNA Molecular Spring”. *Cell*, 1998.
DOI: 10.1016/S0092-8674(00)81449-X [26]

[262] Rachel E. Ashley, Andrew Dittmore, Sylvia A. McPherson, Charles L. Turnbough,

Keir C. Neuman, and Neil Osheroff. “Activities of Gyrase and Topoisomerase IV on Positively Supercoiled DNA.” *Nucleic Acids Research*, 2017.
DOI: 10.1093/nar/gkx649 [27]

[263] Tamara R. Litwin, Maria Solà, Ian J. Holt, and Keir C. Neuman. “A Robust Assay to Measure DNA Topology-Dependent Protein Binding Affinity.” *Nucleic Acids Research*, 2015.
DOI: 10.1093/nar/gku1381 [27]

[264] Michael W. Van Dyke, Robert G. Roeder, and Michèle Sawadogo. “Physical Analysis of Transcription Preinitiation Complex Assembly on a Class II Gene Promoter”. *Science*, 1988.
DOI: 10.1126/science.3413495 [27]

[265] Stephen Buratowski, Steven Hahn, Leonard Guarente, and Phillip A. Sharp. “Five Intermediate Complexes in Transcription Initiation by RNA Polymerase II.” *Cell*, 1989.
DOI: 10.1016/0092-8674(89)90578-3 [27]

[266] Korbinian Liebl and Martin Zacharias. “How Global DNA Unwinding Causes Non-Uniform Stress Distribution and Melting of DNA.” *PLOS ONE*, 2020.
DOI: 10.1371/journal.pone.0232976 [27]

[267] Maxim D. Frank-Kamenetskii and Sergei M. Mirkin. “Triplex DNA Structures”. *Annual Review of Biochemistry*, 1995.
DOI: 10.1146/annurev.bi.64.070195.000433 [27]

[268] Pierre Vekhoff, Alexandre Ceccaldi, David Polverari, Jean Pylouster, Claudio Pisano, and Paola B. Arimondo. “Triplex Formation on DNA Targets: How To Choose the Oligonucleotide”. *Biochemistry*, 2008.
DOI: 10.1021/bi801087g [27]

[269] Sangjin Kim, Erik Broströmer, Dong Xing, Jianshi Jin, Shasha Chong, Hao Ge, Siyuan Wang, Chan Gu, Lijiang Yang, Yi Qin Gao, Xiao-dong Su, Yujie Sun, and X. Sunney Xie. “Probing Allostery Through DNA”. *Science*, 2013.
DOI: 10.1126/science.1229223 [27]

[270] Gabriel Rosenblum, Nadav Elad, Haim Rozenberg, Felix Wiggers, Jakub Jungwirth, and Hagen Hofmann. “Allostery through DNA Drives Phenotype Switching”. *Nature Communications*, 2021.
DOI: 10.1038/s41467-021-23148-2 [27]

[271] Joseph J. Traverso, Valipuram S. Manoranjan, A. R. Bishop, Kim Ø. Rasmussen, and Nikolaos K. Voulgarakis. “Allostery through Protein-Induced DNA Bubbles”. *Scientific Reports*, 2015.
DOI: 10.1038/srep09037 [27]

[272] Stephen P Hancock, Duilio Cascio, and Reid C Johnson. “Cooperative DNA Binding by Proteins through DNA Shape Complementarity”. *Nucleic Acids Research*, 2019.
DOI: 10.1093/nar/gkz642 [27]

[273] Jaspreet Singh and Prashant K. Purohit. “Elasticity as the Basis of Allostery in DNA”. *The Journal of Physical Chemistry B*, 2019.
DOI: 10.1021/acs.jpcb.8b07501 [27]

[274] Tomáš Dršata, Marie Zgarbová, Naďa Špačková, Petr Jurečka, Jiří Šponer, and Filip Lankaš. "Mechanical Model of DNA Allostery". *The Journal of Physical Chemistry Letters*, 2014.
DOI: 10.1021/jz501826q [27]

[275] Xinliang Xu, Hao Ge, Chan Gu, Yi Qin Gao, Siyuan S. Wang, Beng Joo Reginald Thio, James T. Hynes, X. Sunney Xie, and Jianshu Cao. "Modeling Spatial Correlation of DNA Deformation: DNA Allostery in Protein Binding". *The Journal of Physical Chemistry B*, 2013.
DOI: 10.1021/jp4047243 [27]

[276] Nicolas Clauvelin and Wilma K. Olson. "Synergy between Protein Positioning and DNA Elasticity: Energy Minimization of Protein-Decorated DNA Minicircles". *The Journal of Physical Chemistry B*, 2021.
DOI: 10.1021/acs.jpcb.0c11612 [27]

[277] Tahir H. Tahirov, Ko Sato, Emi Ichikawa-Iwata, Motoko Sasaki, Taiko Inoue-Bungo, Masaaki Shiina, Kazumi Kimura, Shioka Takata, Atsushi Fujikawa, Hisayuki Morii, Takashi Kumasaka, Masaki Yamamoto, Shunsuke Ishii, and Kazuhiro Ogata. "Mechanism of C-Myb–C/EBPβ Cooperation from Separated Sites on a Promoter". *Cell*, 2002.
DOI: 10.1016/S0092-8674(01)00636-5 [27]

[278] Jonathan M. Fogg, Allison K. Judge, Erik Stricker, Hilda L. Chan, and Lynn Zechiedrich. "Supercoiling and Looping Promote DNA Base Accessibility and Coordination among Distant Sites". *Nature Communications*, 2021.
DOI: 10.1038/s41467-021-25936-2 [27]

[279] Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith, and Andrej Kosmrlj. "How a Protein Searches for Its Site on DNA: The Mechanism of Facilitated Diffusion". *Journal of Physics A: Mathematical and Theoretical*, 2009.
DOI: 10.1088/1751-8113/42/43/434013 [27, 28]

[280] Arthur D. Riggs, Suzanne Bourgeois, and Melvin Cohn. "The Lac Represser-Operator Interaction: III. Kinetic Studies". *Journal of Molecular Biology*, 1970.
DOI: 10.1016/0022-2836(70)90074-4 [27]

[281] Alexey A. Shvets, Maria P. Kochugaeva, and Anatoly B. Kolomeisky. "Mechanisms of Protein Search for Targets on DNA: Theoretical Insights." *Molecules (Basel, Switzerland)*, 2018.
DOI: 10.3390/molecules23092106 [28]

[282] Tao Hu, A. Yu Grosberg, and B. I. Shklovskii. "How Proteins Search for Their Specific Sites on DNA: The Role of DNA Conformation." *Biophysical Journal*, 2006.
DOI: 10.1529/biophysj.105.078162 [28]

[283] Chris A. Brackley, Mike E. Cates, and Davide Marenduzzo. "Effect of DNA Conformation on Facilitated Diffusion". *Biochemical Society Transactions*, 2013.
DOI: 10.1042/BST20120234 [28]

[284] Galina Obmolova, Changill Ban, Peggy Hsieh, and Wei Yang. "Crystal Structures of Mismatch Repair Protein MutS and Its Complex with a Substrate DNA". *Nature*, 2000.

DOI: 10.1038/35037509 [28]

[285] Meindert H. Lamers, Anastassis Perrakis, Jacqueline H. Enzlin, Herrie H. K. Winterwerp, Niels de Wind, and Titia K. Sixma. "The Crystal Structure of DNA Mismatch Repair Protein MutS Binding to a G·T Mismatch". *Nature*, 2000.
DOI: 10.1038/35037523 [28, 108]

[286] Yaping Huang and Guo-Min Li. "DNA Mismatch Repair Preferentially Safeguards Actively Transcribed Genes". *Cutting-edge Perspectives in Genomic Maintenance V*, 2018.
DOI: 10.1016/j.dnarep.2018.08.010 [28]

[287] Andrew Dittmore, Sumitabha Brahmachari, Yasuharu Takagi, John F. Marko, and Keir C. Neuman. "Supercoiling DNA Locates Mismatches". *Physical Review Letters*, 2017.
DOI: 10.1103/PhysRevLett.119.147801 [28, 105, 106, 136, 145, 152]

[288] Parth Rakesh Desai, Sumitabha Brahmachari, John F Marko, Siddhartha Das, and Keir C Neuman. "Coarse-Grained Modelling of DNA Plectoneme Pinning in the Presence of Base-Pair Mismatches". *Nucleic Acids Research*, 2020.
DOI: 10.1093/nar/gkaa836 [28, 32, 33, 62, 80, 106, 132, 135, 136, 141, 145, 147, 152]

[289] Huijin Lee, Jihee Hwang, Fahad Rashid, James A. London, Richard Fishel, James M. Berger, Sua Myong, and Taekjip Ha. "A High Throughput Single Molecule Platform to Study DNA Supercoiling Effect on Protein-DNA Interactions." *bioRxiv*, 2024.
DOI: 10.1101/2024.10.24.620099 [28, 105, 106]

[290] Wilber Lim, Ferdinando Randisi, Jonathan P. K. Doye, and Ard A. Louis. "The Interplay of Supercoiling and Thymine Dimers in DNA." *Nucleic Acids Research*, 2022.
DOI: 10.1093/nar/gkac082 [29, 62, 106, 152]

[291] Kristina A Makasheva, Anton V Endutkin, and Dmitry O Zharkov. "Requirements for DNA Bubble Structure for Efficient Cleavage by Helix–Two-Turn–Helix DNA Glycosylases". *Mutagenesis*, 2020.
DOI: 10.1093/mutage/gez047 [29]

[292] Bryant C Nelson and Miral Dizdaroglu. "Implications of DNA Damage and DNA Repair on Human Diseases". *Mutagenesis*, 2020.
DOI: 10.1093/mutage/gez048 [29]

[293] Hong Dou, Sankar Mitra, and Tapas K. Hazra. "Repair of Oxidized Bases in DNA Bubble Structures by Human DNA Glycosylases NEIL1 and NEIL2". *Journal of Biological Chemistry*, 2003.
DOI: 10.1074/jbc.M308658200 [29]

[294] Yu-Yuan Hsiao, Woei-Horng Fang, Chia-Chia Lee, Yi-Ping Chen, and Hanna S. Yuan. "Structural Insights Into DNA Repair by RNase T—An Exonuclease Processing 3′ End of Structured DNA in Repair Pathways". *PLOS Biology*, 2014.
DOI: 10.1371/journal.pbio.1001803 [29]

[295] Richard J. Isaacs and H.Peter Spielmann. "A Model for Initial DNA Lesion Recognition by NER and MMR Based on Local Conformational Flexibility". *DNA Repair*, 2004.

DOI: 10.1016/j.dnarep.2004.01.004 [29, 103, 105]

[296] Joke Baute and Anne Depicker. "Base Excision Repair and Its Role in Maintaining Genome Stability". *Critical Reviews in Biochemistry and Molecular Biology*, 2008.
DOI: 10.1080/10409230802309905 [29]

[297] Mengtian Ren, Fabian Gut, Yilan Fan, Jingke Ma, Xiajing Shan, Aysenur Yikilmazsoy, Mariia Likhodeeva, Karl-Peter Hopfner, and Chuanzheng Zhou. "Structural Basis for Human OGG1 Processing 8-oxodGuo within Nucleosome Core Particles." *Nature communications*, 0031.
DOI: 10.1038/s41467-024-53811-3 [29]

[298] Xiaodong Cheng and Richard J Roberts. "Base Flipping". IN: *eLS*. 2014.
DOI: 10.1002/9780470015902.a0002714.pub3 [29]

[299] J. S. Mitchell, C. A. Laughton, and Sarah A. Harris. "Atomistic Simulations Reveal Bubbles, Kinks and Wrinkles in Supercoiled DNA". *Nucleic Acids Research*, 2011.
DOI: 10.1093/nar/gkq1312 [29, 33, 105, 150]

[300] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo. "Stochastic Model of Supercoiling-Dependent Transcription". *Physical Review Letters*, 2016.
DOI: 10.1103/PhysRevLett.117.018101 [29]

[301] Nick Gilbert and Davide Marenduzzo. "Topological Epigenetics: The Biophysics of DNA Supercoiling and Its Relation to Transcription and Genome Instability". *Current Opinion in Cell Biology*, 2025.
DOI: 10.1016/j.ceb.2024.102448 [29, 30]

[302] Porter M. Hall, Lauren A. Mayse, Lu Bai, Marcus B. Smolka, B. Franklin Pugh, and Michelle D. Wang. "High-Resolution Genome-Wide Maps Reveal Widespread Presence of Torsional Insulation". *bioRxiv*, 2025.
DOI: 10.1101/2024.10.11.617876 [30]

[303] Quan Du, Chaim Smith, Nahum Shiffeldrim, Maria Vologodskaia, and Alexander Vologodskii. "Cyclization of Short DNA Fragments and Bending Fluctuations of the Double Helix". *Proceedings of the National Academy of Sciences*, 2005.
DOI: 10.1073/pnas.0500983102 [31]

[304] David Sivak and Phillip Geissler. "Consequences of Local Inter-Strand Dehybridization for Large-Amplitude Bending Fluctuations of Double-Stranded DNA". *The Journal of Chemical Physics*, 2012.
DOI: 10.1063/1.3679654 [31]

[305] T. E. Cloutier and J. Widom. "DNA Twisting Flexibility and the Formation of Sharply Looped Protein–DNA Complexes". *Proceedings of the National Academy of Sciences*, 2005.
DOI: 10.1073/pnas.0409059102 [31]

[306] Alexander Vologodskii and Maxim D. Frank-Kamenetskii. "Strong Bending of the DNA Double Helix". *Nucleic Acids Research*, 2013.
DOI: 10.1093/nar/gkt396 [31]

[307] Alexey K. Mazur and Mounir Maaloum. "DNA Flexibility on Short Length Scales Probed by Atomic Force Microscopy". *Physical Review Letters*, 2014.
DOI: 10.1103/PhysRevLett.112.068104 [31]

[308] Reza Vafabakhsh and Taekjip Ha. "Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization". *Science*, 2012.
DOI: 10.1126/science.1224139 [31]

[309] Paul A. Wiggins, Thijn van der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C. Nelson. "High Flexibility of DNA on Short Length Scales Probed by Atomic Force Microscopy". *Nature Nanotechnology*, 2006.
DOI: 10.1038/nnano.2006.63 [31]

[310] Luke Czapla, David Swigon, and Wilma K. Olson. "Sequence-Dependent Effects in the Cyclization of Short DNA". *Journal of Chemical Theory and Computation*, 2006.
DOI: 10.1021/ct060025+ [31]

[311] Tung T. Le and Harold D. Kim. "Probing the Elastic Limit of DNA Bending". *Nucleic Acids Research*, 2014.
DOI: 10.1093/nar/gku735 [31]

[312] José Pérez-Martín and VíctorYR 1997 de Lorenzo. "Clues and Consequences of DNA Bending in Transcription". *Annual Review of Microbiology,*
DOI: 10.1146/annurev.micro.51.1.593 [31]

[313] Hernan G. Garcia, Paul Grayson, Lin Han, Mandar Inamdar, Jané Kondev, Philip C. Nelson, Rob Phillips, Jonathan Widom, and Paul A. Wiggins. "Biological Consequences of Tightly Bent DNA: The Other Life of a Macromolecular Celebrity". *Biopolymers*, 2007.
DOI: 10.1002/bip.20627 [31]

[314] Timothy J. Richmond and Curt A. Davey. "The Structure of DNA in the Nucleosome Core". *Nature*, 2003.
DOI: 10.1038/nature01595 [31]

[315] Jie Yan and John F. Marko. "Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA". *Physical Review Letters*, 2004.
DOI: 10.1103/PhysRevLett.93.108108 [31]

[316] Paul A. Wiggins, Rob Phillips, and Philip C. Nelson. "Exact Theory of Kinkable Elastic Polymers". *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2005.
DOI: 10.1103/PhysRevE.71.021909 [31]

[317] Jie Yan, Ryo Kawamura, and John F. Marko. "Statistics of Loop Formation along Double Helix DNAs". *Physical Review E*, 2005.
DOI: 10.1103/PhysRevE.71.061905 [31]

[318] Jejoong Yoo, Sangwoo Park, Christopher Maffeo, Taekjip Ha, and Aleksei Aksimentiev. "DNA Sequence and Methylation Prescribe the Inside-out Conformational Dynamics and Bending Energetics of DNA Minicircles." *Nucleic Acids Research*, 2021.
DOI: 10.1093/nar/gkab967 [32, 33]

[319] Peter Gross, Niels Laurens, Lene B. Oddershede, Ulrich Bockelmann, Erwin J. G. Peterman, and Gijs J. L. Wuite. "Quantifying How DNA Stretches, Melts and Changes Twist under Tension". *Nature Physics*, 2011.
DOI: 10.1038/nphys2002 [32]

[320] Theo Odijk. "Stiff Chains and Filaments under Tension". *Macromolecules*, 1995.
DOI: 10.1021/ma00124a044 [32]

[321] John F. Marko and Eric D. Siggia. "Stretching DNA". *Macromolecules*, 1995.
DOI: 10.1021/ma00130a008 [32]

[322] Esmaeil Farshi. *Navigating Genetic Landscapes: A Novel Intersection of Quantum Tunneling, DNA Supercoiling, and Spontaneous Mutations*. 2023.
DOI: 10.13140/RG.2.2.34431.38569 [32]

[323] Yair A. G. Fosado, Davide Michieletto, Chris A. Brackley, and Davide Marenduzzo. "Nonequilibrium Dynamics and Action at a Distance in Transcriptionally Driven DNA Supercoiling". *Proceedings of the National Academy of Sciences of the United States of America*, 2021.
DOI: 10.1073/pnas.1905215118 [32, 60]

[324] David Farré-Gil, Juan Pablo Arcon, Charles A. Laughton, and Modesto Orozco. "CGeNArate: A Sequence-Dependent Coarse-Grained Model of DNA for Accurate Atomistic MD Simulations of Kb-Long Duplexes." *Nucleic Acids Research*, 2024.
DOI: 10.1093/nar/gkae444 [33, 60]

[325] Enrico Skoruppa and Helmut Schiessel. "Systematic Coarse-Graining of Sequence-Dependent Structure and Elasticity of Double-Stranded DNA". *Physical Review Research*, 2025.
DOI: 10.1103/PhysRevResearch.7.013044 [33]

[326] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. "Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model". *The Journal of Chemical Physics*, 2011.
DOI: 10.1063/1.3552946 [33]

[327] Benedict E. K. Snodin, Ferdinando Randisi, Majid Mosayebi, Petr Šulc, John S. Schreck, Flavio Romano, Thomas E. Ouldridge, Roman Tsukanov, Eyal Nir, Ard A. Louis, and Jonathan P. K. Doye. "Introducing Improved Structural Properties and Salt Dependence into a Coarse-Grained Model of DNA". *The Journal of Chemical Physics*, 2015.
DOI: 10.1063/1.4921957 [33]

[328] Gyehyun Park, Myung Keun Cho, and YounJoon Jung. "Sequence-Dependent Kink Formation in Short DNA Loops: Theory and Molecular Dynamics Simulations". *Journal of Chemical Theory and Computation*, 2021.
DOI: 10.1021/acs.jctc.0c01116 [33]

[329] Aditya Sengar, Thomas E. Ouldridge, Oliver Henrich, Lorenzo Rovigatti, and Petr Šulc. "A Primer on the oxDNA Model of DNA: When to Use It, How to Simulate It and How to Interpret the Results". *Frontiers in Molecular Biosciences*, 2021.
DOI: 10.3389/fmolb.2021.693710 [33]

[330] Filip Lankaš, Richard Lavery, and John H. Maddocks. "Kinking Occurs during Molecular Dynamics Simulations of Small DNA Minicircles". *Structure*, 2006.
DOI: 10.1016/j.str.2006.08.004 [33, 105, 150]

[331] Korbinian Liebl, Tomas Drsata, Filip Lankas, Jan Lipfert, and Martin Zacharias. "Explaining the Striking Difference in Twist-Stretch Coupling between DNA and RNA: A Comparative Molecular Dynamics Analysis". *Nucleic Acids Research*, 2015.
DOI: 10.1093/nar/gkv1028 [33]

[332] Minjung Kim, Sehui Bae, Inrok Oh, Jejoong Yoo, and Jun Soo Kim. "Sequence-Dependent Twist-Bend Coupling in DNA Minicircles". *Nanoscale*, 2021.
DOI: 10.1039/D1NR04672A [33]

[333] Christopher Maffeo, Robert Schöpflin, Hergen Brutzer, René Stehr, Aleksei Aksimentiev, Gero Wedemann, and Ralf Seidel. "DNA–DNA Interactions in Tight Supercoils Are Described by a Small Effective Charge Density". *Physical Review Letters*, 2010.
DOI: 10.1103/PhysRevLett.105.158101 [33]

[334] Xipeng Wang, Simón Ramírez-Hinestrosa, Jure Dobnikar, and Daan Frenkel. "The Lennard-Jones Potential: When (Not) to Use It". *Physical Chemistry Chemical Physics*, 2020.
DOI: 10.1039/C9CP05445F [37]

[335] Johann Fischer and Martin Wendland. "On the History of Key Empirical Intermolecular Potentials". *Fluid Phase Equilibria*, 2023.
DOI: 10.1016/j.fluid.2023.113876 [38]

[336] Hendrik A. Lorentz. "Ueber Die Anwendung Des Satzes Vom Virial in Der Kinetischen Theorie Der Gase". *Annalen der Physik*, 1881.
DOI: 10.1002/andp.18812480110 [38]

[337] Daniel Berthelot. "Sur Le Mélange Des Gaz". *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 1898.
[38]

[338] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *Journal of the American Chemical Society*, 1995.
DOI: 10.1021/ja00124a002 [38]

[339] Michael R. Shirts, Christoph Klein, Jason M. Swails, Jian Yin, Michael K. Gilson, David L. Mobley, David A. Case, and Ellen D. Zhong. "Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset". *Journal of Computer-Aided Molecular Design*, 2017.
DOI: 10.1007/s10822-016-9977-1 [38]

[340] Tom Darden, Darrin York, and Lee Pedersen. "Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems". *The Journal of Chemical Physics*, 1993.
DOI: 10.1063/1.464397 [38]

[341] Vickie Tsui. *9.1 NMR Refinement of DNA and RNA Duplexes.* https://ambermd.org/tutorials/advanced/tutorial4/index.php. (Visited on 07/11/2025) [39]

[342] Oliver T. Unke, Stefan Chmiela, Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. "Machine Learning Force Fields". *Chemical Reviews*, 2021. DOI: 10.1021/acs.chemrev.0c01111 [40]

[343] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E. Cheatham III, Charles A. Laughton, and Modesto Orozco. "Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers". *Biophysical Journal*, 2007. DOI: 10.1529/biophysj.106.097782 [40]

[344] Ivan Ivani et al. "Parmbsc1: A Refined Force Field for DNA Simulations". *Nature Methods*, 2016. DOI: 10.1038/nmeth.3658 [40, 61, 64, 147, 148]

[345] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes". *Journal of Computational Physics*, 1977. DOI: 10.1016/0021-9991(77)90098-5 [42]

[346] Shuichi Miyamoto and Peter A. Kollman. "Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models". *Journal of Computational Chemistry*, 1992. DOI: 10.1002/jcc.540130805 [42]

[347] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations". *Journal of Computational Chemistry*, 1997. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H [42]

[348] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water". *The Journal of Chemical Physics*, 1983. DOI: 10.1063/1.445869 [43]

[349] Kahled Toukan and Aneesur Rahman. "Molecular-Dynamics Study of Atomic Motions in Water". *Physical Review B*, 1985. DOI: 10.1103/PhysRevB.31.2643 [43]

[350] Sachini P. Kadaoluwa Pathirannahalage, Nastaran Meftahi, Aaron Elbourne, Alessia C. G. Weiss, Chris F. McConville, Agilio Padua, David A. Winkler, Margarida Costa Gomes, Tamar L. Greaves, Tu C. Le, Quinn A. Besford, and Andrew J. Christofferson. "Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations". *Journal of Chemical Information and Modeling*, 2021. DOI: 10.1021/acs.jcim.1c00794 [43]

[351] Yeyue Xiong, Saeed Izadi, and Alexey V. Onufriev. "Fast Polarizable Water Model for Atomistic Simulations". *Journal of Chemical Theory and Computation*, 2022.
DOI: 10.1021/acs.jctc.2c00378 [43]

[352] Aldi Asmadi, Tom Kirchner, Wael Abdallah, Maxim V. Fedorov, and Mikhail R. Stukan. "Influence of the Drude Charge Value on the Performance of Polarisable Water Model: A Test for Microscopic and Macroscopic Parameters". *Journal of Molecular Liquids*, 2013.
DOI: 10.1016/j.molliq.2013.09.026 [43]

[353] Herman J. C. Berendsen, J. Raul Grigera, and Tjerk P. Straatsma. "The Missing Term in Effective Pair Potentials". *The Journal of Physical Chemistry*, 1987.
DOI: 10.1021/j100308a038 [43]

[354] Jun Wang, Piotr Cieplak, Qin Cai, Meng-Juei Hsieh, Junmei Wang, Yong Duan, and Ray Luo. "Development of Polarizable Models for Molecular Mechanical Calculations. 3. Polarizable Water Models Conforming to Thole Polarization Screening Schemes." *The Journal of Physical Chemistry B*, 2012.
DOI: 10.1021/jp212117d [43]

[355] Laurent David, Ray Luo, and Michael K. Gilson. "Comparison of Generalized Born and Poisson Models: Energetics and Dynamics of HIV Protease". *Journal of Computational Chemistry*, 2000.
DOI: 10.1002/(SICI)1096-987X(200003)21:4<295::AID-JCC5>3.0.CO;2-8 [44]

[356] Michael Feig. "Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity". *Journal of Chemical Theory and Computation*, 2007.
DOI: 10.1021/ct7000705 [44]

[357] Bojan Zagrovic and Vijay Pande. "Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study". *Journal of Computational Chemistry*, 2003.
DOI: 10.1002/jcc.10297 [44]

[358] Ramu Anandakrishnan, Aleksander Drozdetski, Ross C. Walker, and Alexey V. Onufriev. "Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations." *Biophysical Journal*, 2015.
DOI: 10.1016/j.bpj.2014.12.047 [44, 61]

[359] Alexey V. Onufriev and David A. Case. "Generalized Born Implicit Solvent Models for Biomolecules." *Annual Review of Biophysics*, 2019.
DOI: 10.1146/annurev-biophys-052118-115325 [44, 45]

[360] Frederic M. Richards. "Areas, Volumes, Packing, and Protein Structure". *Annual Review of Biophysics*, 1977.
DOI: 10.1146/annurev.bb.06.060177.001055 [44]

[361] Jörg Weiser, Peter S. Shenkin, and W. Clark Still. "Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO)". *Journal of Computational Chemistry*, 1999.
DOI: 10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A [45, 46]

[362] Hai Nguyen, Alberto Pérez, Sherry Bermeo, and Carlos Simmerling. "Refinement of

Generalized Born Implicit Solvation Parameters for Nucleic Acids and Their Complexes with Proteins". *Journal of Chemical Theory and Computation*, 2015.
DOI: 10.1021/acs.jctc.5b00271 [45, 46, 64]

[363] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. "Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics". *Journal of the American Chemical Society*, 1990.
DOI: 10.1021/ja00172a038 [45]

[364] Jayashree Srinivasan, Megan W. Trevathan, Paul Beroza, and David A. Case. "Application of a Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects". *Theoretical Chemistry Accounts*, 1999.
DOI: 10.1007/s002140050460 [45, 60]

[365] Alexey Onufriev, David A. Case, and Donald Bashford. "Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect". *Journal of Computational Chemistry*, 2002.
DOI: 10.1002/jcc.10126 [45]

[366] B. Jayaram, Dennis Sprous, and David L. Beveridge. "Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field". *The Journal of Physical Chemistry B*, 1998.
DOI: 10.1021/jp982007x [45]

[367] Alexey Onufriev, Donald Bashford, and David A. Case. "Modification of the Generalized Born Model Suitable for Macromolecules". *The Journal of Physical Chemistry B*, 2000.
DOI: 10.1021/jp994072s [45, 46]

[368] John Mongan, W. Andreas Svrcek-Seiler, and Alexey Onufriev. "Analysis of Integral Expressions for Effective Born Radii". *The Journal of Chemical Physics*, 2007.
DOI: 10.1063/1.2783847 [46]

[369] Tomasz Grycuk. "Deficiency of the Coulomb-field Approximation in the Generalized Born Model: An Improved Formula for Born Radii Evaluation". *The Journal of Chemical Physics*, 2003.
DOI: 10.1063/1.1595641 [46]

[370] John Mongan, Carlos Simmerling, J. Andrew McCammon, David A. Case, and Alexey Onufriev. "Generalized Born Model with a Simple, Robust Molecular Volume Correction". *Journal of Chemical Theory and Computation*, 2007.
DOI: 10.1021/ct600085e [46, 64]

[371] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular Dynamics with Coupling to an External Bath". *The Journal of Chemical Physics*, 1984.
DOI: 10.1063/1.448118 [47, 49]

[372] Tetsuya Morishita. "Fluctuation Formulas in Molecular-Dynamics Simulations with the Weak Coupling Heat Bath". *The Journal of Chemical Physics*, 2000.
DOI: 10.1063/1.1287333 [48]

[373] Don S. Lemons and Anthony Gythiel. "Paul Langevin's 1908 Paper "On the Theory of

Brownian Motion" ["Sur La Théorie Du Mouvement Brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]". *American Journal of Physics*, 1997.
DOI: 10.1119/1.18725 [48]

[374] M. Paul Langevin. "Sur La Théorie Du Mouvement Brownien". *Comptes-rendus de l'Académie des sciences*, 1908.
[48]

[375] Johan Åqvist, Petra Wennerström, Martin Nervall, Sinisa Bjelic, and Bjørn O. Brandsdal. "Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm". *Chemical Physics Letters*, 2004.
DOI: 10.1016/j.cplett.2003.12.039 [49]

[376] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.
DOI: 10.1093/oso/9780198803195.001.0001 [49]

[377] Eitan Lerner et al. "FRET-based Dynamic Structural Biology: Challenges, Perspectives and an Appeal for Open-Science Practices". *eLife*, 2021. ed. by Olga Boudker.
DOI: 10.7554/eLife.60416 [49]

[378] Antonino Ingargiola, Eitan Lerner, SangYoon Chung, Shimon Weiss, and Xavier Michalet. "FRETBursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET". *PLOS ONE*, 2016.
DOI: 10.1371/journal.pone.0160716 [50, 52, 68]

[379] Björn Hellenkamp et al. "Precision and Accuracy of Single-Molecule FRET Measurements—a Multi-Laboratory Benchmark Study". *Nature Methods*, 2018.
DOI: 10.1038/s41592-018-0085-0 [50, 53, 67]

[380] Simon Sindbert, Stanislav Kalinin, Hien Nguyen, Andrea Kienzler, Lilia Clima, Willi Bannwarth, Bettina Appel, Sabine Müller, and Claus A. M. Seidel. "Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity". *Journal of the American Chemical Society*, 2011.
DOI: 10.1021/ja105725e [52, 53, 54]

[381] Theodor Förster. "Zwischenmolekulare Energiewanderung Und Fluoreszenz". *Annalen der Physik*, 1948.
DOI: 10.1002/andp.19484370105 [52]

[382] B. Wieb van der Meer, George Coker, and S.-Y. Simon Chen. *Resonance Energy Transfer: Theory and Data*. Wiley-Blackwell, 1994.
[52]

[383] Gregory D. Scholes. "Long-Range Resonance Energy Transfer in Molecular Systems". *Annual Review of Physical Chemistry*, 2003.
DOI: 10.1146/annurev.physchem.54.011002.103746 [52]

[384] Bernard Valeur. "Molecular Fluorescence". IN: *Encyclopedia of Applied Physics*. 2009.
DOI: 10.1002/3527600434.eap684 [52]

[385] Stanislav Kalinin, Thomas Peulen, Simon Sindbert, Paul J. Rothwell, Sylvia Berger,

Tobias Restle, Roger S. Goody, Holger Gohlke, and Claus A. M. Seidel. "A Toolkit and Benchmark Study for FRET-restrained High-Precision Structural Modeling". *Nature Methods*, 2012.
DOI: 10.1038/nmeth.2222 [52, 53, 54]

[386] Mykola Dimura, Thomas-Otavio Peulen, Hugo Sanabria, Dmitro Rodnin, Katherina Hemmen, Christian A. Hanke, Claus A. M. Seidel, and Holger Gohlke. "Automated and Optimally FRET-assisted Structural Modeling". *Nature Communications*, 2020.
DOI: 10.1038/s41467-020-19023-1 [54, 67]

[387] Daniel Jost. "Twist-DNA: Computing Base-Pair and Bubble Opening Probabilities in Genomic Superhelical DNA". *Bioinformatics*, 2013.
DOI: 10.1093/bioinformatics/btt415 [54, 56, 67, 69]

[388] Daniel Jost, Asif Zubair, and Ralf Everaers. "Bubble Statistics and Positioning in Superhelically Stressed DNA". *Physical Review E*, 2011.
DOI: 10.1103/PhysRevE.84.031912 [54, 56, 67, 69]

[389] A.V. Vologodskii, A.V. Lukashin, V.V. Anshelevich, and M.D. Frank-Kamenetskii. "Fluctuations in Superhelical DNA". *Nucleic Acids Research*, 1979.
DOI: 10.1093/nar/6.3.967 [54]

[390] Craig J. Benham. "Energetics of the Strand Separation Transition in Superhelical DNA". *Journal of Molecular Biology*, 1992.
DOI: 10.1016/0022-2836(92)90404-8 [54]

[391] Thomas Quail, Stefan Golfier, Maria Elsner, Keisuke Ishihara, Vasanthanarayan Murugesan, Roman Renger, Frank Jülicher, and Jan Brugués. "Force Generation by Protein–DNA Co-Condensation". *Nature Physics*, 2021.
DOI: 10.1038/s41567-021-01285-1 [58]

[392] Sreekala Balasubramanian, Fei Xu, and Wilma K. Olson. "DNA Sequence-Directed Organization of Chromatin: Structure-Based Computational Analysis of Nucleosome-Binding Sequences". *Biophysical Journal*, 2009.
DOI: 10.1016/j.bpj.2008.11.040 [58]

[393] Thana Sutthibutpong, Agnes Noy, and Sarah Harris. "Atomistic Molecular Dynamics Simulations of DNA Minicircle Topoisomers: A Practical Guide to Setup, Performance, and Analysis". IN: *Chromosome Architecture: Methods and Protocols*. ed. by Mark C. Leake. Springer New York, 2016.
DOI: 10.1007/978-1-4939-3631-1_15 [60, 61, 65, 69]

[394] Thana Sutthibutpong. "Molecular Dynamics Study of Supercoiled DNA Minicircles Tightly Bent and Supercoiled DNA in Atomistic Resolution - White Rose eTheses Online". PhD thesis. University of Leeds, 2015.
(Visited on 11/28/2020) [60]

[395] Jonathan M. Fogg, Natalia Kolmakova, Ian Rees, Sergei Magonov, Helen Hansma, John J. Perona, and E. Lynn Zechiedrich. "Exploring Writhe in Supercoiled Minicircle DNA". *Journal of Physics. Condensed Matter: An Institute of Physics Journal*, 2006.
DOI: 10.1088/0953-8984/18/14/S01 [60]

[396] Beata Klejevskaja, Alice L. B. Pyne, Matthew Reynolds, Arun Shivalingam, Richard Thorogate, Bart W. Hoogenboom, Liming Ying, and Ramon Vilar. "Studies of G-quadruplexes Formed within Self-Assembled DNA Mini-Circles". *Chemical Communications*, 2016.
DOI: 10.1039/C6CC07110D [60]

[397] Julián Valero, Nibedita Pal, Soma Dhakal, Nils G. Walter, and Michael Famulok. "A Bio-Hybrid DNA Rotor–Stator Nanoengine That Moves along Predefined Tracks". *Nature Nanotechnology*, 2018.
DOI: 10.1038/s41565-018-0109-z [60]

[398] Sean D Colloms, Jonathan Bath, and David J Sherratt. "Topological Selectivity in Xer Site-Specific Recombination". *Cell*, 1997.
DOI: 10.1016/S0092-8674(00)81931-5 [60]

[399] Troy A. Lionberger, Davide Demurtas, Guillaume Witz, Julien Dorier, Todd D. Lillian, Edgar Meyhofer, and Andrzej Stasiak. "Cooperative Kinking at Distant Sites in Mechanically Stressed DNA". *Nucleic Acids Research*, 2011.
DOI: 10.1093/nar/gkr666 [60, 105]

[400] Hai Nguyen, Daniel R. Roe, and Carlos Simmerling. "Improved Generalized Born Solvent Model Parameters for Protein Simulations". *Journal of Chemical Theory and Computation*, 2013.
DOI: 10.1021/ct3010485 [60, 64]

[401] Saeed Izadi, Ramu Anandakrishnan, and Alexey V. Onufriev. "Implicit Solvent Model for Million-Atom Atomistic Simulations: Insights into the Organization of 30-Nm Chromatin Fiber." *Journal of Chemical Theory and Computation*, 2016.
DOI: 10.1021/acs.jctc.6b00712 [60]

[402] Ruhong Zhou. "Free Energy Landscape of Protein Folding in Water: Explicit vs. Implicit Solvent". *Proteins: Structure, Function, and Bioinformatics*, 2003.
DOI: 10.1002/prot.10483 [61]

[403] Carlos Roldán-Piñero, Juan Luengo-Márquez, Salvatore Assenza, and Rubén Pérez. "Systematic Comparison of Atomistic Force Fields for the Mechanical Properties of Double-Stranded DNA." *Journal of Chemical Theory and Computation*, 2024.
DOI: 10.1021/acs.jctc.3c01089 [61]

[404] Franziska Kriegel, Christian Matek, Tomáš Geršata, Klara Kulenkampff, Sophie Tschirpke, Martin Zacharias, Filip Lankaš, and Jan Lipfert. "The Temperature Dependence of the Helical Twist of DNA". *Nucleic Acids Research*, 2018.
DOI: 10.1093/nar/gky599 [63, 87]

[405] Hana Dohnalová, Eva Matoušková, and Filip Lankaš. "Temperature-Dependent Elasticity of DNA, RNA, and Hybrid Double Helices." *Biophysical Journal*, 2024.
DOI: 10.1016/j.bpj.2024.01.032 [63]

[406] Hana Dohnalová, Tomáš Geršata, Jiří Šponer, Martin Zacharias, Jan Lipfert, and Filip Lankaš. "Compensatory Mechanisms in Temperature Dependence of DNA Double Helical Structure: Bending and Elongation". *Journal of Chemical Theory and Computation*, 2020.

DOI: 10.1021/acs.jctc.0c00037 [63]

[407] Valentin V. Rybenkov, Alexander V. Vologodskii, and Nicholas R. Cozzarelli. "The Effect of Ionic Conditions on DNA Helical Repeat, Effective Diameter and Free Energy of Supercoiling". *Nucleic Acids Research*, 1997.
DOI: 10.1093/nar/25.7.1412 [63]

[408] Biao Wan and Jin Yu. "Two-Phase Dynamics of DNA Supercoiling Based on DNA Polymer Physics." *Biophysical Journal*, 2022.
DOI: 10.1016/j.bpj.2022.01.001 [63]

[409] John F. Marko and Sébastien Neukirch. "Competition between Curls and Plectonemes near the Buckling Transition of Stretched Supercoiled DNA". *Physical Review E*, 2012.
DOI: 10.1103/PhysRevE.85.011908 [63, 147]

[410] Franziska Kriegel, Niklas Ermann, Ruaridh Forbes, David Dulin, Nynke H. Dekker, and Jan Lipfert. "Probing the Salt Dependence of the Torsional Stiffness of DNA by Multiplexed Magnetic Torque Tweezers". *Nucleic Acids Research*, 2017.
DOI: 10.1093/nar/gkx280 [63]

[411] Jae-Hyung Jeon, Jozef Adamcik, Giovanni Dietler, and Ralf Metzler. "Supercoiling Induces Denaturation Bubbles in Circular DNA". *Physical Review Letters*, 2010.
DOI: 10.1103/PhysRevLett.105.208101 [63]

[412] David A. Case et al. "AmberTools". *Journal of Chemical Information and Modeling*, 2023.
DOI: 10.1021/acs.jcim.3c01153 [64, 65, 66, 112]

[413] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. "Pairwise Solute Descreening of Solute Charges from a Dielectric Medium". *Chemical Physics Letters*, 1995.
DOI: 10.1016/0009-2614(95)01082-K [64]

[414] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. "Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium". *The Journal of Physical Chemistry*, 1996.
DOI: 10.1021/jp961710n [64]

[415] V. Tsui and D. A. Case. "Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations". *Biopolymers*, 2000–2001.
DOI: 10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E [64]

[416] David A. Case et al. *Amber 2021*. 2021.
(Visited on 03/14/2025) [64, 65, 112]

[417] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water". *The Journal of Chemical Physics*, 1983.
DOI: 10.1063/1.445869 [65, 112]

[418] Thana Sutthibutpong, Sarah A. Harris, and Agnes Noy. "Comparison of Molecular Contours for Measuring Writhe in Atomistic Supercoiled DNA". *Journal of Chemical Theory and Computation*, 2015.

DOI: 10.1021/acs.jctc.5b00035 [66, 111, 112]

[419] Pablo Daniel Dans, Ignacio Faustino, Federica Battistini, Krystyna Zakrzewska, Richard Lavery, and Modesto Orozco. "Unraveling the Sequence-Dependent Polymorphic Behavior of d(CpG) Steps in B-DNA". *Nucleic Acids Research*, 2014.
DOI: 10.1093/nar/gku809 [66]

[420] Edgar Brunner and Ullrich Munzel. "The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation". *Biometrical Journal*, 2000.
DOI: 10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U [66]

[421] William H. Kruskal and W. Allen and Wallis. "Use of Ranks in One-Criterion Variance Analysis". *Journal of the American Statistical Association*, 1952.
DOI: 10.1080/01621459.1952.10483441 [66]

[422] Olive Jean Dunn. "Multiple Comparisons Using Rank Sums". *Technometrics*, 1964.
DOI: 10.1080/00401706.1964.10490181 [66]

[423] Stanton A. Glantz. *Primer of Biostatistics*. 7th Edition. McGraw Hill, 2012.
(Visited on 03/22/2025) [66]

[424] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods*, 2020.
DOI: 10.1038/s41592-019-0686-2 [66]

[425] Giedrius Gasiunas et al. "A Catalogue of Biochemically Diverse CRISPR-Cas9 Orthologs". *Nature Communications*, 2020.
DOI: 10.1038/s41467-020-19344-1 [67]

[426] Benjamin Ambrose, James M. Baxter, John Cully, Matthew Willmott, Elliot M. Steele, Benji C. Bateman, Marisa L. Martin-Fernandez, Ashley Cadby, Jonathan Shewring, Marleen Aaldering, and Timothy D. Craggs. "The smfBox Is an Open-Source Platform for Single-Molecule FRET". *Nature Communications*, 2020.
DOI: 10.1038/s41467-020-19468-4 [68]

[427] Mahmoud A. S. Abdelhamid, Alice V. Rhind-Tutt, Benjamin Ambrose, and Timothy D. Craggs. "Making Precise and Accurate Single-Molecule FRET Measurements Using the Open-Source smfBox". *JoVE*, 2021.
DOI: 10.3791/62378 [68]

[428] Eyal Nir, Xavier Michalet, Kambiz M. Hamadani, Ted A. Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. "Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments". *The Journal of Physical Chemistry B*, 2006.
DOI: 10.1021/jp063483n [68]

[429] Ivan Terterov, Daniel Nettels, Dmitrii E. Makarov, and Hagen Hofmann. "Time-Resolved Burst Variance Analysis". *Biophysical Reports*, 2023.
DOI: 10.1016/j.bpr.2023.100116 [68]

[430] Amir Bar, Alkan Kabakç çı ıoğ ğlu, and David Mukamel. "Denaturation of Circular DNA: Supercoils and Overtwist". *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2012.

DOI: 10.1103/PhysRevE.86.061904 [75]

[431] Karl O. Stetter. "History of Discovery of the First Hyperthermophiles". *Extremophiles*, 2006.
DOI: 10.1007/s00792-006-0012-7 [89]

[432] Tribeni Chatterjee, Anupama Tiwari, Ritika Gupta, Himadri Shukla, Aastha Varshney, Satish Mishra, and Saman Habib. "A Plasmodium Apicoplast-Targeted Unique Exonuclease/FEN Exhibits Interspecies Functional Differences Attributable to an Insertion That Alters DNA-binding". *Nucleic Acids Research*, 2024.
DOI: 10.1093/nar/gkae512 [99]

[433] D. H. Williamson, P. R. Preiser, P. W. Moore, S. McCready, M. Strath, and R. J. M. (Iain) Wilson. "The Plastid DNA of the Malaria Parasite Plasmodium Falciparum Is Replicated by Two Mechanisms". *Molecular Microbiology*, 2002.
DOI: 10.1046/j.1365-2958.2002.03033.x [99]

[434] Christopher D. Goodman, Hayley D. Buchanan, and Geoffrey I. McFadden. "Is the Mitochondrion a Good Malaria Drug Target?" *Trends in Parasitology*, 2017.
DOI: 10.1016/j.pt.2016.10.002 [99]

[435] Marco Biddau and Lilach Sheiner. "Targeting the Apicoplast in Malaria." *Biochemical Society Transactions*, 2019.
DOI: 10.1042/BST20170563 [99]

[436] Liting Lim and Geoffrey Ian McFadden. "The Evolution, Metabolism and Functions of the Apicoplast". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2010.
DOI: 10.1098/rstb.2009.0273 [99]

[437] Joshua D. Hawk, Lela Stefanovic, Jayne C. Boyer, Thomas D. Petes, and Rosann A. Farber. "Variation in Efficiency of DNA Mismatch Repair at Different Sites in the Yeast Genome." *Proceedings of the National Academy of Sciences of the United States of America*, 2005.
DOI: 10.1073/pnas.0503415102 [103, 105]

[438] Thuy T. M. Ngo, Bailey Liu, Feng Wang, Aakash Basu, Carl Wu, and Taekjip Ha. "Dependence of Nucleosome Mechanical Stability on DNA Mismatches." *eLife*, 2024.
DOI: 10.7554/eLife.95514 [103]

[439] Scott A. Lujan, Jessica S. Williams, Zachary F. Pursell, Amy A. Abdulovic-Cui, Alan B. Clark, Stephanie A. Nick McElhinny, and Thomas A. Kunkel. "Mismatch Repair Balances Leading and Lagging Strand DNA Replication Fidelity." *PLoS Genetics*, 2012.
DOI: 10.1371/journal.pgen.1003016 [103]

[440] Brian D Harfe and Sue Jinks-Robertson. "Sequence Composition and Context Effects on the Generation and Repair of Frameshift Intermediates in Mononucleotide Runs in Saccharomyces Cerevisiae". *Genetics*, 2000.
DOI: 10.1093/genetics/156.2.571 [103]

[441] Barbara Kramer, Wilfried Kramer, and Hans-Joachim Fritz. "Different Base/Base

Mismatches Are Corrected with Different Efficiencies by the Methyl-Directed DNA Mismatch-Repair System of E. Coli". *Cell*, 1984.
DOI: 10.1016/0092-8674(84)90283-6 [103, 104, 105, 108, 145]

[442] Tunc Kayikcioglu, Jasmin S. Zarb, Sonisilpa Mohapatra, Chang-Ting Lin, James A. London, Kasper D. Hansen, Richard Fishel, and Taekjip Ha. "Massively Parallel Single Molecule Tracking of Sequence-Dependent DNA Mismatch Repair in Vivo". *bioRxiv*, 2023.
DOI: 10.1101/2023.01.08.523062 [103, 104, 105, 108, 144, 145, 146]

[443] Christiane Dohet, Robert Wagner, and Miroslav Radman. "Repair of Defined Single Base-Pair Mismatches in Escherichia Coli." *Proceedings of the National Academy of Sciences of the United States of America*, 1985.
DOI: 10.1073/pnas.82.2.503 [103, 104, 108, 145]

[444] Hong Wang, Yong Yang, Mark J. Schofield, Chunwei Du, Yonatan Fridman, Susan D. Lee, Erik D. Larson, James T. Drummond, Eric Alani, Peggy Hsieh, and Dorothy A. Erie. "DNA Bending and Unbending by MutS Govern Mismatch Recognition and Specificity". *Proceedings of the National Academy of Sciences*, 2003.
DOI: 10.1073/pnas.2433654100 [103, 104]

[445] Madeleine Jones, Robert Wagner, and Miroslav Radman. "Repair of a Mismatch Is Influenced by the Base Composition of the Surrounding Nucleotide Sequence." *Genetics*, 1987.
DOI: 10.1093/genetics/115.4.605 [103]

[446] Isabelle Varlet, Miroslav Radman, and Peter Brooks. "DNA Mismatch Repair in Xenopus Egg Extracts: Repair Efficiency and DNA Repair Synthesis for All Single Base-Pair Mismatches." *Proceedings of the National Academy of Sciences of the United States of America*, 1990.
DOI: 10.1073/pnas.87.20.7883 [103]

[447] Anna Tikhomirova, Irina V. Beletskaya, and Tigran V. Chalikian. "Stability of DNA Duplexes Containing GG, CC, AA, and TT Mismatches". *Biochemistry*, 2006.
DOI: 10.1021/bi060304j [103]

[448] Ganesh Natrajan, Meindert H. Lamers, Jacqueline H. Enzlin, Herrie H. K. Winterwerp, Anastassis Perrakis, and Titia K. Sixma. "Structures of Escherichia Coli DNA Mismatch Repair Enzyme MutS in Complex with Different Mismatches: A Common Recognition Mode for Diverse Substrates". *Nucleic Acids Research*, 2003.
DOI: 10.1093/nar/gkg677 [103, 107, 111]

[449] Ingrid Tessmer, Yong Yang, Jie Zhai, Chungwei Du, Peggy Hsieh, Manju M. Hingorani, and Dorothy A. Erie. "Mechanism of MutS Searching for DNA Mismatches and Signaling Repair." *The Journal of Biological Chemistry*, 2008.
DOI: 10.1074/jbc.M805712200 [104, 107]

[450] Greg L. Hura, Chi-Lin Tsai, Shelley A. Claridge, Marc L. Mendillo, Jessica M. Smith, Gareth J. Williams, Alexander J. Mastroianni, A. Paul Alivisatos, Christopher D. Putnam, Richard D. Kolodner, and John A. Tainer. "DNA Conformations in Mismatch Repair Probed in Solution by X-ray Scattering from Gold

Nanocrystals." *Proceedings of the National Academy of Sciences of the United States of America*, 2013.
DOI: 10.1073/pnas.1308595110 [104]

[451] James Brown, Tom Brown, and Keith R. Fox. "Affinity of Mismatch-Binding Protein MutS for Heteroduplexes Containing Different Mismatches." *The Biochemical journal*, 2001.
DOI: 10.1042/0264-6021:3540627 [104, 146]

[452] Anthony Mazurek, Christopher N. Johnson, Markus W. Germann, and Richard Fishel. "Sequence Context Effect for hMSH2-hMSH6 Mismatch-Dependent Activation". *Proceedings of the National Academy of Sciences*, 2009.
DOI: 10.1073/pnas.0808572106 [104]

[453] Minseon Cho, Min Su Han, and Changill Ban. "Detection of Mismatched DNAsvia the Binding Affinity of MutS Using a Gold Nanoparticle-Based Competitive Colorimetric Method". *Chemical Communications*, 2008.
DOI: 10.1039/B811346G [104]

[454] Vanessa C. DeRocco, Lauryn E. Sass, Ruoyi Qiu, Keith R. Weninger, and Dorothy A. Erie. "Dynamics of MutS–Mismatched DNA Complexes Are Predictive of Their Repair Phenotypes". *Biochemistry*, 2014.
DOI: 10.1021/bi401429b [104, 121]

[455] Alexander P. Fields, Elisabeth A. Meyer, and Adam E. Cohen. "Euler Buckling and Nonlinear Kinking of Double-Stranded DNA." *Nucleic Acids Research*, 2013.
DOI: 10.1093/nar/gkt739 [104, 121, 146]

[456] Michael A. Edelbrock, Saravanan Kaliyaperumal, and Kandace J. Williams. "DNA Mismatch Repair Efficiency and Fidelity Are Elevated during DNA Synthesis in Human Cells." *Mutation research*, 2009.
DOI: 10.1016/j.mrfmmm.2008.12.006 [105]

[457] Flavia C Hasenauer, Hugo C Barreto, Chantal Lotton, and Ivan Matic. "Genome-Wide Mapping of Spontaneous DNA Replication Error-Hotspots Using Mismatch Repair Proteins in Rapidly Proliferating Escherichia Coli". *Nucleic Acids Research*, 2025.
DOI: 10.1093/nar/gkae1196 [105]

[458] Cai Li and Nicholas M. Luscombe. "Nucleosome Positioning Stability Is a Modulator of Germline Mutation Rate Variation across the Human Genome". *Nature Communications*, 2020.
DOI: 10.1038/s41467-020-15185-0 [105]

[459] Feng Li, Lei Tian, Liya Gu, and Guo-Min Li. "Evidence That Nucleosomes Inhibit Mismatch Repair in Eukaryotic Cells*". *Journal of Biological Chemistry*, 2009.
DOI: 10.1074/jbc.M109.049874 [105]

[460] Peng Mao, Alexander J. Brown, Ewa P. Malc, Piotr A. Mieczkowski, Michael J. Smerdon, Steven A. Roberts, and John J. Wyrick. "Genome-Wide Maps of Alkylation Damage, Repair, and Mutagenesis in Yeast Reveal Mechanisms of Mutational Heterogeneity." *Genome Research*, 2017.
DOI: 10.1101/gr.225771.117 [105]

[461] Hope A. Cole, Jenna M. Tabor-Godwin, and Jeffrey J. Hayes. "Uracil DNA Glycosylase Activity on Nucleosomal DNA Depends on Rotational Orientation of Targets." *The Journal of Biological Chemistry*, 2010.
DOI: 10.1074/jbc.M109.073544 [105]

[462] John M Hinz, Yesenia Rodriguez, and Michael J Smerdon. "Rotational Dynamics of DNA on the Nucleosome Surface Markedly Impact Accessibility to a DNA Repair Enzyme". *Proceedings of the National Academy of Sciences*, 2010.
DOI: 10.1073/pnas.0914443107 [105]

[463] Abhilash Jayaraj, Kelly M. Thayer, David L. Beveridge, and Manju M. Hingorani. "Molecular Dynamics of Mismatch Detection—How MutS Uses Indirect Readout to Find Errors in DNA". *Biophysical Journal*, 2023.
DOI: 10.1016/j.bpj.2023.06.006 [105, 108]

[464] Scott J Witte, Isabella M Rosa, Bryce W Collingwood, Jonathan M Piscitelli, and Carol M Manhart. "The Mismatch Repair Endonuclease MutLα Tethers Duplex Regions of DNA Together and Relieves DNA Torsional Tension". *Nucleic Acids Research*, 2023.
DOI: 10.1093/nar/gkad096 [106]

[465] Thomas A. Kunkel and Dorothy A. Erie. "Eukaryotic Mismatch Repair in Relation to DNA Replication." *Annual Review of Genetics*, 2015.
DOI: 10.1146/annurev-genet-112414-054722 [106, 145]

[466] Farid A. Kadyrov, Shannon F. Holmes, Mercedes E. Arana, Olga A. Lukianova, Mike O'Donnell, Thomas A. Kunkel, and Paul Modrich. "Saccharomyces Cerevisiae MutLalpha Is a Mismatch Repair Endonuclease." *The Journal of Biological Chemistry*, 2007.
DOI: 10.1074/jbc.M707617200 [106]

[467] Anna Pluciennik, Leonid Dzantiev, Ravi R. Iyer, Nicoleta Constantin, Farid A. Kadyrov, and Paul Modrich. "PCNA Function in the Activation and Strand Direction of MutLα Endonuclease in Mismatch Repair". *Proceedings of the National Academy of Sciences*, 2010.
DOI: 10.1073/pnas.1010662107 [106]

[468] Gerald T. Marsischky and Richard D. Kolodner. "Biochemical Characterization of the Interaction between the Saccharomyces Cerevisiae MSH2-MSH6 Complex and Mispaired Bases in DNA *". *Journal of Biological Chemistry*, 1999.
DOI: 10.1074/jbc.274.38.26668 [106]

[469] Mark C. Hall, Hong Wang, Dorothy A. Erie, and Thomas A. Kunkel. "High Affinity Cooperative DNA Binding by the Yeast Mlh1-Pms1 Heterodimer." *Journal of Molecular Biology*, 2001.
DOI: 10.1006/jmbi.2001.4958 [106]

[470] Sumitabha Brahmachari, Andrew Dittmore, Yasuharu Takagi, Keir C. Neuman, and John F. Marko. "Defect-Facilitated Buckling in Supercoiled Double-Helix DNA." *Physical review. E*, 2018.
DOI: 10.1103/PhysRevE.97.022416 [106, 152]

[471] Ariel Afek, Honglue Shi, Atul Rangadurai, Harshit Sahay, Alon Senitzki, Suela Xhani, Mimi Fang, Raul Salinas, Zachery Mielko, Miles A. Pufall, Gregory M. K. Poon, Tali E. Haran, Maria A. Schumacher, Hashim M. Al-Hashimi, and Raluca Gordân. "DNA Mismatches Reveal Conformational Penalties in Protein-DNA Recognition." *Nature*, 2020.
DOI: 10.1038/s41586-020-2843-2 [106]

[472] Luciana M. Oliveira, Adam S. Long, Tom Brown, Keith R. Fox, and Gerald Weber. "Melting Temperature Measurement and Mesoscopic Evaluation of Single, Double and Triple DNA Mismatches". *Chemical Science*, 2020.
DOI: 10.1039/D0SC01700K [107, 108, 111, 113, 125, 126, 137, 139, 144, 236]

[473] Natália Kolozsvári and Martin R. Gill. "Targeting DNA Mismatches with Metal Complexes". *Journal of Inorganic Biochemistry*, 2025.
DOI: 10.1016/j.jinorgbio.2025.112977 [107]

[474] WikiLectures contributors. *Basic Components of Nucleic Acids — WikiLectures,* https://www.wikilectures.eu/index.php?title=Basic_components_of_nucleic_acids&oldid=61358. 2022.
(Visited on 06/25/2025) [107]

[475] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. "UCSF Chimera–a Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry*, 2004.
DOI: 10.1002/jcc.20084 [110]

[476] Joyce H.G. Lebbink, Alexander Fish, Annet Reumer, Ganesh Natrajan, Herrie H.K. Winterwerp, and Titia K. Sixma. "Magnesium Coordination Controls the Molecular Switch Function of DNA Mismatch Repair Protein MutS *". *Journal of Biological Chemistry*, 2010.
DOI: 10.1074/jbc.M109.066001 [111]

[477] Tomáš Bouchal, Ivo Durník, Viktor Illík, Kamila Réblová, and Petr Kulhánek. "Importance of Base-Pair Opening for Mismatch Recognition." *Nucleic Acids Research*, 2020.
DOI: 10.1093/nar/gkaa896 [111, 118]

[478] Gilbert G. Privé, Kazunori Yanagi, and Richard E. Dickerson. "Structure of the B-DNA Decamer C-C-A-A-C-G-T-T-G-G and Comparison with Isomorphous Decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G". *Journal of Molecular Biology*, 1991.
DOI: 10.1016/0022-2836(91)90619-H [111]

[479] Roshan Satange, Shih-Hao Kao, Ching-Ming Chien, Shan-Ho Chou, Chi-Chien Lin, Stephen Neidle, and Ming-Hon Hou. "Staggered Intercalation of DNA Duplexes with Base-Pair Modulation by Two Distinct Drug Molecules Induces Asymmetric Backbone Twisting and Structure Polymorphism". *Nucleic Acids Research*, 2022.
DOI: 10.1093/nar/gkac629 [111]

[480] John Jr. Santa Lucia and Douglas H. Turner. "Structure of (rGGCGAGCC)2 in Solution from NMR and Restrained Molecular Dynamics". *Biochemistry*, 1993.

DOI: 10.1021/bi00210a009 [111]

[481] Y. Boulard, J.A.H. Cognet, and G.V. Fazakerley. "Solution Structure as a Function of pH of Two Central Mismatches, C·T and C·C, in the 29 to 39 K-ras Gene Sequence, by Nuclear Magnetic Resonance and Molecular dynamics11Edited by T. Tinoco". *Journal of Molecular Biology*, 1997.
DOI: 10.1006/jmbi.1997.0975 [111, 121]

[482] Stephanie C. C. van der Lubbe and Célia Fonseca Guerra. "Hydrogen-Bond Strength of CC and GG Pairs Determined by Steric Repulsion: Electrostatics and Charge Transfer Overruled". *Chemistry – A European Journal*, 2017.
DOI: 10.1002/chem.201701821 [111]

[483] Cathal T. Gallagher and Mark S. Searle. "Drug-Induced Stabilisation of a Mismatched C-T Base Pair in a DNA Hairpin." *Chemical Communications*, 2003.
DOI: 10.1039/b305337g [111]

[484] Roshan Satange, Chien-Ying Chuang, Stephen Neidle, and Ming-Hon Hou. "Polymorphic G:G Mismatches Act as Hotspots for Inducing Right-Handed Z DNA by DNA Intercalation". *Nucleic Acids Research*, 2019.
DOI: 10.1093/nar/gkz653 [111]

[485] William N. Hunter, Geoff Kneale, Tom Brown, Dov Rabinovich, and Olga Kennard. "Refined Crystal Structure of an Octanucleotide Duplex with G · T Mismatched Base-Pairs". *Journal of Molecular Biology*, 1986.
DOI: 10.1016/0022-2836(86)90246-9 [111]

[486] William N. Hunter, Tom Brown, Geoff Kneale, Naveen N. Anand, Dov Rabinovich, and Olga Kennard. "The Structure of Guanosine-Thymidine Mismatches in B-DNA at 2.5-A Resolution." *Journal of Biological Chemistry*, 1987.
DOI: 10.1016/S0021-9258(18)61060-9 [111]

[487] Guoyun He, Chun Kit Kwok, and Sik Lok Lam. "Preferential Base Pairing Modes of T·T Mismatches". *FEBS Letters*, 2011.
DOI: 10.1016/j.febslet.2011.10.044 [111]

[488] Sean J. Johnson and Lorena S. Beese. "Structures of Mismatch Replication Errors Observed in a DNA Polymerase". *Cell*, 2004.
DOI: 10.1016/S0092-8674(04)00252-1 [111]

[489] Xiang-Jun Lu and Wilma K Olson. "3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures". *Nature Protocols*, 2008.
DOI: 10.1038/nprot.2008.104 [112]

[490] Marla S. Babcock, Edwin O. D. Pednault, and Wilma K. Olson. "Nucleic Acid Structure Analysis. Mathematics for Local Cartesian and Helical Structure Parameters That Are Truly Comparable between Structures." *Journal of Molecular Biology*, 1994.
DOI: 10.1006/jmbi.1994.1213 [113]

[491] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. "MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations". *Journal of Computational Chemistry*, 2011.

DOI: 10.1002/jcc.21787 [113]

[492] Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein. "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations". IN: *Proceedings of the 15th Python in Science Conference*. ed. by Sebastian Benthall and Scott Rostrup. 2016.
DOI: 10.25080/Majora-629e541a-00e [113]

[493] Monika Sharma, Alexander V. Predeus, Shayantani Mukherjee, and Michael Feig. "DNA Bending Propensity in the Presence of Base Mismatches: Implications for DNA Repair". *The Journal of Physical Chemistry B*, 2013.
DOI: 10.1021/jp403127a [118]

[494] Jason D. Kahn. "DNA, Flexibly Flexible". *Biophysical Journal*, 2014.
DOI: 10.1016/j.bpj.2014.06.007 [120]

[495] Ebenezer O. Otokiti and Richard D. Sheardy. "Effect of Base Pair A/C and G/T Mismatches on the Thermal Stabilities of DNA Oligomers That Form B-Z Junctions". *Biochemistry*, 1997.
DOI: 10.1021/bi970972s [125]

[496] Tapas Paul, Subhas C. Bera, and Padmaja P. Mishra. "Direct Observation of Breathing Dynamics at the Mismatch Induced DNA Bubble with Nanometre Accuracy: A smFRET Study". *Nanoscale*, 2017.
DOI: 10.1039/C6NR09348E [130]

[497] Amita Joshi and Basuthkar J. Rao. "MutS Recognition: Multiple Mismatches and Sequence Context Effects". *Journal of Biosciences*, 2001.
DOI: 10.1007/BF02704758 [137]

[498] John A. Birdsell. "Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution". *Molecular Biology and Evolution*, 2002.
DOI: 10.1093/oxfordjournals.molbev.a004176 [146]

[499] Midas Segers, Enrico Skoruppa, Helmut Schiessel, and Enrico Carlon. "Statistical Mechanics of Multiplectoneme Phases in DNA". *Physical Review E*, 2025.
DOI: 10.1103/PhysRevE.111.044408 [155]

Appendix

A.1 Sequences

Control339:

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATGA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 GGACCTGCAT GGCTGCGTAC GGAGGAAGGA ACTCTTGCGT GGTAGTATGT TGACCACTGT
181 ATCACGGATG CGGGTAGAAG ATGTGAGCAG AGACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTGCA CCATGCGACC AGACCACTCG AGCATGGCAC TGTGTACGCT
301 GTCTCCAGTC TGGTAGTCAT CATCCCTAGC ACGGTCTCG
```

Mod339:

Mutations from Control339 are highlighted in blue.

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATAA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 TGACTTGCAT AGCTGCGTAT GGAGGAAGGA ACTCTTGCGT GTTAGTATGT TGACCACTGT
181 ATCACGGATG CGGGTAGAAG ATGTGAGCAG AGACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTTCA CCATACGACC ATACCATTCG AGCATGGCAC TATGTACGCT
301 GTCTCCATTC TGGTAGTCAT CATCCCTATC ACGGTTTCG
```

Strong339:

```
  1 TCTCTCTCTC TCTCTCACCA CCGCGTACAA GAAAGTTTGT TCTTTCATTT CGCCATACAG
 61 CTGCCTCGCG CGTTTCGGTG ATATCGGTGA AAACCCTTCC CGTTTCGCTC AAGTTAGTAT
121 AAAAAAGCTG AACGAGAAAC GTAAAATGAT ATAAATATCA ATATATTAAA TTAGATTTTG
181 CATAAAAAAC AGACTACATA ATACTGTAAA ACACAACATA TGCAGTCACT ATGAATCAAC
241 TACTTAGATG GTATTAGTGA CCTGTAACAG AGCTGAGGCA TCTTTCTCAA TGAAAGTGTC
301 GGAACTCGGT ACATAGGCTT CCGGCTGATA GTAGTCGTA
```

Curve75-1:

The curved 75-mer, as taken from the Dekker paper [41], is highlighted in blue.

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATGA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 GGACCTGCAT GGCTGCGCGA ATCAGCCTAT TTAGGCTATT TTTTCCACCA TTTCTGGCGT
181 TATTTCCGGT TTTTACTGAG ATCTCTCCCA CTACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTGCA CCATGCGACC AGACCACTCG AGCATGGCAC TGTGTACGCT
301 GTCTCCAGTC TGGTAGTCAT CATCCCTAGC ACGGTCTCG
```

Curve75-2:

The curved 75-mer, as taken from the Dekker paper [41], is highlighted in blue.

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATGA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 GGACCTGCAT GGCTGCGGAT GCTCACCGCA TTTCCTGAAA ATTCACGCTG TATCTTGAAA
181 AATCGACGTT TTTTACGTGG TTTTCCGTCG AAACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTGCA CCATGCGACC AGACCACTCG AGCATGGCAC TGTGTACGCT
301 GTCTCCAGTC TGGTAGTCAT CATCCCTAGC ACGGTCTCG
```

Flat75-1:

The flat 75-mer, as taken from the Dekker paper [41], is highlighted in blue.

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATGA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 GGACCTGCAT GGCTGCGAGG GGGGCGGCTA TTGAAGGGGT CCCGTTGGTC TGCCCACAAA
181 AGGCCCCCTG AGTCTTTCAC TCAGTTTGTC CTACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTGCA CCATGCGACC AGACCACTCG AGCATGGCAC TGTGTACGCT
301 GTCTCCAGTC TGGTAGTCAT CATCCCTAGC ACGGTCTCG
```

Flat75-2:

The flat 75-mer, as taken from the Dekker paper [41], is highlighted in blue.

```
  1 CGCTACACAC ATCGTACTGC AACTGCACCT AGCTGAGATG GTTCCATAGA TGACTACGCA
 61 TCCCTCTAGG CCTTACATGA CCTGATACAG TGACTTTGAC AGGTTTGCGA GGTACAGCAA
121 GGACCTGCAT GGCTGCGACA GGTCCCGGGT GTTCACGCAT CCACAGACCC CTCGGAGGGA
181 GCATTAGAGG CTGGGGCTCA TCTGGGGGTC TGACACCCAG GTCAAGTTCT CGACCTTCTC
241 GGAGGAGGTG TTCCAGTGCA CCATGCGACC AGACCACTCG AGCATGGCAC TGTGTACGCT
301 GTCTCCAGTC TGGTAGTCAT CATCCCTAGC ACGGTCTCG
```

126-mer:

```
  1 CCCTTCTTCT TCTGCTCGGA CTCTTTAATA GTTAATAATT TAATAGTTTG AAATTTAATA
 61 GTTAACAATT TAATAGTTAT TAATTTAATA GTTTAATATT TAATCTTAAA CAACAGATAT
121 CATCTC
```

Control55:

The position of the mismatch is highlighted in blue.

```
  1 ATGACCTGAT ACAGTGACTT TGACAGGTTT GCGAGGTACA GCAAGGACCT GCATG
```

Mod55:

The position of the mismatch is highlighted in blue.

```
  1 GTTTCGCGCT ACACACATCG TACTGCAACT GCACCTAGCT GAGATGGTTC CATAG
```

A.2 Supplementary Figures and Tables

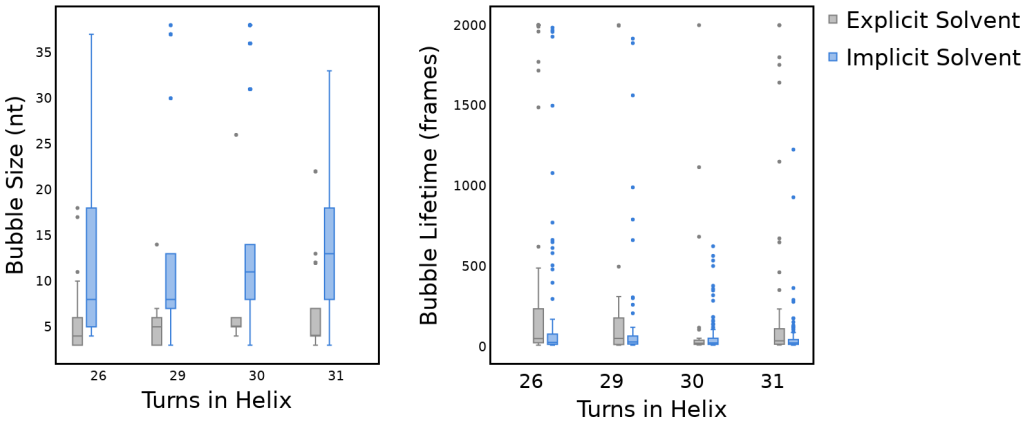

Figure A.1: The bubbles are larger ($p = 5.34 \times 10^{-60}$) in implicit solvent but longer lived ($p = 1.58 \times 10^{-4}$) in explicit solvent.
Box plots of the bubble sizes **(left)** and lifetimes **(right)** across all simulations at the specified superhelical densities. The box plots compare the simulations of Strong339 in implicit (blue) and explicit (grey) solvent. The bubble size is calculated in terms of the number of consecutive nucleotides (nt) disrupted by the formation of the bubble. The bubble lifetime is calculated by the number of frames each bubble exists for where the maximum is 2000.

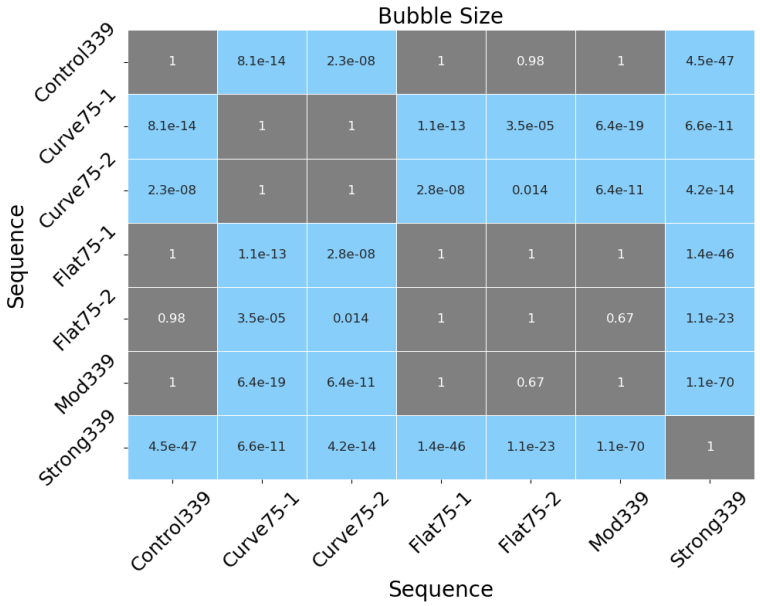

Figure A.2: The sequence influences the size of bubbles formed in line with the bubble-driven plectoneme propensity profiles.
The p values of Dunn's post-hoc test with Bonferroni corrections comparing bubble size against sequence. Significant ($p < 0.05$) values are highlighted in blue whilst non-significant values are in grey. The pairwise comparisons were calculated by including the data for all superhelical densities for each corresponding sequence.

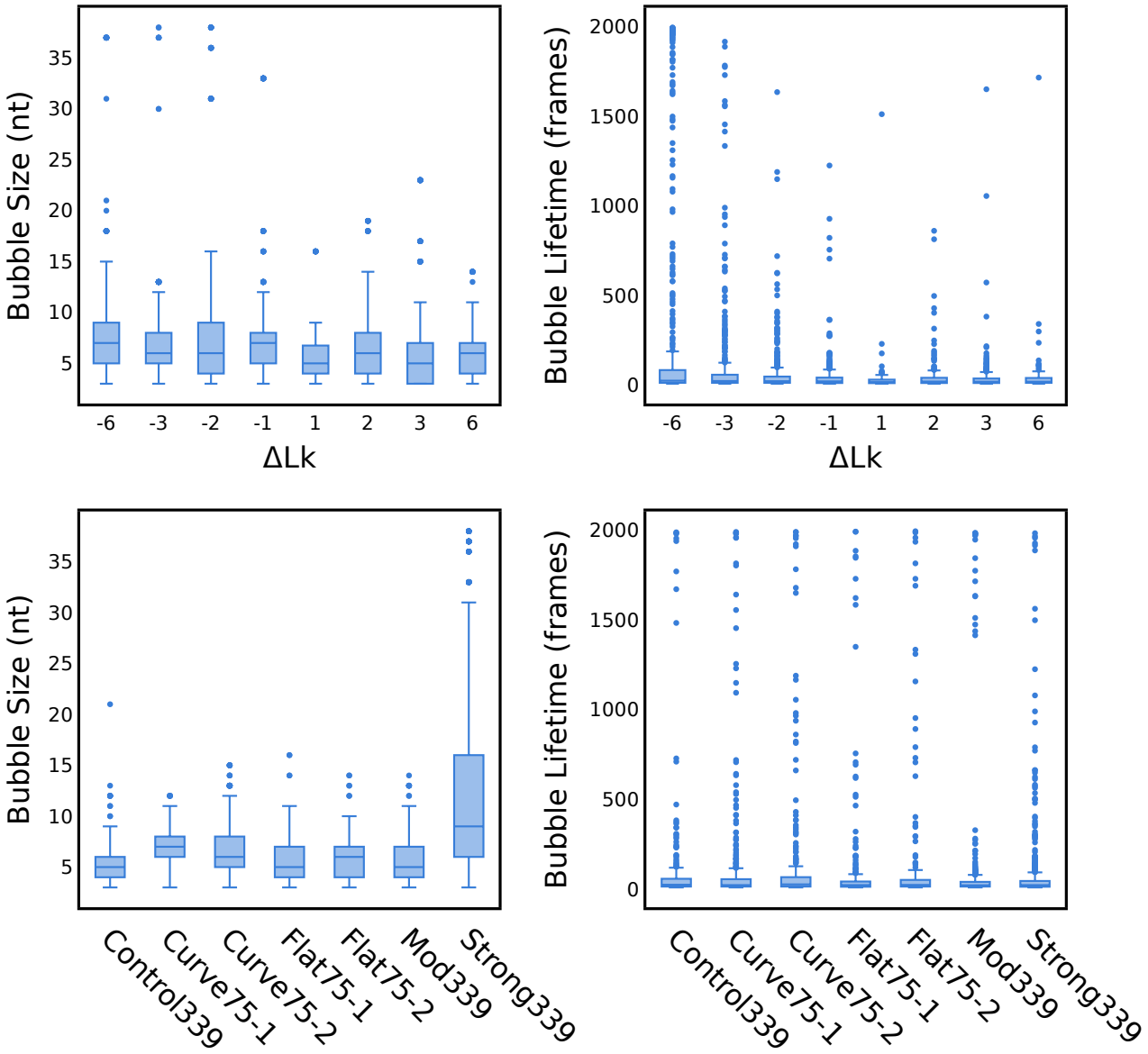

Figure A.3: Both bubble size and lifetime are influenced by the degree of supercoiling; however, sequence only influences the size of bubbles and not their lifetime.
Box plots of the bubble sizes **(left)** and bubble lifetimes **(right)** compared across superhelical density **(top)** and across sequence **(bottom)**. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There is a possible total of 2000 frames. For comparison by superhelical density, the data was plotted for all sequences at each specified number of turns. See fig. A.4 for pairwise comparisons.

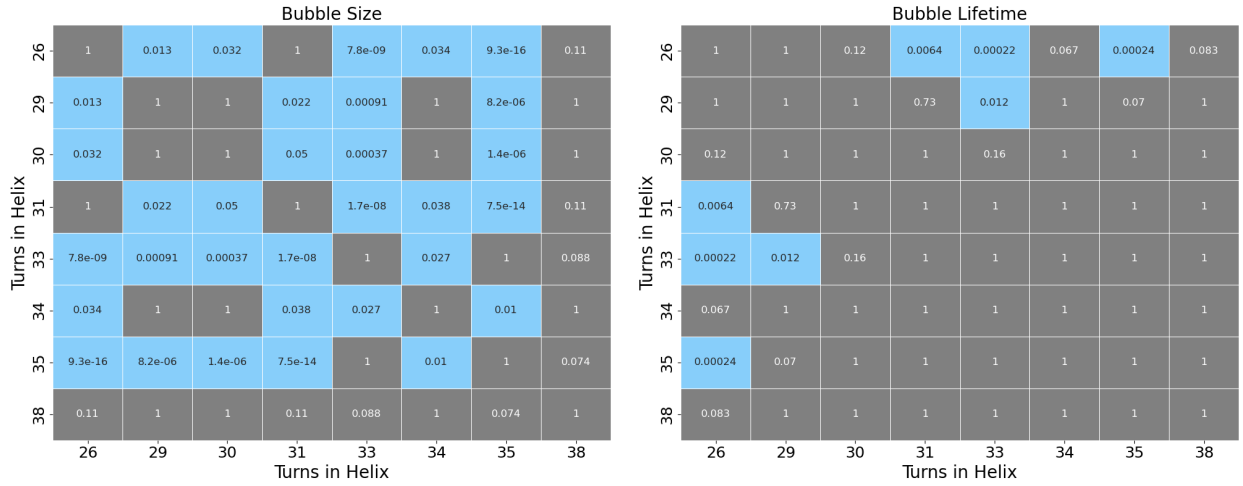

Figure A.4: The degree of superhelical density affects both bubble lifetime and bubble size. The p values of Dunn's post-hoc test with Bonferroni corrections comparing bubble size **(left)** and lifetime **(right)** by the number of turns in the helix. The relaxed topoisomer has 32 turns. Significant ($p < 0.05$) values are highlighted in blue whilst non-significant values are in grey. The pairwise comparisons were calculated by including the data for all 339-mer sequences at the specified superhelical densities.

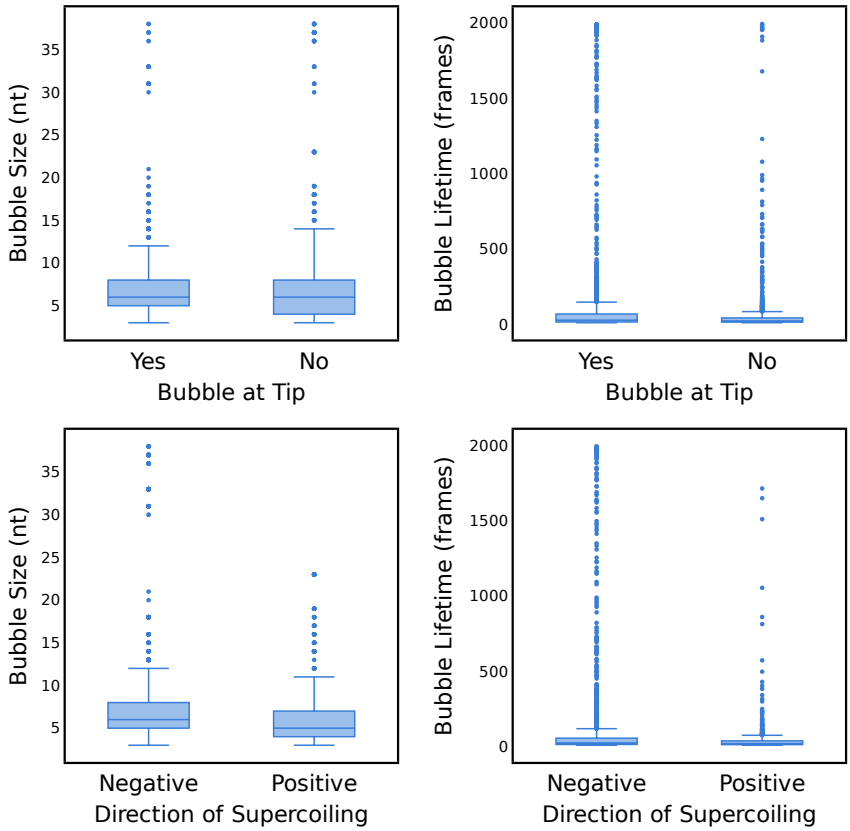

Figure A.5: Bubbles at the tip of the plectoneme are shorter and longer lived than those nucleated elsewhere, whilst bubbles under negative supercoiling are larger and longer lived than those under positive supercoiling.
Box plots of the bubble sizes **(left)** and bubble lifetimes **(right)** compared across location **(top)** and across direction of supercoiling **(bottom)**. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are defined as the number of frames each bubble persists for throughout the simulation after nucleation. There is a possible total of 2000 frames. For comparison by bubble location, a bubble was considered to be at the tip of the plectoneme if it was within 20 nucleotides of the two base pairs with the maximum average distance from the centre of mass of the DNA over the frames in which it exists.

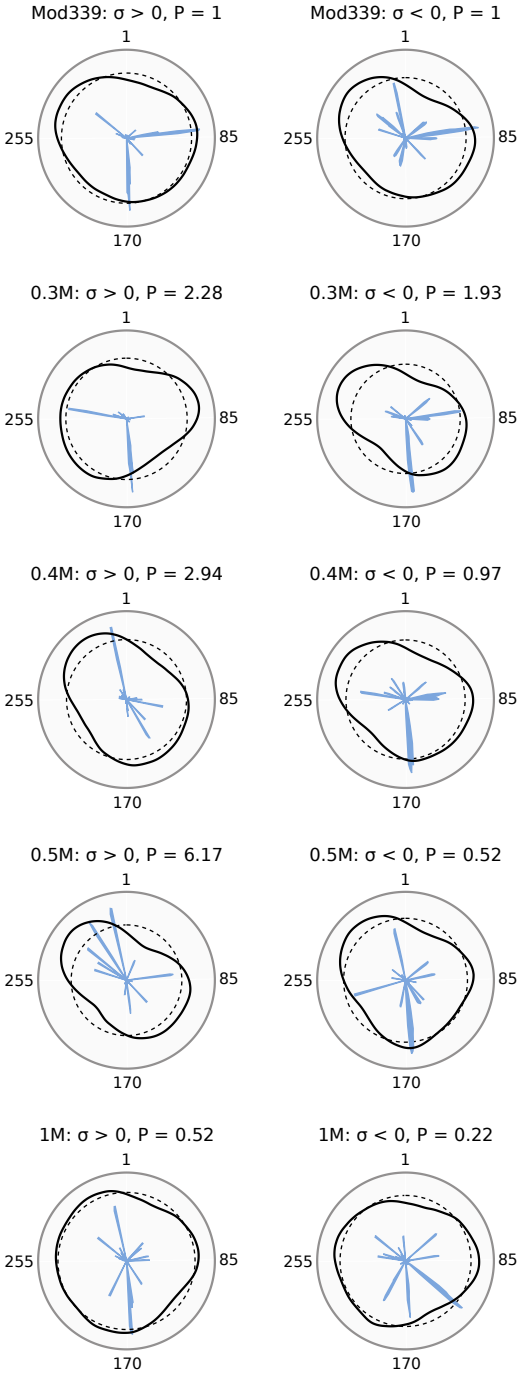

Figure A.6: As salt concentration increases, the pinning landscapes under negative supercoiling become more in line with that under positive supercoiling.
The plectoneme propensity (black), average plectoneme propensity (black, dashed) and bubble density (blue, filled) of the Mod339 sequence under positive **(left)** and negative **(right)** supercoiling at increased salt concentration. The bubble and plectoneme densities are normalised such that the maximum value for each parameter is 1.

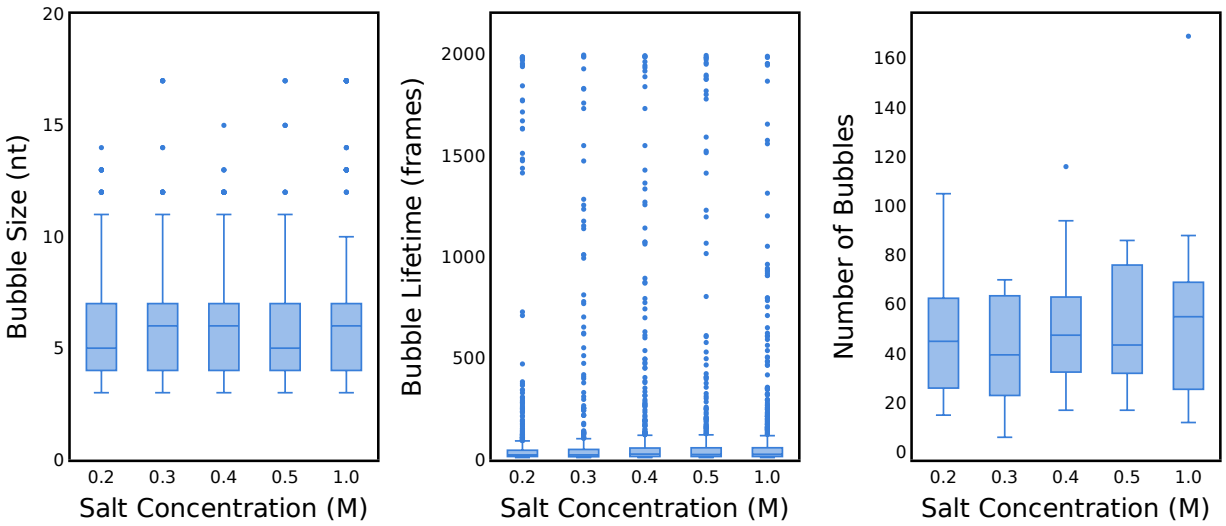

Figure A.7: Increasing salt concentration has little effect on bubble properties, despite directly influencing twist.
Box plots of the bubble sizes **(left)**, bubble lifetimes **(middle)**, and bubble count **(right)** compared across salt concentration. The data is combined for both Control339 and Mod339. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There is a possible total of 2000 frames. The number of bubbles is the number of unique bubbles across all simulations.

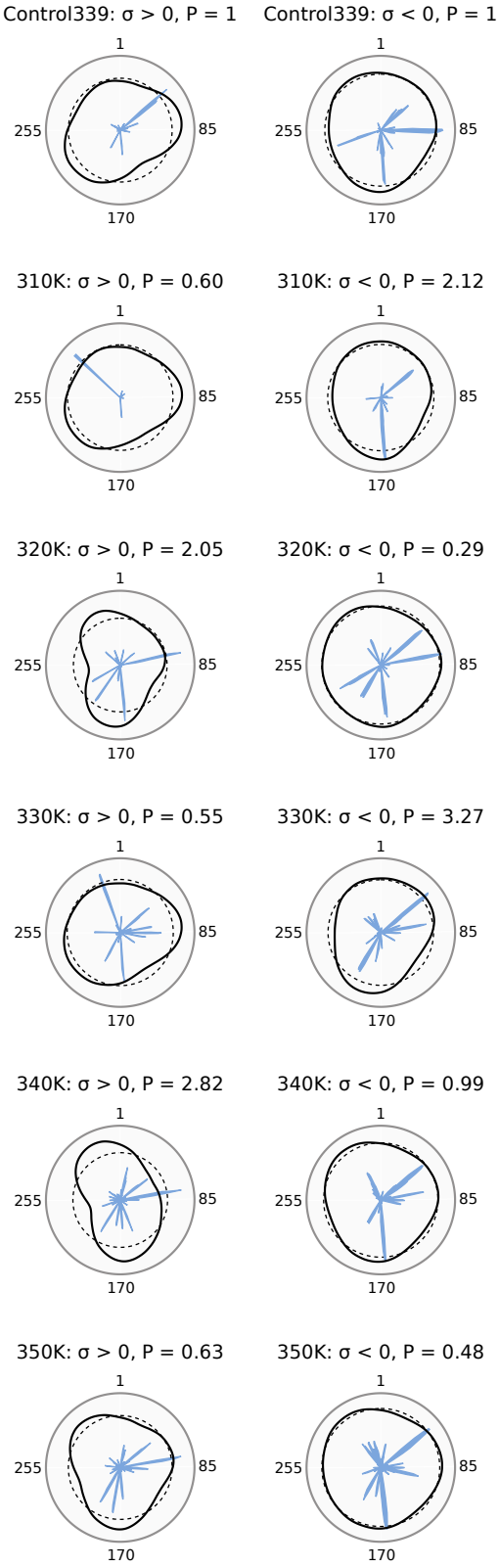

Figure A.8: As temperature increases, the pinning landscapes under negative supercoiling are increasingly driven by bubble formation.
The plectoneme propensity (black), average plectoneme propensity (black, dashed) and bubble density (blue, filled) of the Control339 sequence under positive **(left)** and negative **(right)** supercoiling at increased temperature. The bubble and plectoneme densities are normalised such that the maximum value for each parameter is 1.

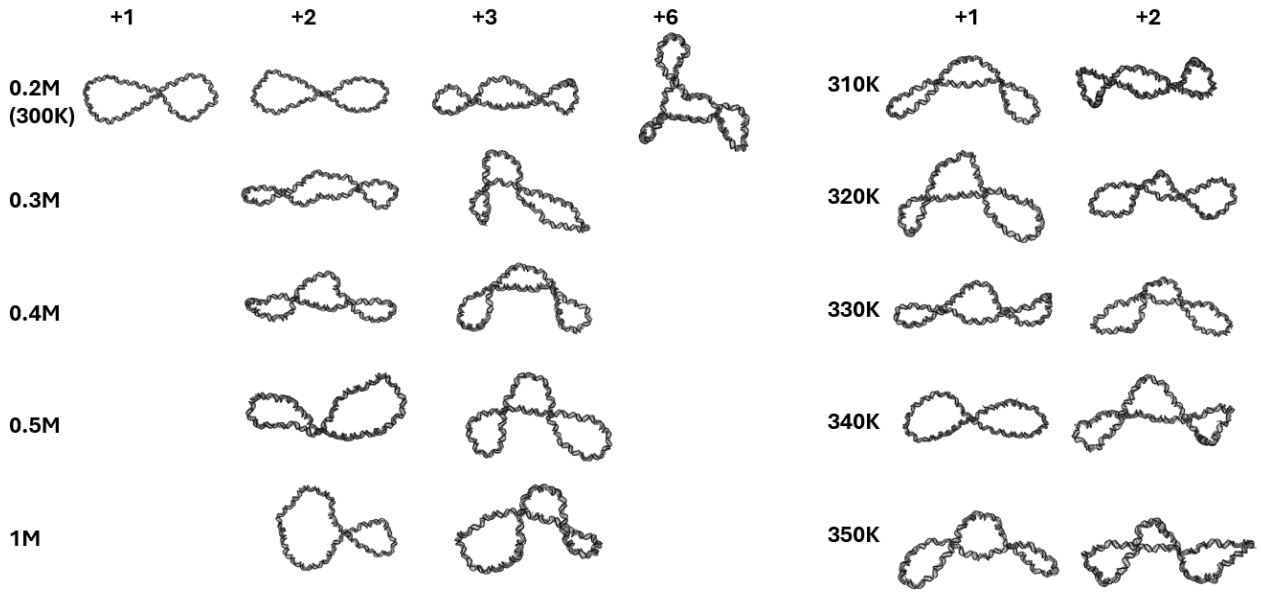

Figure A.9: As salt concentration and temperature increases, the minicircle buffers against large changes in twist resulting in little conformational change.
Representative snapshots from the simulations at each superhelical density and salt concentration.

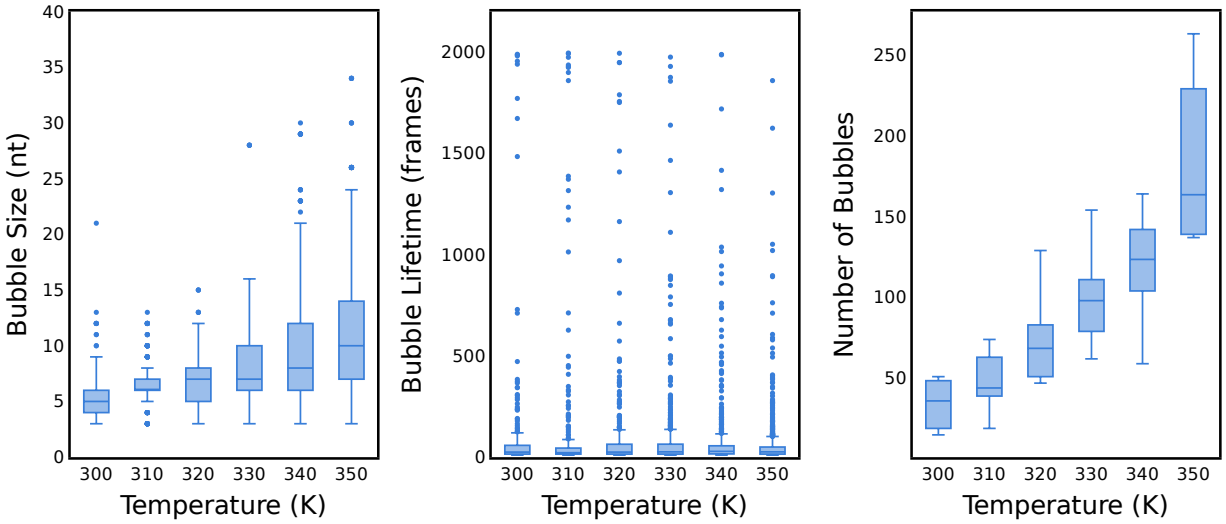

Figure A.10: Increasing the temperature increases the size and number of bubbles formed.
Box plots of the bubble sizes **(left)**, bubble lifetimes **(middle)**, and bubble count **(right)** compared across temperature. The data is combined for both Control339 and Mod339. The bubble size is plotted in terms of the number of nucleotides (nt) disrupted by the bubble. The bubble lifetimes are plotted by the number of frames each bubble persists for throughout the simulation after nucleation. There is a possible total of 2000 frames. The number of bubbles is the number of unique bubbles across all simulations.

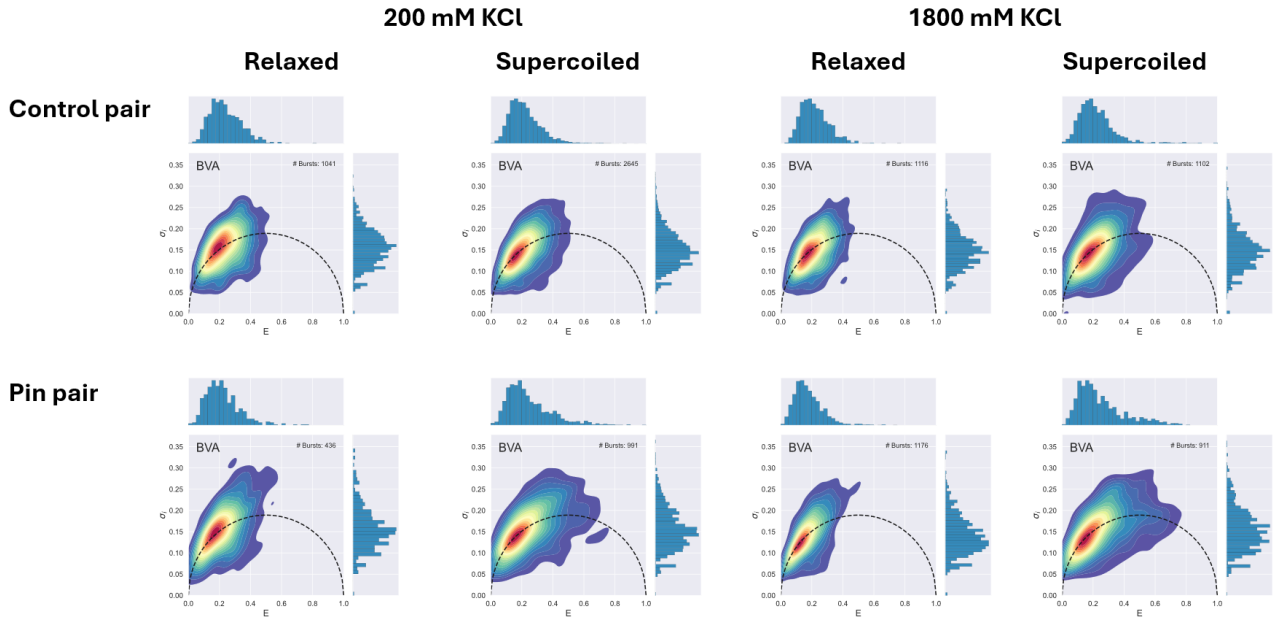

Figure A.11: All constructs were more dynamic at an increased ionic concentration; and, at each concentration, the supercoiled constructs were more dynamic than the relaxed topoisomers.
Burst variance analysis (BVA) plots of the experimental FRET data of the control dye pair **(top)** and pin dye pair **(bottom)** both in the relaxed **(left)** amd $\Delta Lk = -2$ states **(right)**. The data is shown for experiments at 200 mM KCl **(left)** and 1800 mM KCl **(right)**. The expected standard deviation from shot noise is plotted as a dashed semi-circle and the calculated variance from the bursts plotted as a heatmap. Thus, a large variance and more dynamic molecule can be identified by the heatmap laying largely above the plotted semi-circle.

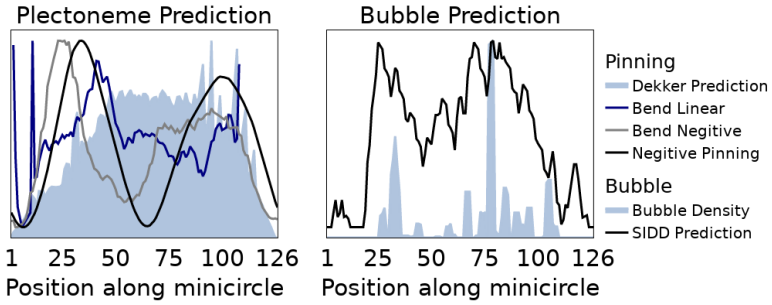

Figure A.12: There is no one region of highly-predicted plectoneme or bubble location in the 126-mer minicircle.
Left: The Dekker prediction of plectoneme positioning under positive supercoiling of the 126-mer is plotted (blue, filled) along with the flattened plectoneme propensity polar plot under positive supercoiling (black), the bend angle over all simulations of positively supercoiled minicircles (grey) and the bend angle over all repeats of the linear Control339 simulation (navy). **Right:** The SIDD prediction (black) of bubble density and the actual bubble densities over all simulations of the 126-mer under negative supercoiling (blue, filled) are plotted.

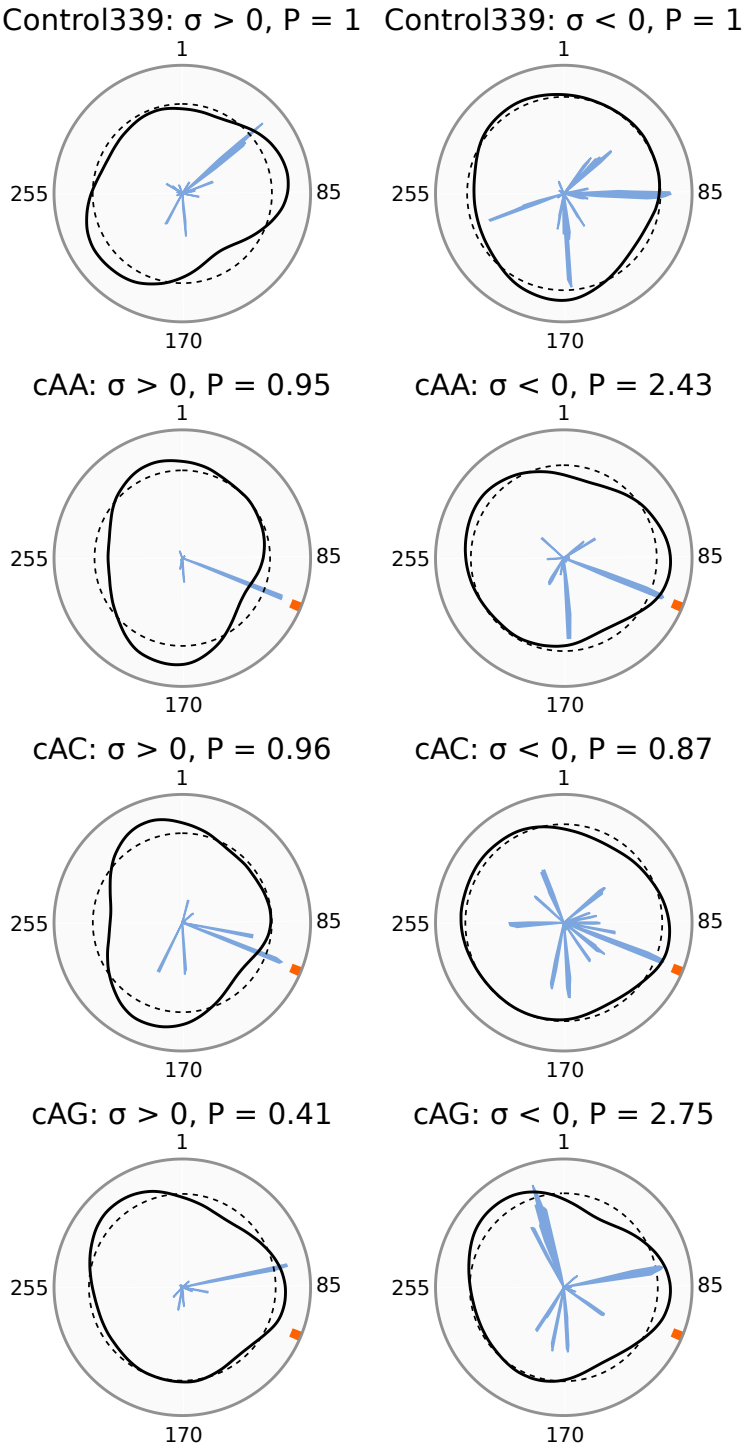

Figure A.13: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cAX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Control339 construct under the corresponding direction of supercoiling is equal to 1.

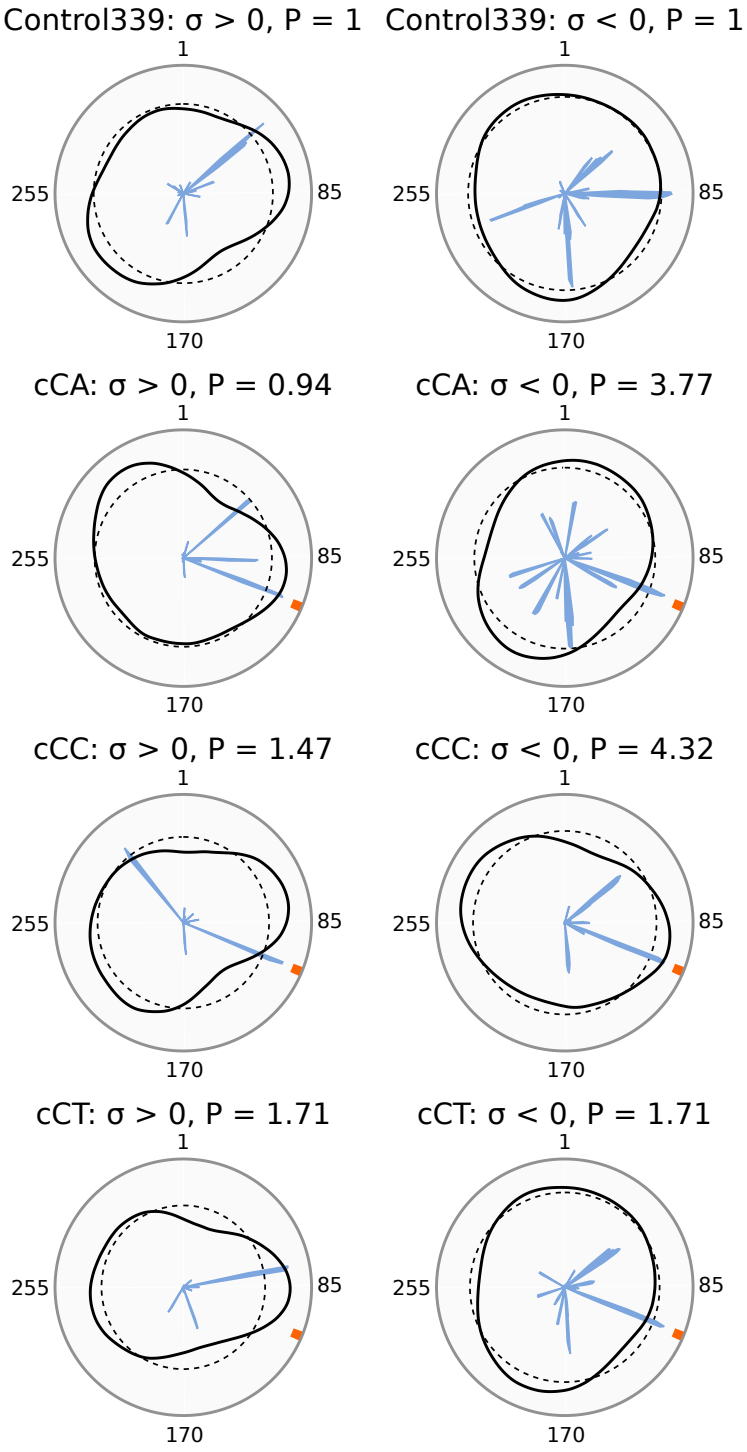

Figure A.14: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cCX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Control339 construct under the corresponding direction of supercoiling is equal to 1.

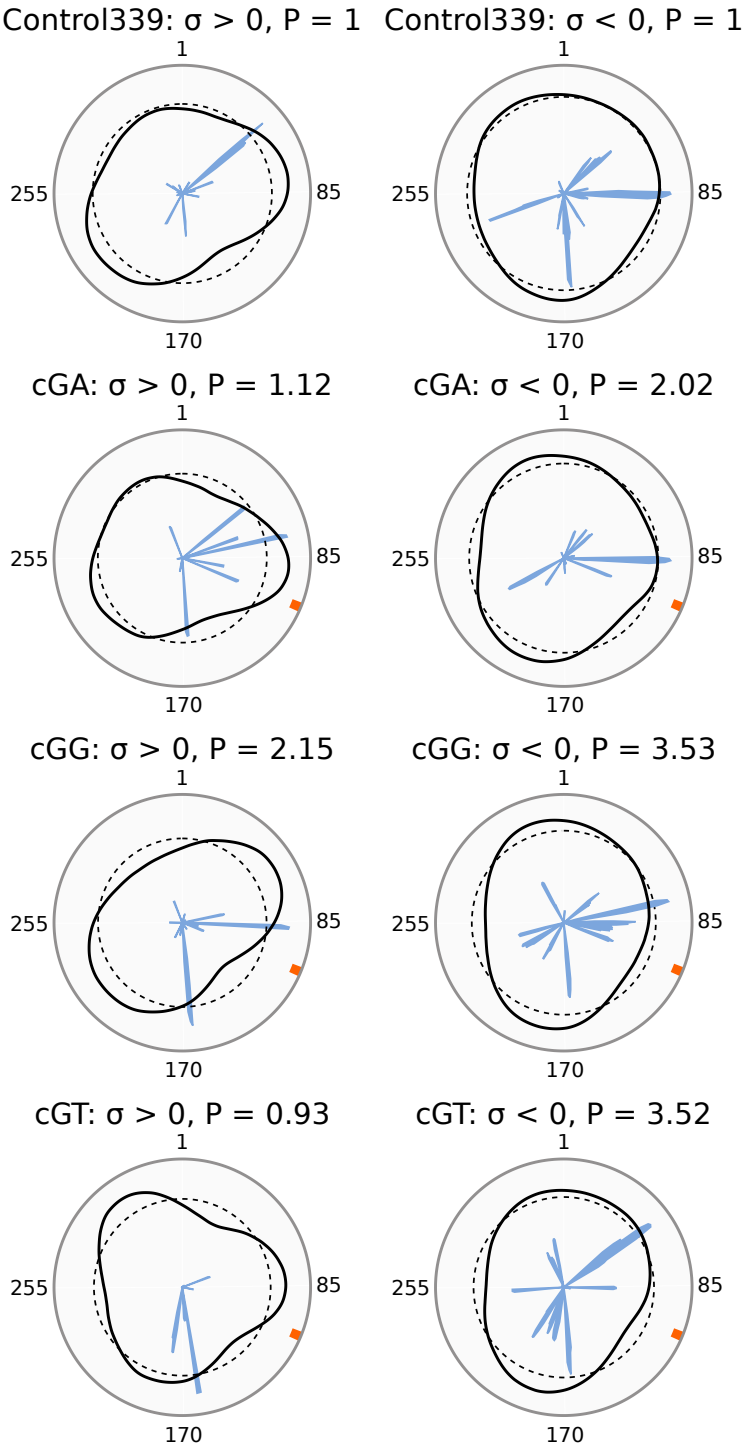

Figure A.15: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cGX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Control339 construct under the corresponding direction of supercoiling is equal to 1.

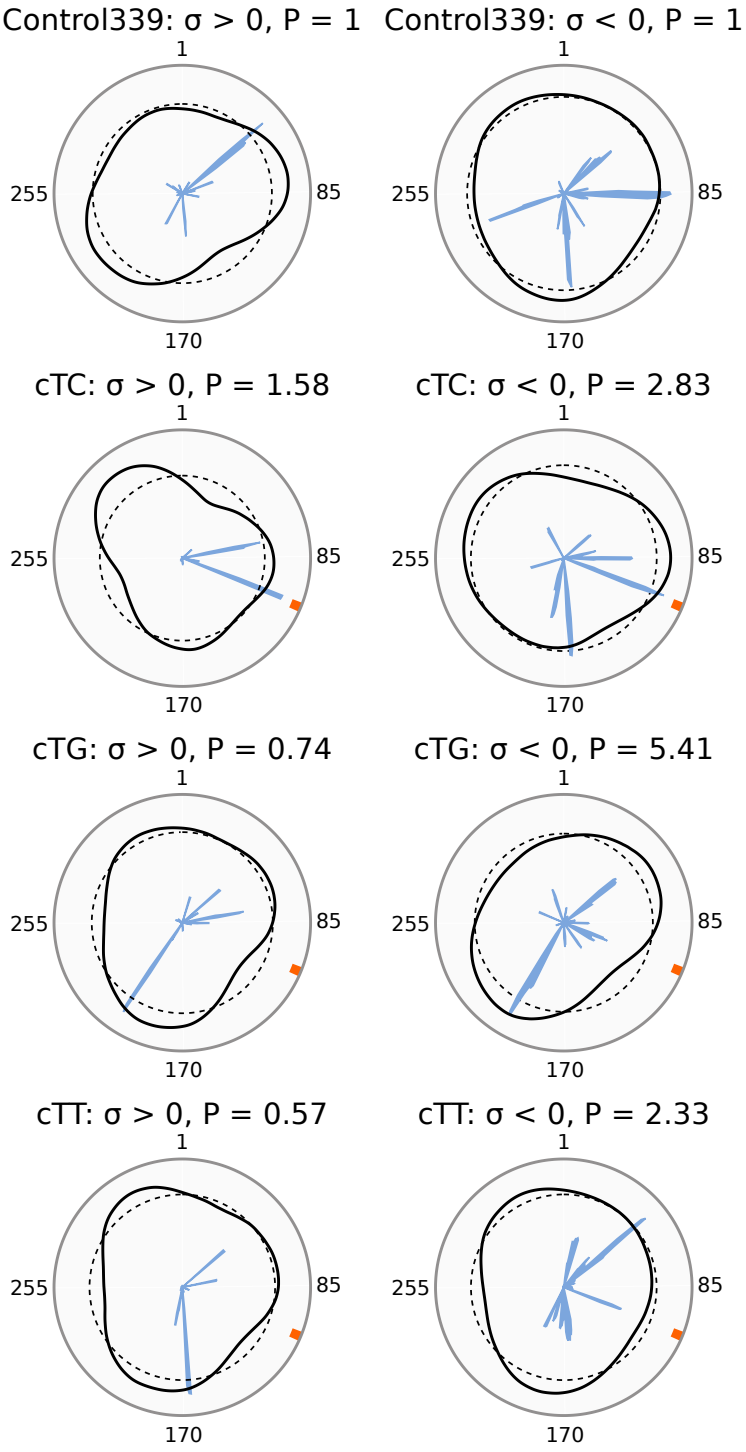

Figure A.16: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cTX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Control339 construct under the corresponding direction of supercoiling is equal to 1.

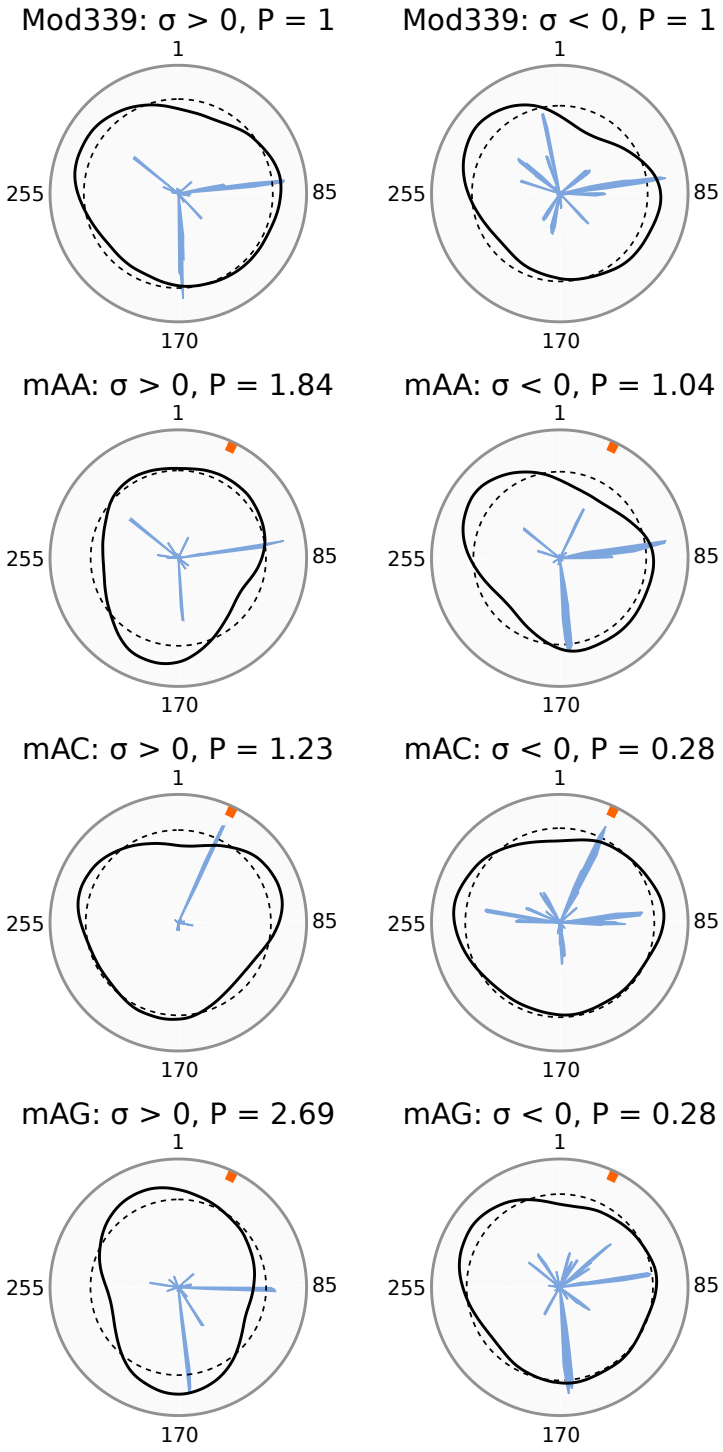

Figure A.17: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mAX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Mod339 construct under the corresponding direction of supercoiling is equal to 1.

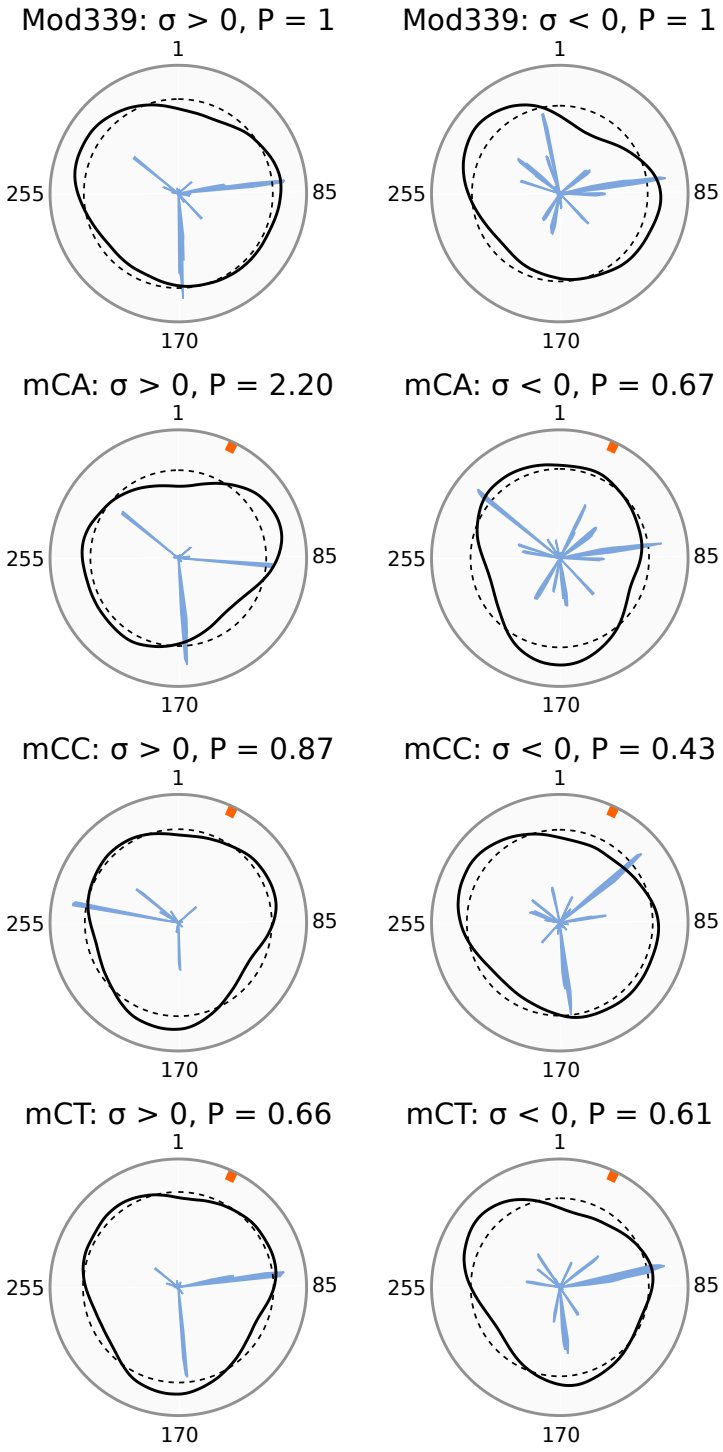

Figure A.18: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mCX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Mod339 construct under the corresponding direction of supercoiling is equal to 1.

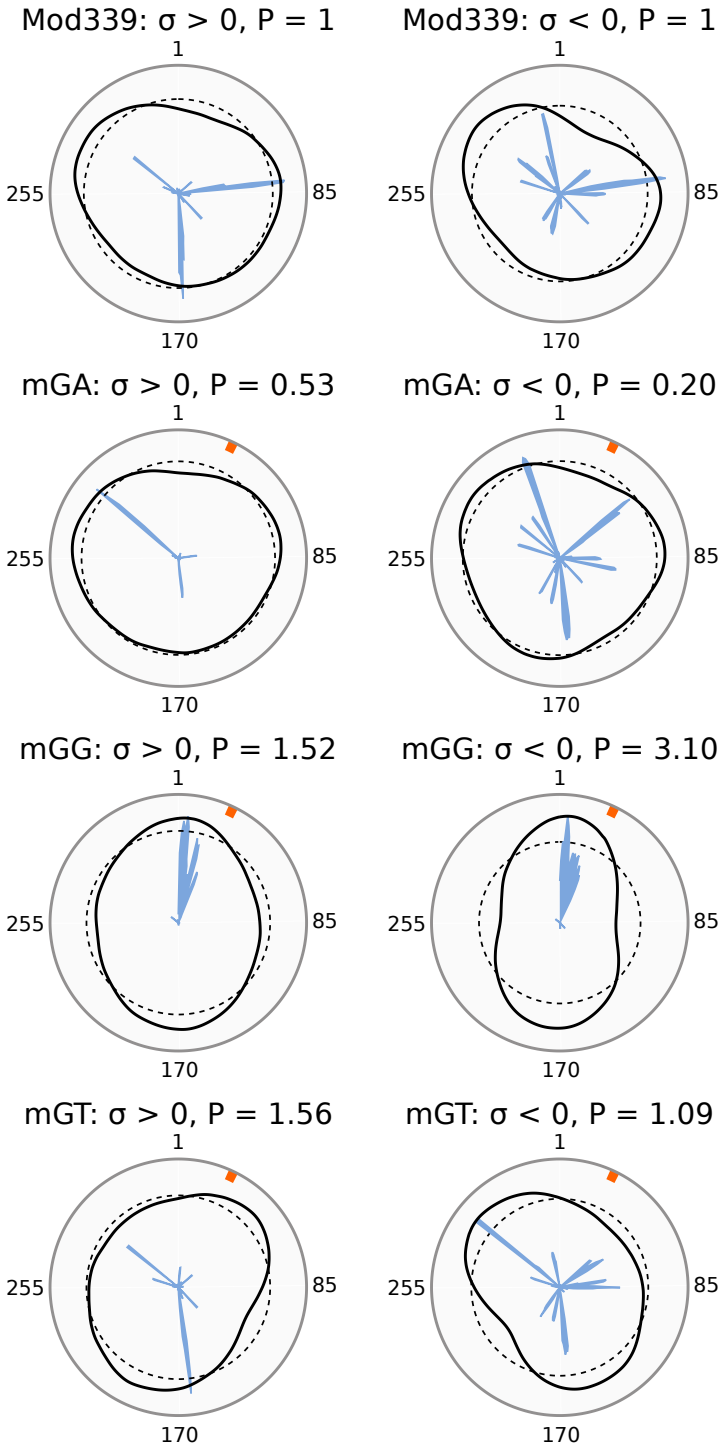

Figure A.19: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mGX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Mod339 construct under the corresponding direction of supercoiling is equal to 1.

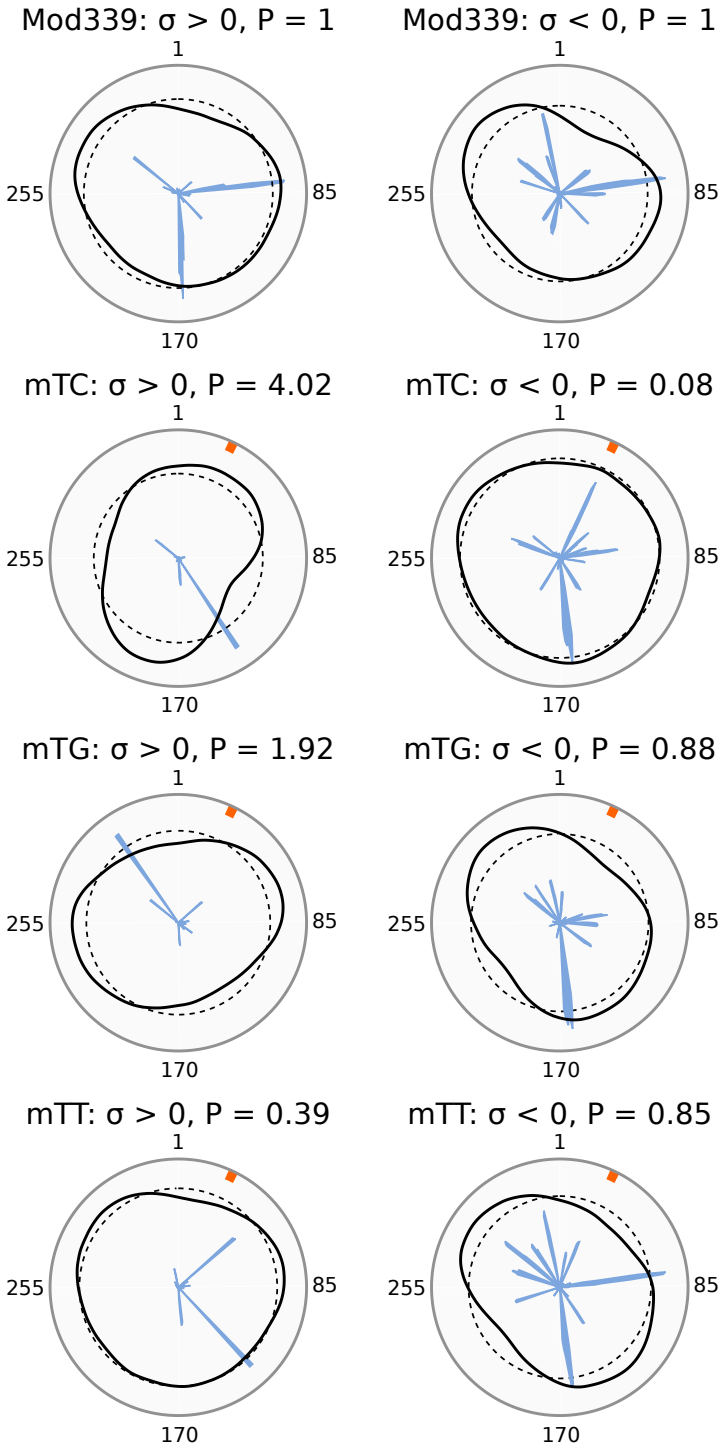

Figure A.20: The single mismatches have different plectoneme pinning propensities depending on the local sequence context and direction of supercoiling.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the mTX mismatches under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the Mod339 construct under the corresponding direction of supercoiling is equal to 1.

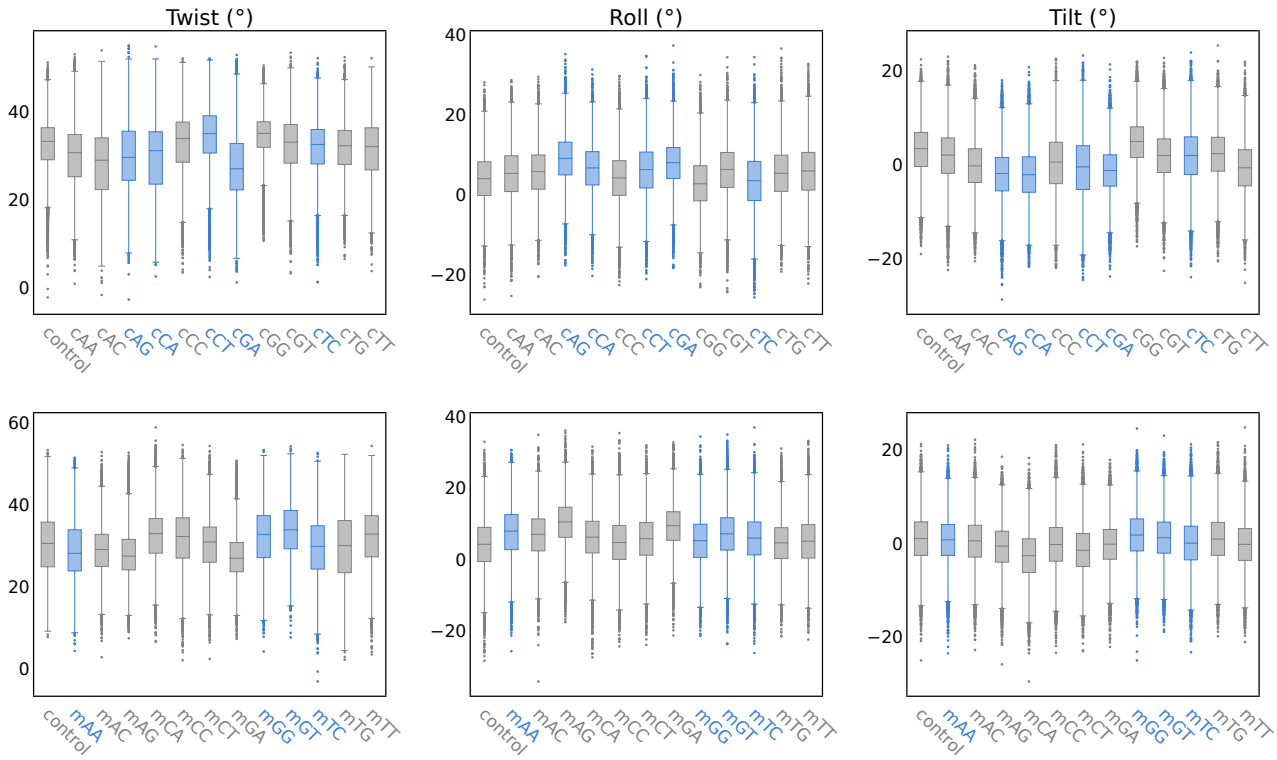

Figure A.21: The induced flexibility at the site of the mismatch is buffered beyond the nearest neighbours.
Box plots of the twist, roll, and tilt over the three replica simulations of the mismatch-containing Control339 **(top)** and Mod339 **(bottom)** sequences. Parameters are measured between the nearest and next nearest neighbour of the mismatches in the 5' direction. Mismatches that pin the plectoneme under positive supercoiling are depicted with blue boxes. All mismatch parameters are within the variance seen in the control simulations.

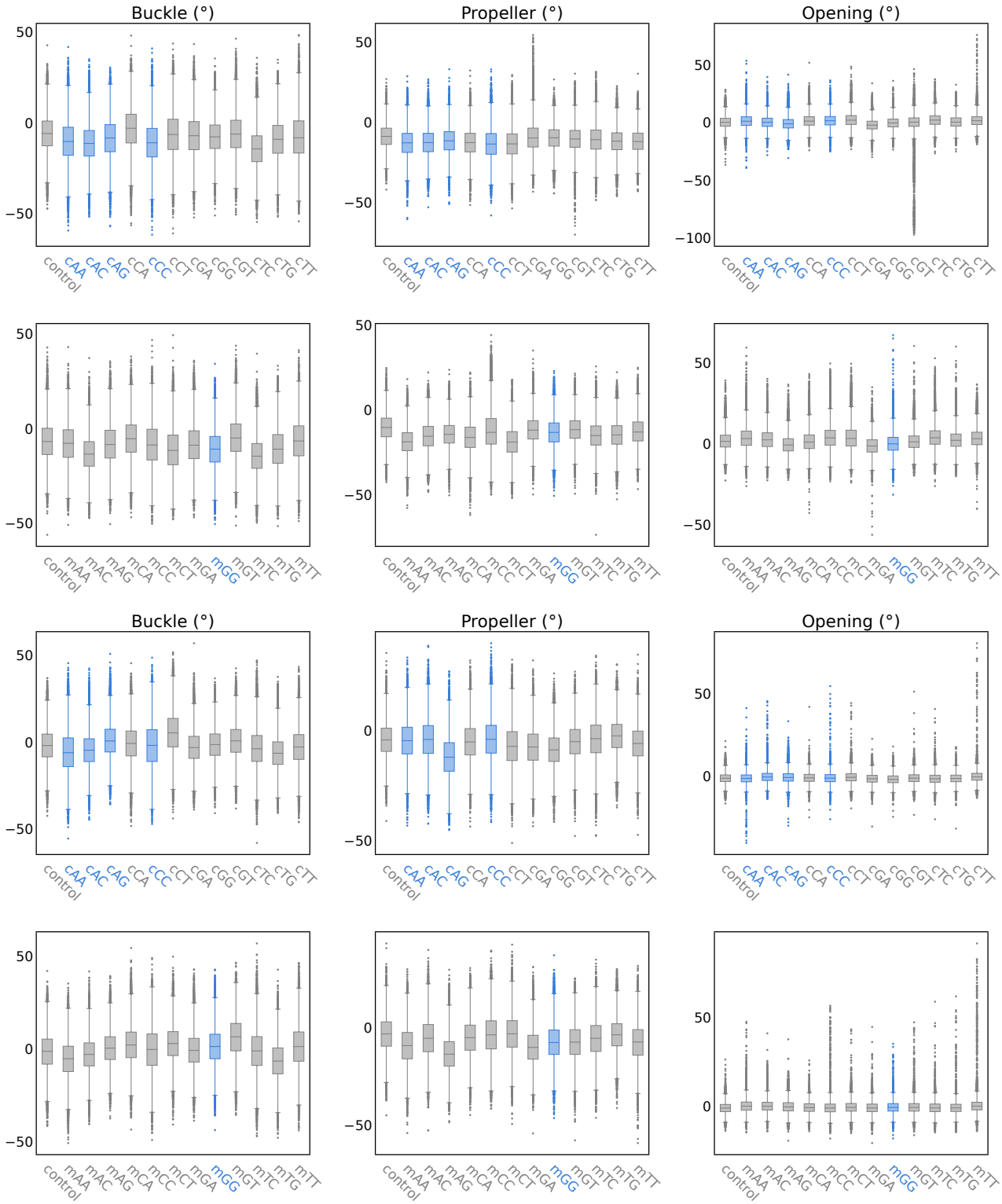

Figure A.22: The parameters describing bubble formation at the mismatch's nearest neighbours are not indicative of pinning under negative supercoiling.
Box plots describing the buckle, propeller, and opening of the nearest neighbour to the mismatch in the 5' **(top)** and 3' **(bottom)** direction for the mismatch-containing Control339 and Mod339 sequences. The parameters were calculated over the three replicas for each simulation and boxes are represented in blue where the mismatch pins the plectoneme under negative supercoiling in the corresponding simulation. All parameters are within the variance seen in the control simulations.

Table A.23: *The percentage of time within the simulations that there are the given number of hydrogen bonds between the base pairs either side of the mismatch. Non-Watson-Crick hydrogen bonding schemes were also considered. The rows that pin the plectoneme under negative supercoiling are shaded grey. Mismatch-1 refers to the base pair in the 5' direction of the mismatch, whilst Mismatch+1 refers to the base pair in the 3' direction of the mismatch.*

Mismatch		Mismatch -1, T:A			Mismatch +1, C:G					
		0	1	2	0	1	2	3	4	5
Control	Control55	0.01%	1.28%	98.71%	0.00%	0.01%	0.17%	56.54%	43.26%	0.02%
	Mod55	0.02%	3.05%	96.94%	0.00%	0.01%	0.25%	55.18%	44.53%	0.03%
AA	Control55	0.04%	2.73%	97.23%	0.00%	0.11%	0.34%	55.54%	44.00%	0.02%
	Mod55	0.09%	10.01%	89.91%	0.00%	0.06%	1.15%	51.40%	47.30%	0.09%
AC	Control55	0.03%	1.64%	98.33%	0.00%	0.07%	0.82%	53.87%	45.18%	0.06%
	Mod55	0.04%	5.78%	94.18%	0.00%	0.10%	0.92%	53.78%	45.15%	0.05%
AG	Control55	0.07%	2.13%	97.80%	0.00%	0.03%	0.41%	53.67%	45.76%	0.13%
	Mod55	0.02%	3.46%	96.51%	0.00%	0.01%	0.52%	54.40%	44.93%	0.15%
CA	Control55	0.03%	2.93%	97.04%	0.00%	0.01%	0.32%	54.43%	45.22%	0.02%
	Mod55	0.21%	6.50%	93.29%	0.00%	0.02%	0.44%	55.21%	44.30%	0.03%
CC	Control55	0.03%	2.08%	97.89%	0.00%	0.01%	0.43%	53.16%	46.37%	0.04%
	Mod55	0.15%	7.89%	91.96%	0.02%	0.04%	0.65%	54.19%	45.06%	0.04%
CT	Control55	0.04%	3.35%	96.61%	0.00%	0.01%	0.44%	52.00%	47.46%	0.10%
	Mod55	0.07%	9.68%	90.25%	0.00%	0.02%	0.53%	53.96%	45.45%	0.04%
GA	Control55	0.04%	1.11%	98.85%	0.00%	0.01%	0.25%	58.32%	41.39%	0.02%
	Mod55	0.03%	2.14%	97.83%	0.00%	0.01%	0.55%	58.39%	40.99%	0.06%
GG	Control55	0.02%	1.05%	98.93%	0.00%	0.01%	0.26%	57.07%	42.61%	0.06%
	Mod55	0.06%	2.21%	97.73%	0.00%	0.04%	0.64%	54.55%	44.72%	0.06%
GT	Control55	0.11%	8.97%	90.92%	0.00%	0.01%	0.25%	56.11%	43.60%	0.04%
	Mod55	0.09%	2.28%	97.63%	0.00%	0.03%	0.54%	54.05%	45.34%	0.05%
TC	Control55	0.02%	2.78%	97.20%	0.00%	0.01%	0.20%	54.65%	45.09%	0.05%
	Mod55	0.16%	7.41%	92.43%	0.00%	0.01%	0.45%	53.73%	45.76%	0.05%
TG	Control55	0.01%	1.33%	98.66%	0.00%	0.00%	0.13%	55.63%	44.22%	0.02%
	Mod55	0.01%	3.60%	96.39%	0.00%	0.00%	0.27%	53.68%	46.01%	0.03%
TT	Control55	0.08%	1.50%	98.42%	0.00%	0.03%	0.43%	52.61%	46.87%	0.05%
	Mod55	0.01%	4.18%	95.81%	0.00%	0.05%	0.83%	49.95%	49.05%	0.12%

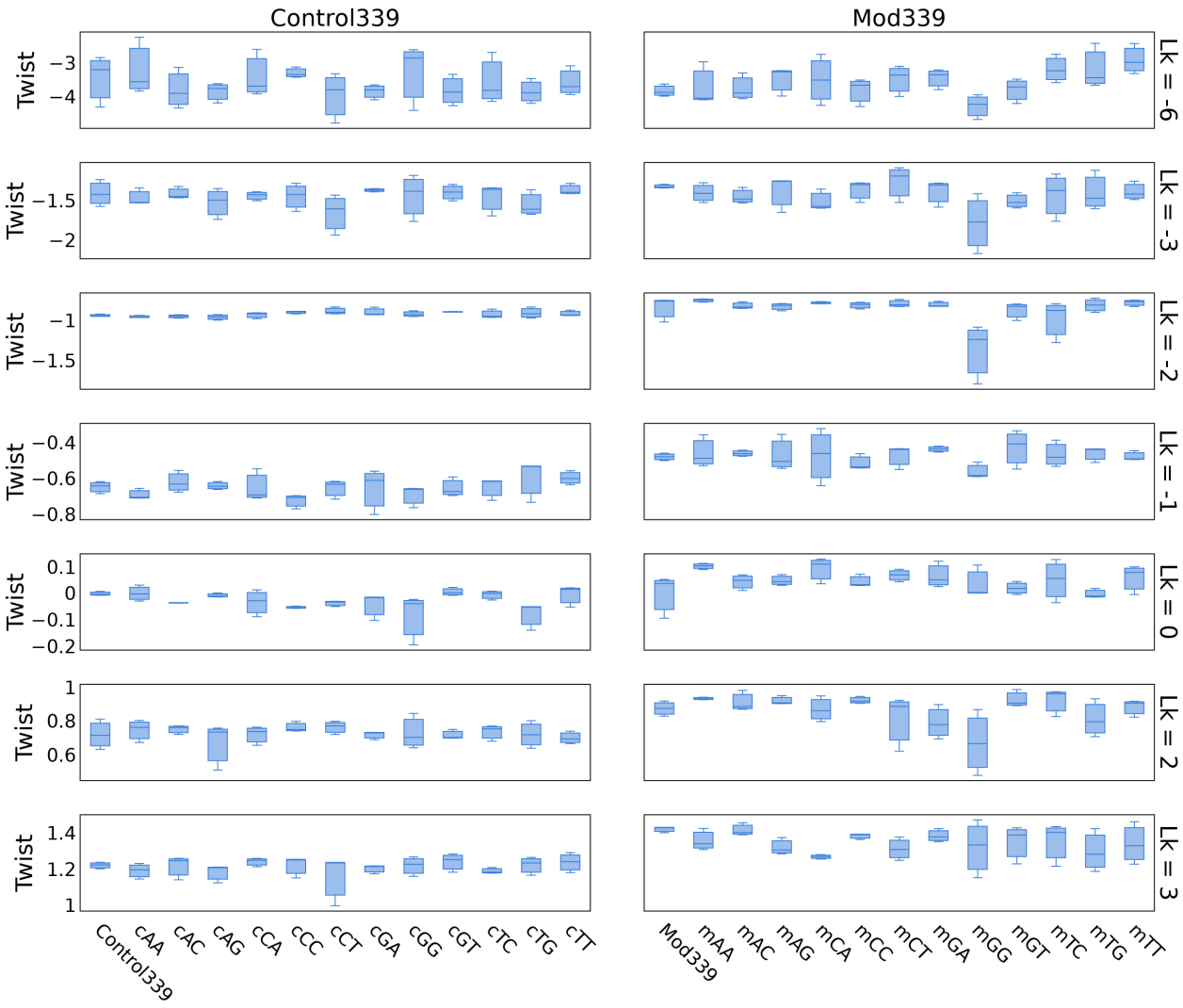

Figure A.24: In general, the twist at each superhelical density is sequence- and mismatch-independent
The twist over all simulations at each linking number of the mismatched constructs in Control339 **(left)** and Mod339 **(right)**. All sequences generally have the same value of twist despite the twist absorption varying by mismatch. The only appreciable difference is the gGG mismatch at highly negative superhelical densities.

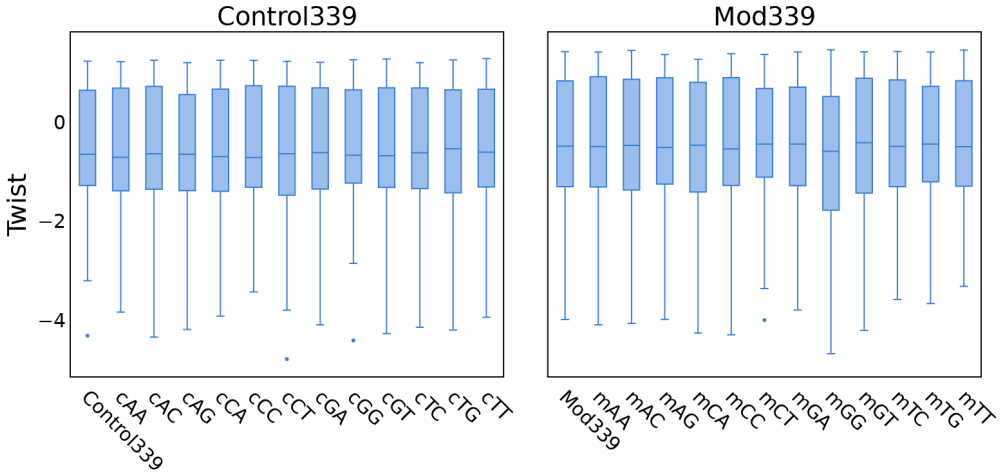

Figure A.25: DNA globally buffers against local changes in conformation, such as twist absorption. The twist over all simulations of all linking numbers for the mismatch-containing Control339 **(left)** and Mod339 **(right)**. All sequences have the same value of twist despite the twist absorption varying by mismatch.

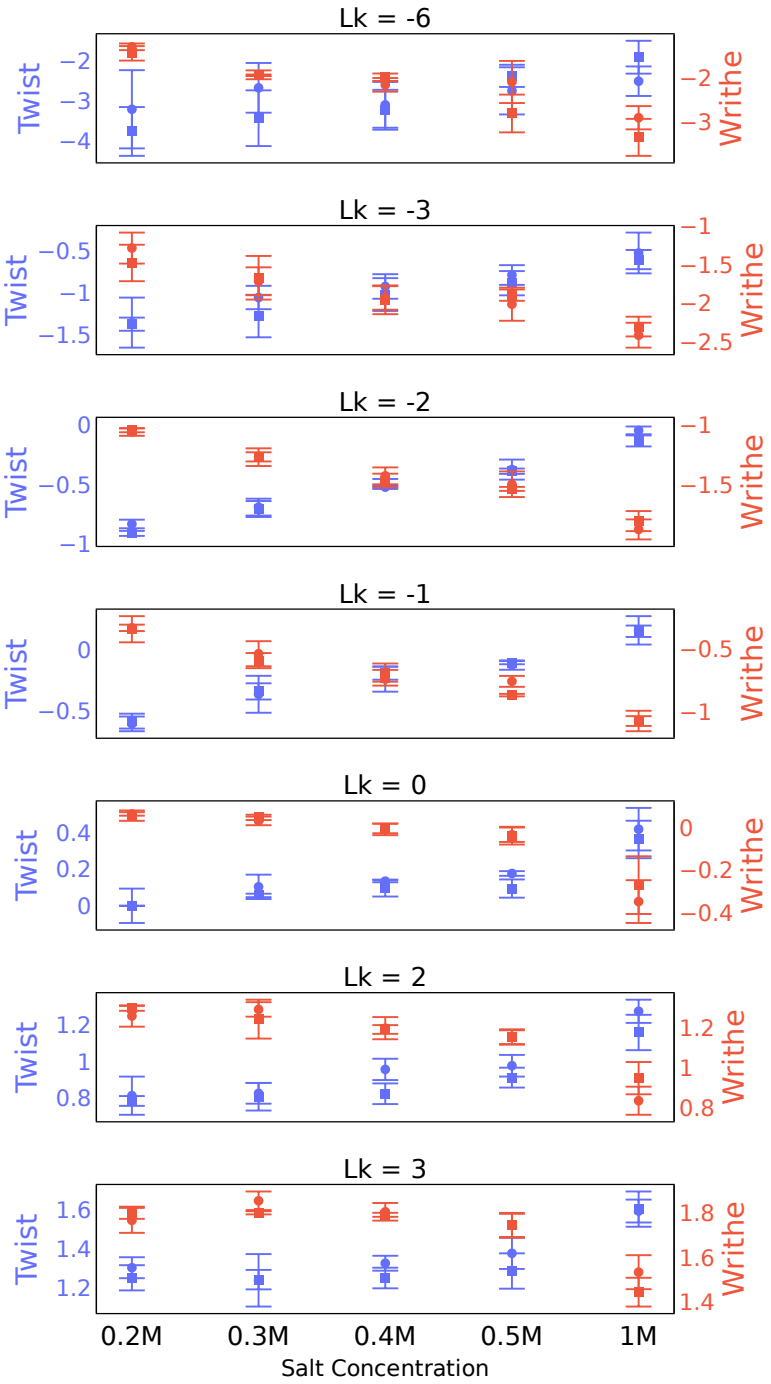

Figure A.26: At increasing salt, the twist becomes more positive and the writhe decreases as a result.
The average twist (blue) and writhe (red) are plotted for the cAC (square) and cGG (circle) mismatches at increasing salt concentration. The error bars represent the standard deviation of the twist and writhe over all the three replicas. The plots are divided by superhelical density with the most negatively supercoiled plotted at the top as the linking number becomes increasingly positive downwards.

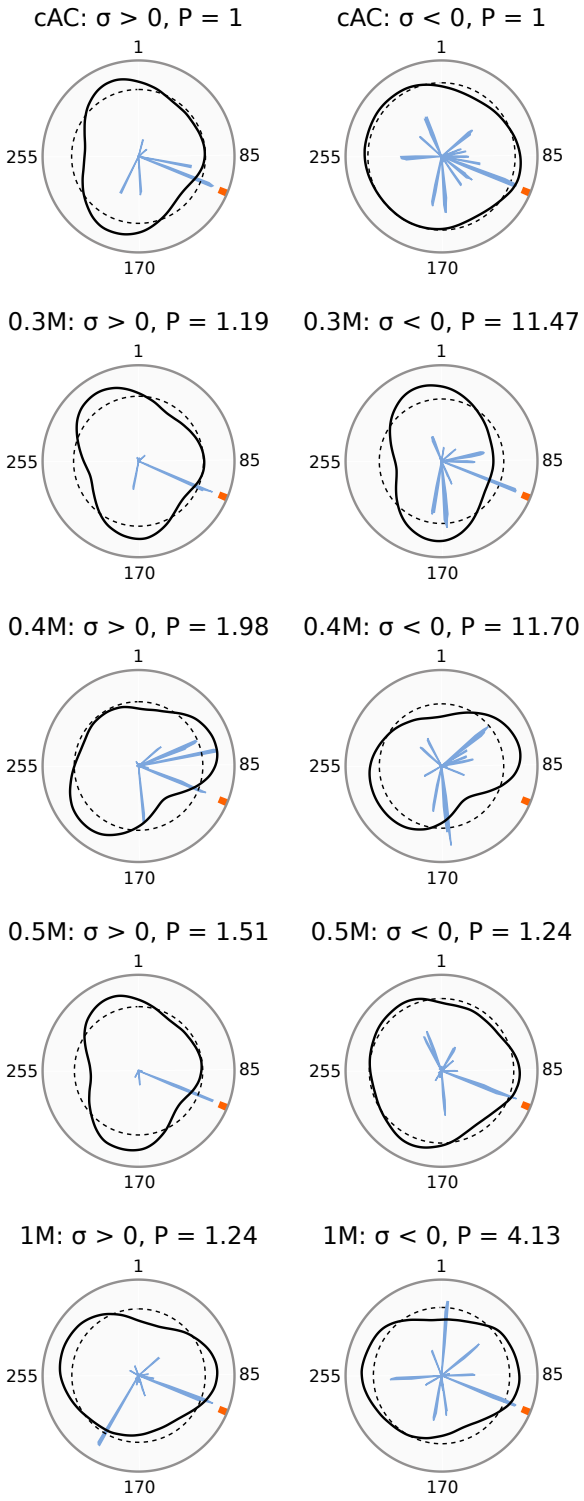

Figure A.27: Increasing the salt concentration increases the local pinning propensity of cAC; however, the overall pinning is not driven by bubble localisation.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cAC mismatch at increasing salt concentration under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct under is equal to 1.

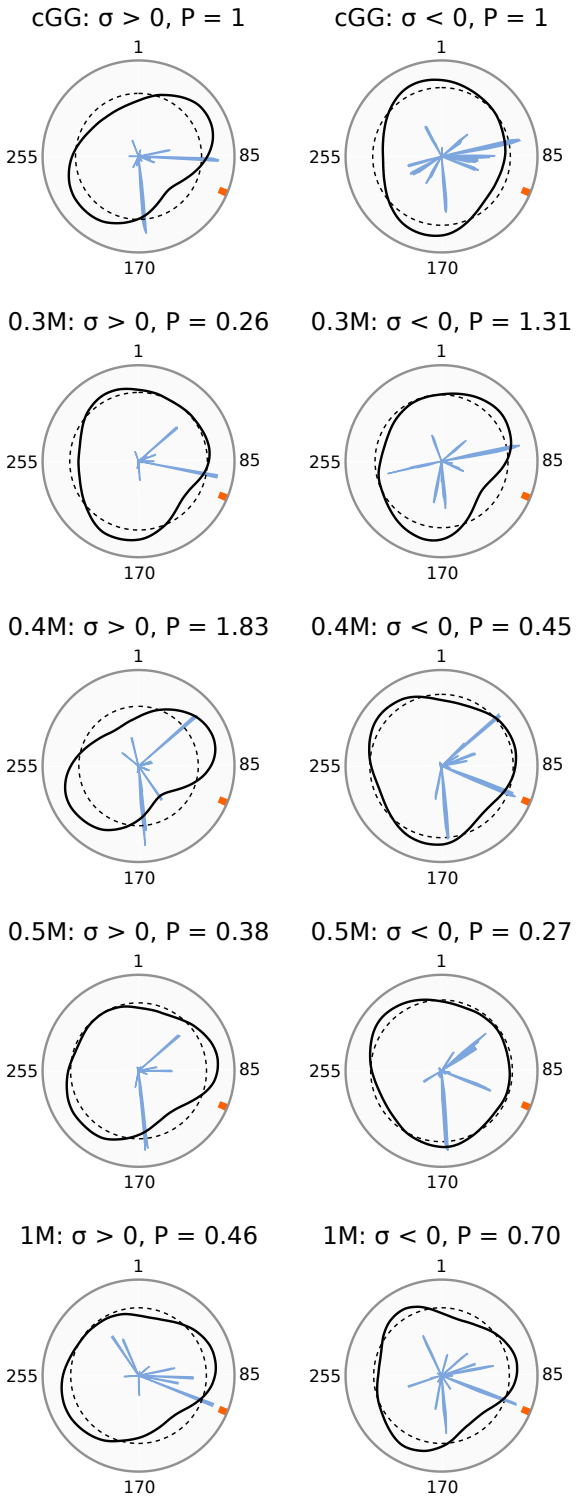

Figure A.28: Increasing the salt concentration increases the pinning propensity of cGG; however, the mismatch still fails to pin the plectoneme overall.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the cGG mismatch at increasing salt concentration under positive **(left)** and negative **(right)** supercoiling. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

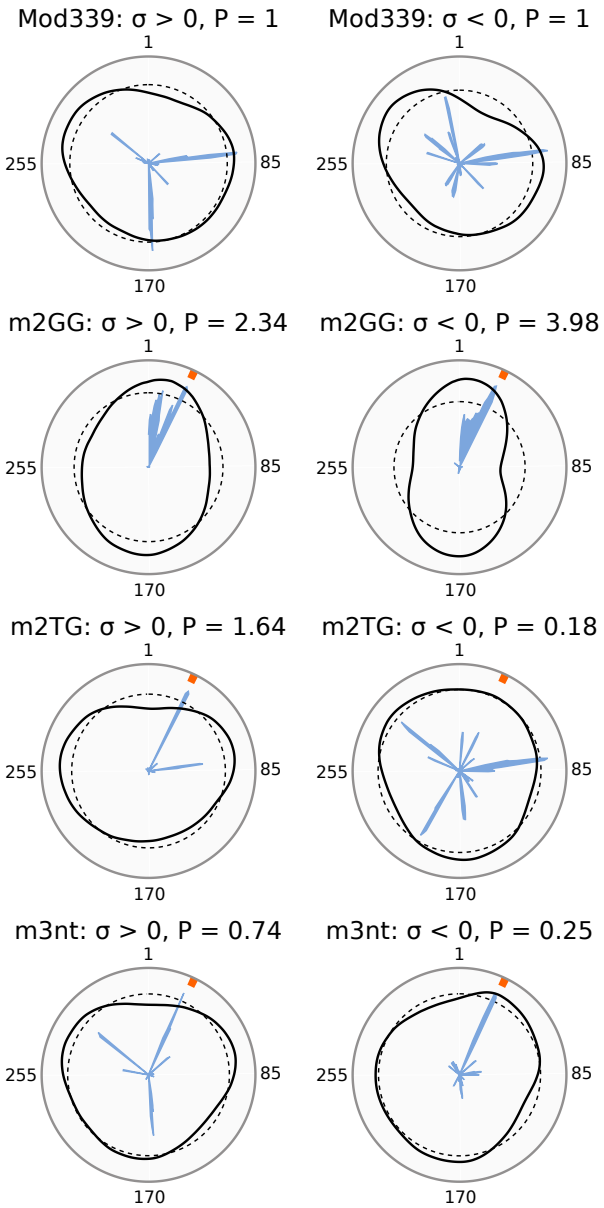

Figure A.29: The GG/GG double mismatch pins the plectoneme under both directions of supercoiling, whilst the CTG/CTG triple mismatch pins the plectoneme only under negative supercoiling in Mod339.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the two double and triple mismatches under positive **(left)** and negative **(right)** supercoiling in the Mod339 sequence. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

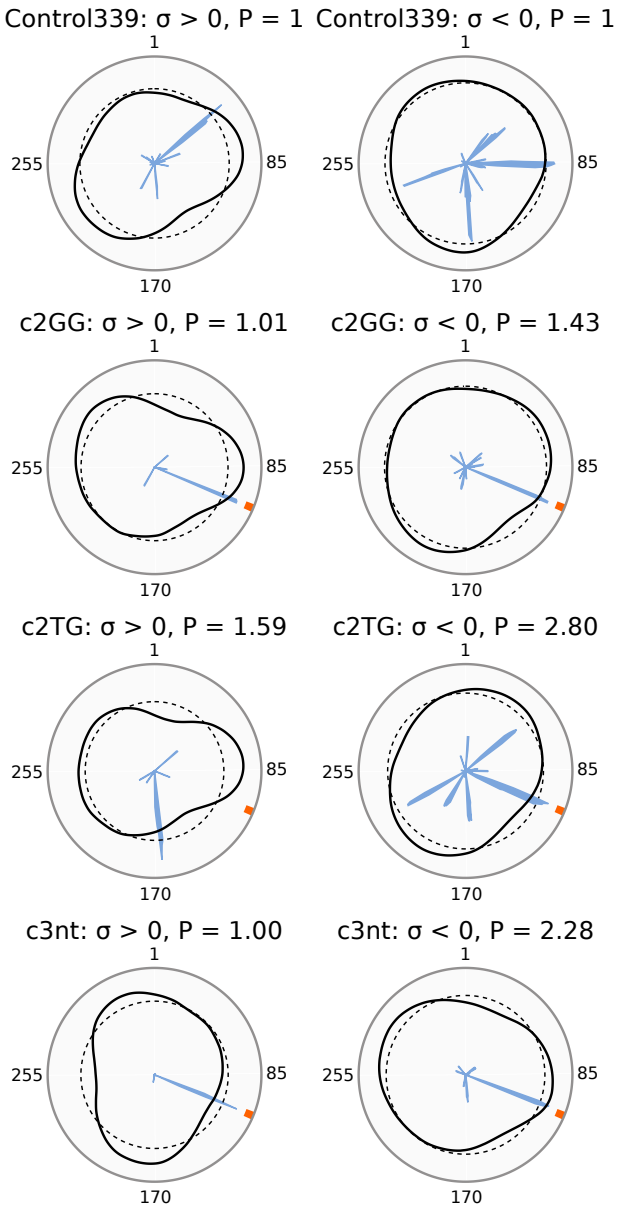

Figure A.30: The GG/GG double mismatch pins the plectoneme under positive supercoiling, whilst the CTG/CTG triple mismatch pins the plectoneme under negative supercoiling in Control339.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the two double and triple mismatches under positive **(left)** and negative **(right)** supercoiling in the Control339 sequence. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

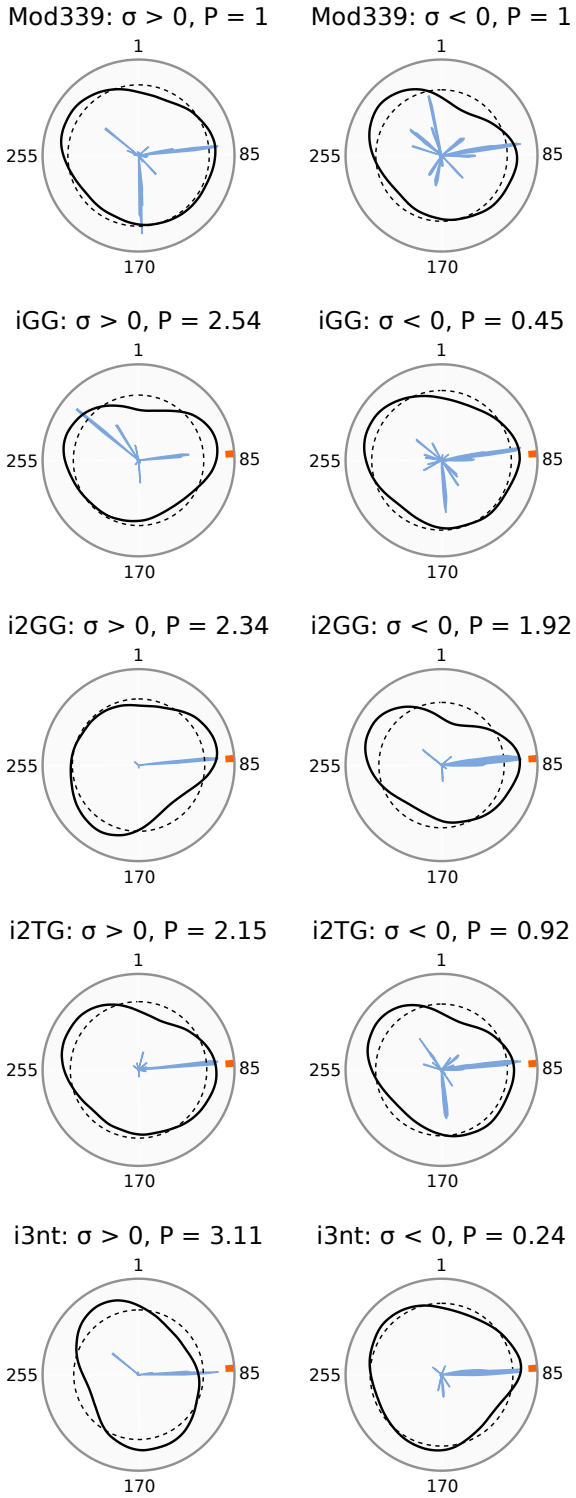

Figure A.31: All mismatches except, surprisingly, the CTG/CTG triple mismatch under positive supercoiling pin the plectoneme when inserted directly into a region of high pinning propensity.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the single GG, two double and one triple mismatches under positive **(left)** and negative **(right)** supercoiling in the Mod339 sequence. The mismatches are inserted directly into a region of high plectoneme density. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

Table A.32: *The individual morse potentials, D, and elastic constants, k, for each base pair and step of the double and triple mismatches in the nearest neighbour sequence context. The D and k values are summed over each base step in the given sequence context. The raw data is taken from data from Oliveira et al. [472]. The rows where the mismatch pins the plectoneme under negative supercoiling in either sequence are shaded grey.*

Mismatch		D_1	D_2	D_3	D_4	D_5	$\sum D$ / meV	$k_{1\text{-}2}$	$k_{2\text{-}3}$	$k_{3\text{-}4}$	$k_{4\text{-}5}$	$\sum k$ / eV nm^{-2}
GG/GG	Control339	33.71	19.97	19.97	33.71		107.36	2.81	1.95	2.81		7.57
	Mod339	33.71	19.97	20.96	69.53		144.17	2.81	1.77	2.21		6.79
GT/GT	Control339	69.53	4.06	19.94	69.53		163.06	0.21	1.45	2.6		4.26
	Mod339	33.71	2.1	19.94	69.53		125.28	2.45	1.46	2.55		6.46
CTG/CTG	Control339	33.71	2.08	4.56	19.94	69.53	129.82	1.27	0.77	1.61	2.55	6.2
	Mod339	33.71	2.08	4.56	19.94	69.53	129.82	0.29	0.77	1.61	2.55	5.22

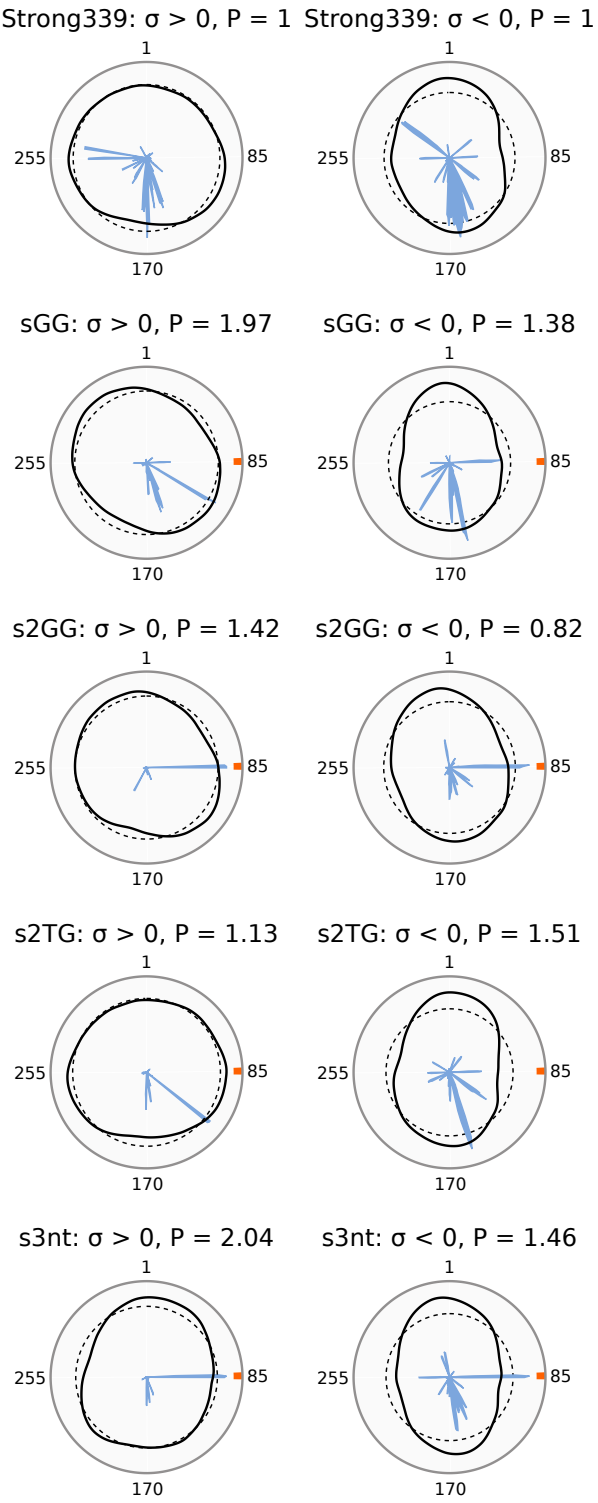

Figure A.33: The TG/TG double mismatch pins the plectoneme under positive supercoiling in the Strong339 sequence despite little local bubble formation.
Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the single GG, two double and one triple mismatches under positive **(left)** and negative **(right)** supercoiling in the Strong339 sequence. The mismatches are inserted directly into a region of high plectoneme density. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

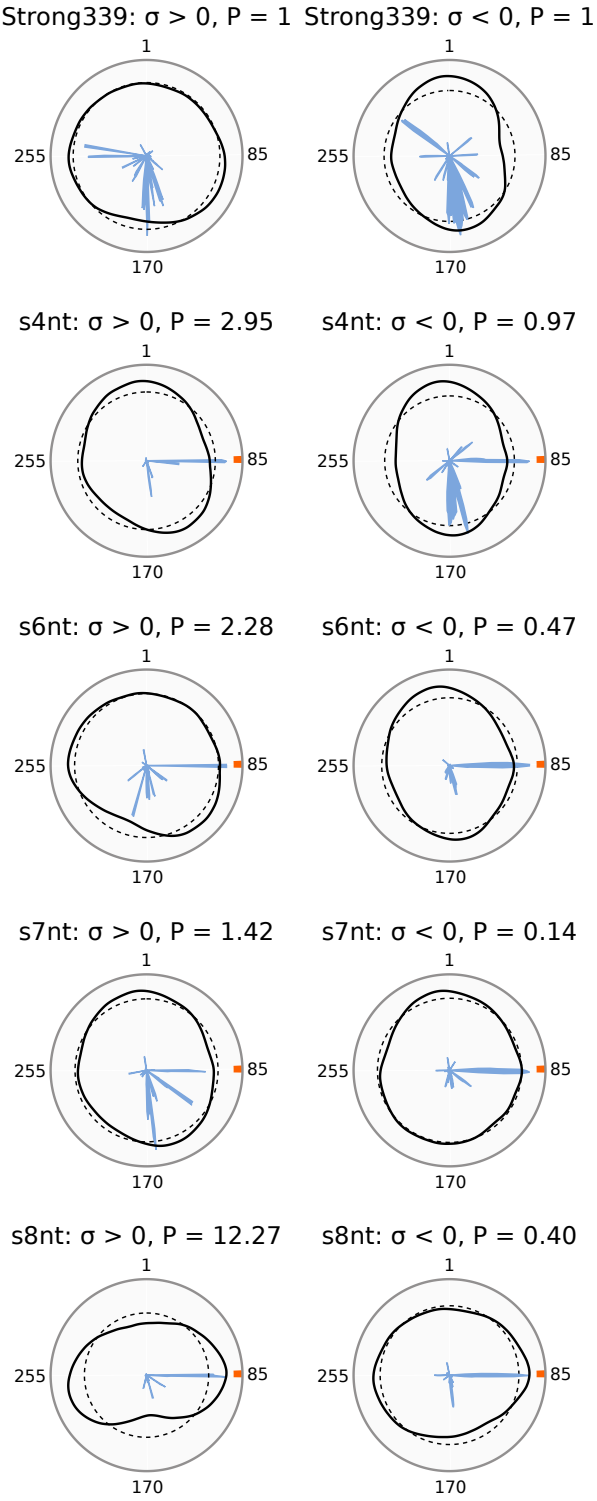

Figure A.34: A consecutive run of eight mismatches is required to overcome the strongly-pinning bubble region in Strong339 and pin the plectoneme under negative supercoiling. Polar plots of the plectoneme propensity (black), average plectoneme propensity (black, dashed), and bubble densities (blue, filled) of the four, six, seven, and eight tandem mismatches under positive **(left)** and negative **(right)** supercoiling in the Strong339 sequence. The mismatches are inserted directly into a region of high plectoneme density. The densities are normalised such that the maximum value for both parameters is 1. The location of the mismatches is represented by an orange band and the P values are normalised such that the P value for the corresponding undamaged construct is equal to 1.

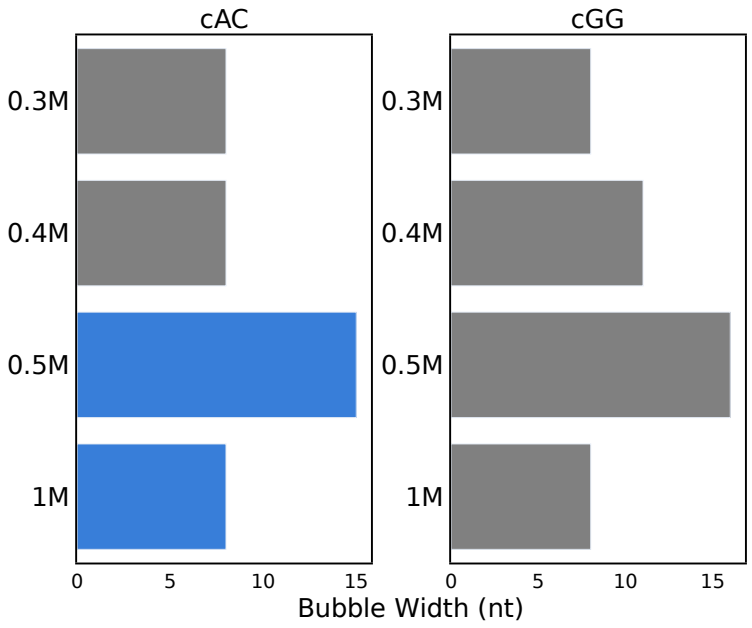

Figure A.35: The size of the bubble at the mismatch does not correlate to pinning propensity nor salt concentration.
The width of the bubble density peak at the mismatch site under negative supercoiling for the cAC and cGG mismatches at increased salt concentrations. The width is given in terms of the number of nucleotides (nt) that the bubble spans. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

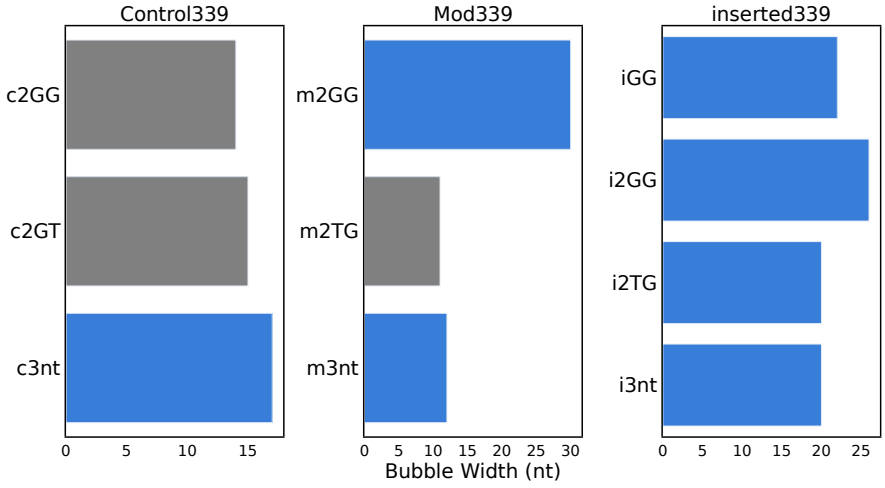

Figure A.36: The size of the bubble at the mismatch does not linearly increase with the number of consecutive mismatches.
The width of the bubble density peak at the mismatch site under negative supercoiling for the multiple mismatches in the Control339, Mod339, and inserted339 sequences. The width is given in terms of the number of nucleotides (nt) that the bubble spans. The blue bars represent the simulations where the mismatch pins the plectoneme under negative supercoiling.

A.3 Supplementary Methods: Simulation Input Files

File A.1: Implicit Solvent Production Run

```
&cntrl
        ntf=2, ntc=2, ntb=0, cut=1000,
        nstlim=50000000, dt=0.002,
        ntpr=5000, ntwx=5000, ntwe=5000, ntwr=5000,
        temp0=300.0, ntt=3, gamma_ln=0.01,
        imin=0, irest=1, ntx=5,
        igb=8, gbsa=0, saltcon=0.2,
        ntr=0,
&end
```

File A.2: Explicit Solvent Production Run (Minicircles)

```
 &cntrl
        ntf=2, ntc=2, ntb=2, cut=8.0,
        nstlim=500000, dt=0.002,
        ntwx=5000, ntwe=5000, ntwr=5000, ntpr=5000,
        temp0=300.0, ntt=1,
        ntp=1,
        imin=0, irest=1, ntx=5,
 &end
```

File A.3: Explicit Solvent Production Run (55-mers)

```
 &cntrl
        ntf=2, ntc=2, ntb=2, cut=12.0,
        nstlim=50000000, dt=0.002,
        ntpr=5000, ntwx=5000, ntwe=5000, ntwr=5000,
        temp0=300.0, ntt=3, gamma_ln=1.0,
        ntp=1,
        imin=0, irest=1, ntx=5,
 &end
```

ntf Force evaluation
= 2 bond interactions involving H-atoms omitted (as SHAKE used)

ntc Flag for SHAKE to perform bond length constraints
=2 bonds involving hydrogen are constrained

ntb Periodic boundaries flag
= 0 No periodic boundary conditions (due to implicit solvent)
= 2 Constant pressure

cut The non-bonded cutoff in Angstroms.

nstlim The number of molecular dynamics steps to be performed

dt The time step in picoseconds

ntpr Frequency (in steps) of energy information printing to the mdout file

ntwx Frequency (in steps) of co-ordinates being written to the mdcrd file

ntwe Frequency (in steps) of the energies and temperatures being written to the mden file.

ntwr Frequency (in steps) of the restart file being updated.

temp0 The reference temperature for the simulation in K.

ntt Flag to specify the thermostat
=1 Constant temperature, using the weak-coupling algorithm
=3 Langevin dynamics

gamma_ln Collision frequency in ps^{-1}

ntb Flag for constant pressure dynamics
= 1 MD with isotropic position scaling

imin Flag to run minimization
=0 Run molecular dynamics with no minimization (production run)

irest Flag to restart a simulation
=1 Restart reading coordinates and velocities from restart file

ntx Option to read starting coordinates, velocities, and box size
=5 Coordinates and velocities read from input file

igb GB model flag
=8 GBneck2

gbsa Option to carry out GB/SA simulations
=0 Surface area not included in solvation term

saltcon Salt concentration (M) in implicit solvent simulation

ntr Flag for Cartesian restraints
=0 No atoms restrained

A.4 Supplementary Results: Coil Formation

In ∼12% of our simulations of the most supercoiled topoisomer, $\Delta Lk = -6$, we observed the formation of a tightly wound coil-like structure (fig. A.37). We could not find any evidence of similar behaviour reported in the literature, so we concluded that it was likely an artifact of high strain, and possibly due to being simulated in implicit solvent.

Once formed, the coils did not unwind. Furthermore, there was no apparent trend to which simulations they formed within (table A.39), and when they did form, they ranged in severity. The formation of the coil was identified by visual inspection of the most populated cluster. Once identified, the whole trajectory was inspected to confirm that the coil did not unwind. If this were the case, the simulation was repeated, and the results of the coil-forming simulation were discarded. This was due to the negative impact this structure would have on bubble statistics and, potentially, bias the conformations formed and the sequence-positioning of plectonemes and bubbles. No coils were found to unwind in any of the simulations, regardless of their severity.

Structurally, from the least severe coil formation, we can see that base pairing is maintained throughout the column of the structure whilst bases flip to accommodate the sharp bend at the tip of the column (fig. A.37). This would likely be more favourable in implicit solvent due to the absence of shielding, screening ions.

The column of the coil is a quadruplex-like structure where each layer is formed of two base

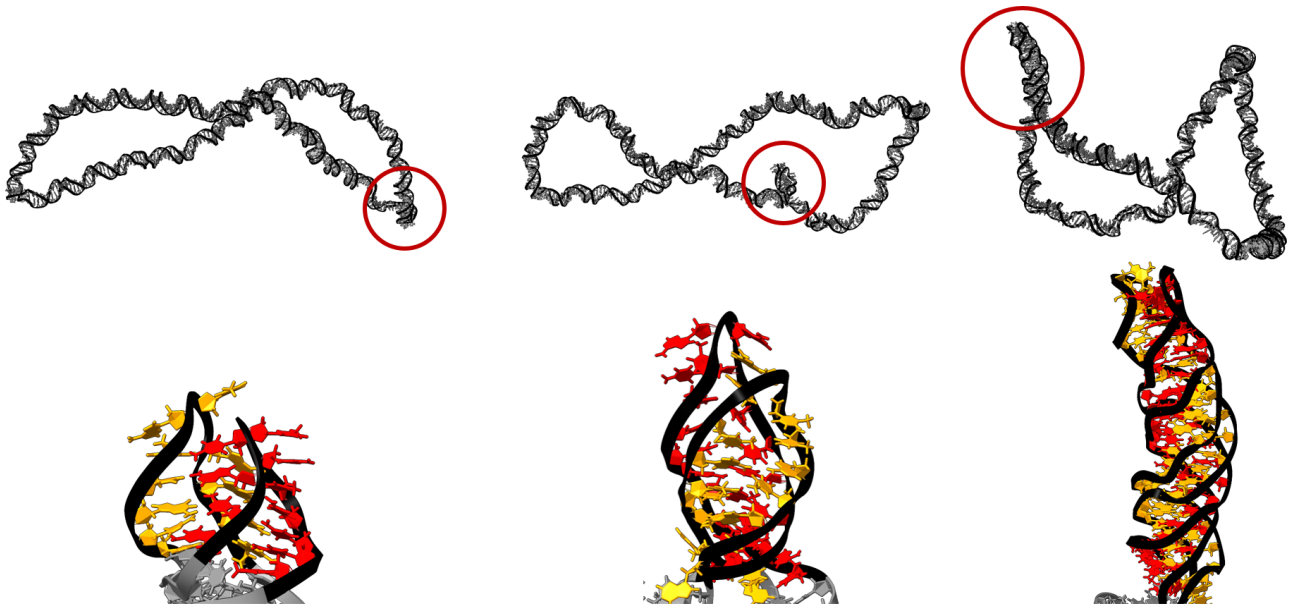

Figure A.37: Coiled structures formed in some simulations of the $\Delta Lk = -6$ topoisomer. Snapshots of the coil structures from the simulations. The top shows the overall structure of the minicircle with coil formation, whilst the bottom shows zoomed-in structures of the coil itself. The coil regions are circled. The two strands of the bottom structures have been coloured separately for clarity. The severity of the coil formed increases from left to right, as shown by the increasing high of the deformation.

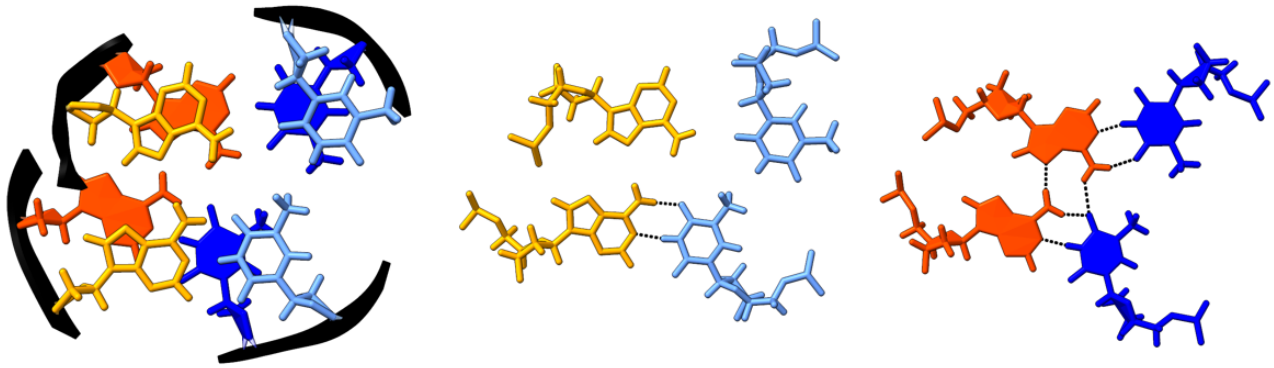

Figure A.38: Some stacks in the column are able to form hydrogen bonds outside of their base pairs.
Snapshots of eight base pairs within the coil column from the simulations. The same strand is coloured yellow or orange, and the other strand is coloured blue. Looking down at the base pairs, the bottom four base pairs have solid ring-filled structures for clarity. In all cases, a base pair is formed by either an orange or yellow base, and a blue base. The two layers are separated and the structures of only the two base pairs that form that layer are shown. The hydrogen bonds within that layer are depicted with black dashed lines.

pairs in a quartet (fig. A.38). Due to the sharp bend at the apex of the coil, these base pairs were distant, in terms of their sequential position in the minicircle. The height of the column depicted how these regions were brought into close proximity. Within a single quartet layer, hydrogen bonds could form outside of each base pair (fig. A.38); however, this was not the case for all quartets. The additional stabilisation afforded by these hydrogen bonds, in addition to the increased stacking interactions, is likely the cause of the stable nature of the column. Visual inspection of the affected trajectories shows that the coil forms immediately in the simulation, whilst the hydrogen bond restraints are applied. Furthermore, it forms at the height established within that simulation; it neither grows nor shrinks once nucleated. The formation during the application of the hydrogen bond restraints suggests that the formation is an artefact of the restraints, and the requirement to maintain hydrogen bonding. As bubbles are unable to nucleate, the DNA pursues an alternative direction to dissipate the torsional strain, further evidence that this structure is not experimentally valid.

Table A.39: *The simulations that formed the coil structure. Where* (x2) *is stated, two different replicas of that simulation formed the coil and both were repeated. Where* (ii) *is stated, two repeats were needed as the first repeat also formed the coil structure. All coils were formed at only the highest superhelical density ($\Delta Lk = -6$).*

Sequence	Simulation	Conditions
Control339	Control339	0.2 M, 0.5 M, 1 M
	cAC	0.4 M, 0.5 M
	cCC	0.2 M (ii)
	cCT	0.2 M
	cGG	0.2 M (ii), 0.4 M (x2)
	cGA	0.2 M
Mod339	Mod339	0.5 M
	mAA	0.2 M
	mAG	0.2 M
	mCA	0.2 M
	mCC	0.2 M
	mGA	0.2 M
	mTG	0.2 M
	mTT	0.2 M
	m3nt	0.2 M
inserted339	i2GG	0.2 M
	i3nt	0.2 M
Strong339	sGG	0.2 M
	s6nt	0.2 M (x2)
	s7nt	0.2 M