Theory and simulation of twisted DNA duplexes

Ferdinando Randisi

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hillary 2019

Ai miei nonni,
soprattutto agli
scienziati mancati
Franca e Giovanni.

Acknowledgements

I never filled in acknowledgements in neither of my past theses, as I thought of it as a section mostly filled with self-congratulatory, vanaglourious fluff. More recently, I came to believe that the conclusion of my DPhil is a pretty good occasion to look back at my life journey, and to thank the people that were instrumental in leading me where I am.

Obtaining a PhD from a such a highly respected and world famous institution as the University of Oxford has been a dream of mine ever since I decided to study Physics. I am deeply grateful to all the people that helped me get this far, and I hope that I will be able to make it up to them and to society in general with my future contributions to the world. While I did put in a lot of work to get to this point, many people that were smarter, worked harder, and in general deserved far greater praise than me didn't quite get to a similarly satisfying point. Life is not fair, and luck has a major part in determining the outcome of our blood, sweat and tears. It is by acknowledging the huge luck that has accompanied most of my hard (as well as the not-so-hard) work that I choose to begin these acknowledgements.

Then, I should thank my thesis supervisor Ard Louis and Jon Doye. Many of my colleagues have had abysmal experiences with punishing and shaming supervisors that made their doctoral experience as crushing as it appears in popular culture, and more. But I was given interesting problems to work on, I received guidance on what tools to use to do that, and I experienced a positive and cheerful working atmosphere within the research group. I have also been very lucky to have been mentored by exceptional early-career scientists Petr Šulc, Flavio Romano, and Lorenzo Rovigatti, that through example and words taught me a lot, especially about how to conduct the technical aspects of my research. Specifically, Flavio and Lorenzo have had a strong influence on my programming style. I also learnt a lot from fellow DPhil students Chico Camargo and Megan Engel, with whom I shared the office for most of my DPhil, and from my collaborator Romain Rollin. Jonathan Patterson has been the best computing cluster administrator I could hope for, always doing all he possibly could to make sure that our computing cluster hydra was running smoothingly and thinking about ways to make our simulation environment the best possible one, at times even well beyond usual working hours.

I also want to express my gratitude to my *Laurea Magistrale* and *Laurea* thesis advisors Guido Martinelli and Andrea Pelissetto, as well as Fabio Sciarrino for helping me in taking the very first steps in the world of research, and supporting

me while deciding on my career beyond my first two degrees at Sapienza Università di Roma. I'm thankful to all my university teachers both here in Oxford and at Sapienza, as well as DTC director David Gavaghan for providing me with the background I needed to move into bioscience. I was not minted from scratch on my first day of University, so I also want to thank my school teachers, especially my high school teachers proff. Donati and Paesani at Liceo Scientifico Statale A. Avogadro in Rome, Mr. Ziesmer and Fowler at Sultana High School in Hesperia (California), and maestra Luciana at Scuola Elementare Lante della Rovere in Rome. My friend Alessandro Manacorda has also been a great teacher, showing me how a relaxed attitude to study can yield greater results than stressing out too much about grades and study metrics, maybe saving me from a nerve breakdown in the following years. Among my scientific peers, he's definitely the one that has influenced me the most.

Finally, a huge amount of gratitude goes to my family and friends for supporting me (and often putting up with me) throughout my studies, starting with my grandparents Pia, Ferdinando, and especially the missed scientists Giovanni and Franca, to whom I dedicate this thesis. My sister Elena and my parents Desirée and Michele, that provided me with love and plenty of encouragement to pursue my dreams, and put up with my absence in the past few years, I can't really thank enough. Joelle has kept me sane and focused in the incredibly stressful year before thesis submission: her unwavering support and love really drove me to punch through that. My friends, old and recent alike, are truly one of things I am most thankful for, and though I would wish to explicitly thank all of them I probably should not, as their names would fill several pages. Regardless of whether we've met in kindergarten, at middle or high school, travelling in places far away, or at a pizzeria near home last month, I cherish you and I'm grateful to have you.

Abstract

We use basic statistical mechanics and computer simulations with coarse-grained models to investigate the response of (inhomogeneous) DNA duplexes to linear and torsional mechanical stress. While the response of homogeneous DNA to more modest external stresses has been studied in detail, much less is know about DNA response to very strong over or under twisting, or about the biologically relevant case of inhomogeneous DNA under external torsion. By simulating geometries that resemble single-molecule molecular tweezers experiments previously reported in the literature, we validate our basic models, and measure the end-to-end duplex extension, torque, and the denaturation states of individual base-pairs, as well as the position of plectonemes when present in the post-buckling state. We generalise previous observations that plectonemes are preferentially localised (i.e. pinned) on sequence mismatches to predict that any sequence that is either significantly more bent or more flexible than the rest of the duplex should present a similar effect. We develop a simple theory to quantify this preference and test it in simulations, observing a semi-quantitative agreement. We also propose a general protocol to extend the popular oxDNA coarse-grained model of DNA to treat such bent/flexible sequences, and apply it to model thymine dimers, the most common DNA photoproducts. The theory provides almost quantitative agreement with simulations; in particular, the prediction that the pinning is dependent only on the bending angle and flexibility of the sequence, but not on the detail of how these are generated, is confirmed. Some consequences of the presence of a thymine dimer in biological DNA are also proposed. Finally, we use oxDNA to investigate the boundaries between the pre-buckling, post-buckling, and torsionally melted states, as well as the features of torsionally-melted underwound and overwound DNA, respectively called L-DNA and P-DNA. Unexpectedly, we observe that both torsionally-melted forms preferentially relax writhe by forming solenoids, rather than plectonemes. We compare our results with previous experimental and theoretical work and propose some experiments to confirm or deny this peculiar feature.

Contents

Gather 'round, you people, the event starts! Take your seats on best places, and witness the little Macromolecule that could. The ledger of life, it is written on it, but how do we even read it? We understand the words it spells, but what about the sentences? And is there more to them beween the lines, in the different strokes that depict its letters? Come and find out, in the way most obvious since ancient times: restrain it, bully it, pull it, twist it, until it speaks. But fear not for the little Macromolecule, for it's used to it, and maybe even enjoys it. That's what the event is all about.

1 Introduction

Contents

1.1 Chemical and structural properties of DNA

The discovery of the role of deoxyribonucleic acid (DNA) as the carrier of genetic information is one of the main scientific achievements of the 20th century. DNA was first isolated in 1868, when (Miescher 1897) isolated the substance contained in the cell nucleus and called it "nuclein". Because nuclein contained no sulphur, he concluded that nuclein was not a protein. Further work by (Kossel 1891) showed that

nuclein contained four distinct nitrogenous bases, as well as sugar. (Levene 1919) was later able to discover the structure of the repeating unit of DNA, called *nucleotide*: a sugar, phosphate, and a nitrogenous base among adenine (A), cytosine (C), guanine (G) and thymine (T). At first, DNA was dismissed as a genetic information carrier due to the belief that the chains it formed were too small, or repetitive, to contain significant information. It was (Avery et al. 1944) who showed that DNA was indeed the molecule responsible for the transfer of traits from dead bacteria to living ones. The determination of the structure of DNA was hindered by the fact that DNA could assume two different double-helical structures (named 'A' and 'B'). This fact was shown by (Franklin and Gosling 1953), and their data was used by (Watson and Crick 1953) to propose their famous DNA model. (Chargaff et al. 1952) had previously discovered that the relative content of A and C was respectively equal to the one of T and G, and (Creeth et al. 1947) had proposed that bases were linked by hydrogen-bonding, but it wasn't until the work of Watson and Crick that these facts were linked. They proposed a structure of DNA were the nucleotides were arranged in a double helix (fig. 1.1): nucleotides on a single strand are separated by alternating sugar (deoxyribose) and phosphate, all linked with covalent bonds, and opposite nucleotides on distinct strands are linked by hydrogen bonds. Each A is linked with two hydrogen bonds to a T on the opposite strand, and each C is linked with three hydrogen bonds to a G. Later, other binding patterns have also been observed, such as Hoogsteen base-pairs (Hoogsteen 1963; Neidle 2007), but the former are the vastly most common ones. The double-helical structure allows for the characteristic semi-conservative replication mechanism discussed in section 1.2. Several double-helical structures are possible in DNA (Ghosh and Bansal 2003), including the right-handed A-DNA and B-DNA, and the left-handed Z-DNA (fig. 1.1), as well as several non-double-helical ones (Ghoshdastidar and Bansal 2018). B-DNA is the most biologically-relevant configuration, observed in the majority of the nuclei of bacteria, archaea, and eukarya, as well as in most viruses, while Z-DNA occurs *in vivo* in some GC-rich sequences during strong twisting such as the one generated during transcription (Ha et al. 2005; Sutthibutpong et al. 2016). A-DNA is also present in biological solvent conditions, but its role in biology is still unclear.

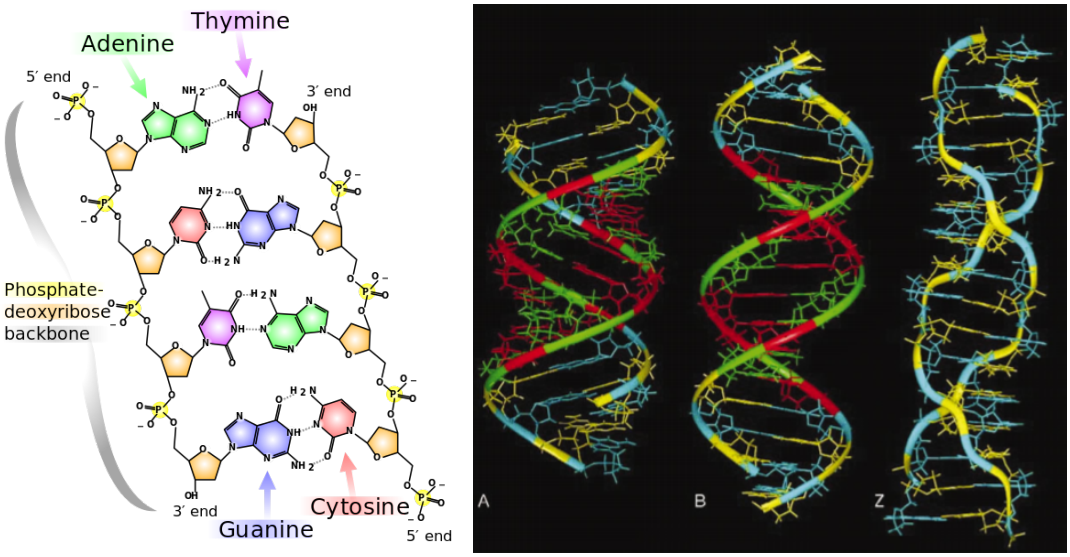

Figure 1.1: Left: schematic DNA structure showing the structure of the different nuclotides, from (Ball 2018). Right: three dimensional structures of the most famous DNA helices, the right-handed A-DNA and B-DNA, and the left-handed Z-DNA, from (Ghosh and Bansal 2003).

1.2 Inheritance and gene expression

In biology, inheritance is defined as the process by which genetic characteristics are transmitted from the parent to the offspring. Several theories of inheritance have been proposed through history (Lonie 2011), but it wasn't until the work of (Mendel 1865) that genetics was rigorously formalised. The cardinal role of DNA inheritance was first recognised by (Avery et al. 1944), who demonstrated how DNA was responsible for the induction of virulence in bacteria. In the same year, (Schrödinger 1944) postulated that genetic information should be carried by an "aperiodic crystal". In this context, (Watson and Crick 1953) proposed their double-helical DNA structure, with the order of nitrogenous bases DNA encoding the genetic information. The specificity of base-pairing also suggested a mechanism for DNA replication: as the two strands forming a double helix are complementary to the other, each strands can behave as a template to build the other. Therefore a double-helix (or duplex, as it will be called for brevity in the rest of the thesis) can be separated in its constituent strands, and each strand can be used as a template to synthesise a copy of the missing strand, generating two double helices from one (Watson and Crick 1953). *In vivo*, this task is carried out by a class

of enzymes known as DNA *polymerases*.

The information encoded in the DNA influences the chemical syntheses taking place in the cell. First, the DNA is used as a template to transcribe ribonucleic acid (RNA), a nucleic acid very similar to DNA: at the most fundamental level, the molecules differ only in the fact that they use a different sugar in the backbone (ribose for RNA, deoxyribose for DNA) and in the base used as complementary to A (T for DNA, and uracil (U) for RNA). RNA is first synthesised by enzymes called RNA polymerases, using a complementary DNA strand as a template. Then, the RNA can be used to perform specific tasks within the cells (such as ribozymes, which perform catalysis) or used themselves as a template to produce proteins.

Information storage for RNA and protein synthesis has been recognised as the key function of DNA. However, the DNA of most organisms also contains a large fraction of non-coding regions (Pennisi 2012). The role of these regions is still not fully understood, but there are several results pointing to important functions of these non-coding regions, from structural maintenance of the genetic material, to regulation of gene expression (Ludwig 2002; Ding et al. 2014). Likewise, chemical modifications of DNA are used to influence several functions inside the cell: for example, methylation of DNA bases has been recognised to influence several different processes, from sequence readout (Jeltsch 2002) to DNA damage repair (Modrich 2016). Therefore, there is more to DNA than the mere order in which the bases are distributed along a given strand: a structural and chemical code, along a more obvious digital one, encodes the information that is necessary for the biochemical basis of life (Noble 2013).

The elucidation of how the response of DNA structure to external factors like force and torque and internal factors like structural inhomogeneities is one of the general scopes of this thesis. Understanding this mechanical response may shed new light on how DNA encodes information beyond a simple digital code.

1.3 Topological properties of DNA

As discussed in section 1.2, strand separation is a process necessary in DNA replication and transcription. It was clear since early on in the study of DNA

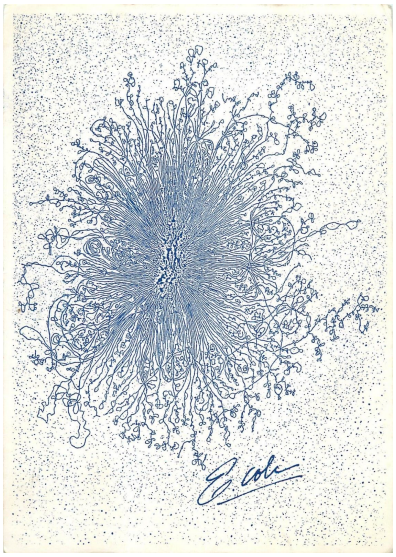

Figure 1.2: Electron microscopy image of a lysed, stained *E. Coli* bacterium. Notice the intricate supercoiled structures formed by the bacterial chromosome. From (Kavenoff 1972).

structural biology that the highly intertwined nature of the DNA duplex would hinder strand separation. Consider, for example, the classic *DNA replication problem* (Holmes 1998): it was proposed that the breaking of bonds (commonly referred to as *melting*, or *denaturation*) would induce torsional stress in regions of the strand adjacent to the breakage. When this melting was extended over several contiguous base-pairs, such as by action of polymerases, this could cause the duplex to writhe against itself, mechanically hindering the progression of the polymerase. Considerations of this kind prompted scientists to consider the topological properties of DNA. These were corroborated from experiments (Weil and Vinograd 1963), suggesting a "twisted circular" microscopic configuration in the DNA of the Polyoma virus: genomes with the same sequence but with different topological properties would segregate in a centrifuge, because of the influence of these topological properties on the DNA structure.

Later on, it was shown that most bacteria kept their circular genomes in a strongly unwound state, causing them to assume characteristic coiled structures, as seen in fig. 1.2. Because the two strands forming the double-helix are already coiled around each other, these higher-order structures are often called *supercoiled*, and the twisting of DNA is referred to as *supercoiling*. In both eukaryotes and bacteria, the topological properties of the genome are preserved by binding to the

nuclear matrix and DNA binding proteins: eukaryotic genomes wrap around proteins called *histones*, and far-away segments of DNA are bound together in bacteria by proteins forming a structural maintenance complex (SMC) (Sinden 1994; Hacker et al. 2017). Later on (Bates and Maxwell 2005), the topological properties of DNA were shown to regulate its expression and control many other processes that involve strand separation, such as transcription, replication, and recombination. As one might expect, the degree of twisting in a portion of genome influences the energy cost required to break a base-pair. However, the twisting also determines the way the genome forms looped domains.

1.4 Single-molecule *in vitro* experiments

Experiments on the behaviour of supercoiled DNA *in vivo* can yield useful insight into how supercoiling impacts cellular processes. However, the large number of different actors in the crowded cellular environment makes it difficult to investigate the behaviour of DNA, and to separate the phenomena that are due to DNA alone from the ones that involve its interaction with the other constituents of the cytosol. *In vivo* experiments can therefore be complemented by so called *in vitro* experiments, where the behaviour of biological molecules is investigated in a controlled environment, outside of a living systems. Several experiments of this class have been crucial in elucidating cardinal aspects of DNA, from the X-ray crystallography (Franklin and Gosling 1953) that determined the helical structure of DNA, to gel electrophoresis, optical absorption studies and more. All of these techniques perform some kind of averaging on a whole population (or ensemble) of molecules, which yields only indirect information about the individual molecule of the ensemble, and often mostly neglects the variability observed within a population. This severely limits the applicability of such techniques to the study of the kinetics of a specific molecule. Moreover, ensemble techniques do not distinguish the effects of static and dynamical disorder, a factor often crucial to investigate biomolecule conformational dynamic (Kapanidis and Strick 2009). Investigating the details of an individual molecule and circumventing the drawbacks above are some of the strong points of single molecule experiments. Single molecule methods can extract information on structure

and dynamics of an individual molecule and investigate its response to external perturbations, thus probing structural, mechanical, and dynamical properties of individual biopolymers. In this section we focus on molecular tweezers experiments, a single-molecule technique extensively modelled through the rest of the thesis.

1.4.1 Molecular tweezers (MT) experiments

MT are one of the key single-molecule techniques employed to study the response of DNA to external torques and forces. The most commonly used designs are magnetic and optical tweezers, although several variants of both techniques have been used (Bustamante et al. 2003; De Vlaminck and Dekker 2012; Bryant, Oberstrass, et al. 2012; Forties and Wang 2014).

1.4.1.1 Optical tweezers

Optical tweezers use lasers to apply force to a micrometer-size transparent bead. The radiation pressure exerted by the laser pushes the bead to the focal point of the laser, which acts as a sort of optical trap. The position of the trap can be controlled with great precision, and the position of the bead can be monitored through microscopy. The beads are coated with a chemical with a high binding affinity to the functionalised ends of the DNA strands; for example, a bead can be coated with streptavidin, which in turn binds to DNA conjugated with biotin. Several other techniques can be used, reviewed in (Janissen et al. 2014). In this way DNA is attached to the bead, which in turn can be moved around to apply torque and force, as shown schematically in fig. 1.3. The distance between the centre of the bead and the trap position can then be mapped on the force exerted on the bead. Keeping the focal point fixed in space allows for experimentation with a constant duplex extension (also called end-to-end distance), while the focal point can be moved according to a feedback mechanism to perform experiments at fixed force (Bustamante et al. 2003). This allows for the application of a linear force, investigating the stretching behaviour of DNA (Wang, Yin, et al. 1997; Smith et al. 1996), measuring the force required to disrupt DNA structures such as hairpins (Engel, Ritchie, et al. 2014) and artificial constructs (Engel, Smith, et al. 2018). The setup can be also extended in order to apply torque, such as by using three

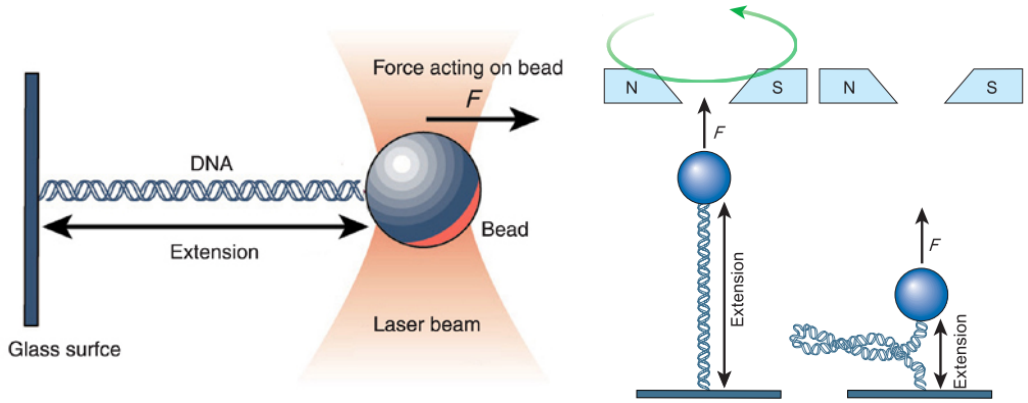

Figure 1.3: Left: in an optical tweezers experiment, a focused laser beam is used to pull and twist a magnetic bead in order to apply force and torque to the DNA duplex. Reproduced from (Geller et al. 2011). Right: in a magnetic tweezers experiment, magnetic fields replace the laser as a way to control the bead. Notice the presence of the plectoneme in the twisted state. From (Bustamante et al. 2003).

beads instead of two (Bryant, Stone, et al. 2003), by using a bead shaped like a truncated cone (Deufel et al. 2007), or by detection of the variation in angular momentum of the light shining on the bead.

1.4.1.2 Magnetic tweezers

Magnetic tweezers are similar to optical tweezers, but they use a magnetic bead in a magnetic field instead of a dielectric one hit by a laser beam (see fig. 1.3). In this case, the force acting on the bead can be measured by the fluctuations in its transverse position (Strick, Dessinges, et al. 2003). Magnetic tweezers were first used to investigate the elastic properties of supercoiled DNA extracted from the λ-bacteriophage (Strick, Allemand, Bensimon, Bensimon, et al. 1996; Strick, Allemand, Bensimon, and Croquette 1998; Strick, Croquette, et al. 2000), but in the past decades they have been used by several groups to probe the mechanical response of DNA in a varying range of solvent conditions, including salt concentration, pH, temperature, and for duplexes of varying sequence and length (Mosconi et al. 2009; Brutzer et al. 2010; Salerno et al. 2012; Tempestini et al. 2013; Meng et al. 2014). Like for optical tweezers, variations of this experimental technique can directly measure the torque exerted by the molecule, thereby permitting comparison with theory and computation (Bryant, Stone, et al. 2003; Lipfert, Kerssemakers, et al. 2010; Lipfert, Wiggin, et al. 2011; Janssen et al. 2012; Oberstrass et al. 2012).

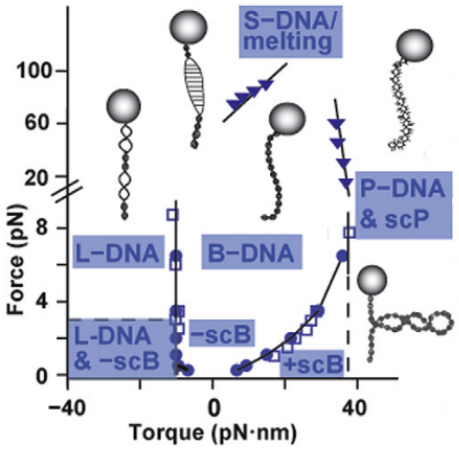

Figure 1.4: Phase diagram of a DNA duplex under fixed force and torque. A duplex in a MT experiment can assume the standard B-DNA phase or a denatured phase (L-DNA and P-DNA, respectively at negative and positive supercoiling), potentially supercoiled (+/-scB and scP, respectively for B-DNA and P-DNA). At high force and low torque, the elusive S-DNA phase appears. From (Kriegel, Ermann, and Lipfert 2017).

The technique has also been applied to study transcription (Howan et al. 2012), and the action of helicases (Manosas et al. 2012) and topoisomerases (Koster et al. 2010; Nöllmann et al. 2007).

1.4.1.3 Some shortcomings of MT

MT have been proven a versatile tool, capable of measuring the torque and force response of DNA, either isolated or bound by other bio-molecules, as discussed in the previous subsections. However, they do have their limits: when used alone, MT can only measure global properties of DNA, such as the duplex extension and torque. Local structural changes that do not appreciably impact these properties remain hard to study (Kim, Ganji, et al. 2017). Besides, the optical or magnetic beads attached to the DNA have significant inertia, which can significantly influence the dynamical behaviour of DNA (Daniels and Sethna 2011). The interpretation of the results by MT can also be guided theoretical methods, such as the ones discussed in section 1.5.

1.4.1.4 DNA phase diagram in MT experiments

MT experiments on torsionally unconstrained molecules usually control the pulling force F applied on the DNA duplex, and measure the duplex response in terms of the variation of its extension. The torsional response of a duplex can also be

investigated, most commonly by either measuring the torque applied to the DNA, or by controlling the number of turns stored in the molecule. A DNA double helix in the straight, torsionally unstressed state has an equilibrium linking number proportional to its contour length L. If the helical repeat of the duplex is d, then the equilibrium linking number is

$$\mathrm{Lk}_0 = \frac{L}{d}\,, \tag{1.1}$$

where d has been measured to be about 10.5 bp/turn in experiments, depending on solvent conditions (Kriegel, Matek, et al. 2018), so that for example a 1.05 kilo-base-pair (kbp) DNA duplex has Lk_0=100 according to eqn. 1.1. The number of turns added to a DNA duplex is therefore the excess linking number $\Delta\mathrm{Lk} = \mathrm{Lk} - \mathrm{Lk}_0$. Because the excess turns are spread throughout the molecule length, ΔLk cannot be reliably used to compare the properties of duplexes of different length. Instead, a better measurement is the superhelical density σ,

$$\sigma = \frac{\mathrm{Lk} - \mathrm{Lk}_0}{\mathrm{Lk}_0} = \frac{\Delta\mathrm{Lk}}{\mathrm{Lk}_0} = \frac{\mathrm{Lk}}{\mathrm{Lk}_0} - 1\,. \tag{1.2}$$

also called supercoiling density or excess linking number density. The precise control of torque and force, as well as measurements of the response of DNA duplexes in terms of its extension, has been used to elucidate the conditions at which DNA transitions from the unsupercoiled B-DNA phase to other forms (Marko and Neukirch 2013; Kriegel, Ermann, and Lipfert 2017). A simplified picture of the phase-diagram of double-helical DNA is shown in fig. 1.4.

Let us first consider the regime of comparatively low pulling forces F: when the magnitude of the torque acting on the DNA is increased, the DNA does not appreciably change structure at first, its extension decreasing only slightly. Then, as the torque reaches a particular, force-dependent threshold, the DNA suddenly buckles, forming one or more (Emanuel et al. 2013) supercoiled structures called *plectonemes* (see fig. 1.3). This phase is observed in negatively supercoiled DNA for forces of up to about 1 pN and torques above -10 pN nm (Sheinin, Forth, et al. 2011), but in positively supercoiled DNA it can be present up until about $F = 7$ pN and torques of about 40 pN nm, depending on a multitude of factors such as the length of the duplex, its sequence, and solvent conditions like temperature and salt concentration.

At about $F = 65$ pN, for torques typical of the unsupercoiled B-DNA phase, one of the first exotic structures to be characterised can be observed: this is S-DNA, also called *overstretched* DNA. This structure is characterised by a significantly elongated structure with a contour length about 1.7 times that of B-DNA. The details of its structure are still unclear, including whether S-DNA is base-paired or not. Recent studies have observed multiple states of S-DNA, constituted of base-paired and unpaired strands (Zhu et al. 2016).

Duplex underwinding experiments have shown the presence of a phase with roughly the same helical repeat as B-DNA, but left-handed rather than right-handed. This phase, termed L-DNA (Bryant, Stone, et al. 2003) due to its left-handed twist, is formed by applying a torque of about -10 to -20 pN nm or greater. The net helicity of this structure is similar to that of Z-DNA. However, as shown with MT experiments on L-DNA in a bath with a fluorescently labelled, Z-DNA-binding protein (Vlijm et al. 2015), L-DNA is not significantly composed of Z-DNA. Instead, it is assumed of being formed mostly of denatured DNA. Some simulations of L-DNA are presented in chapter 5.

A similar structure to L-DNA is observed when DNA is subjected to a positive torque: this structure is termed P-DNA, since it is expected to have the backbone in the centre of its helical structure with exposed, unpaired bases, similarly to an early proposal by (Pauling and Corey 1952) (hence the "P" for "Pauling") of a triple-helical DNA structure.

Finally, coexistence between these phases is not limited to the phase boundaries: for example, bubbles and plectonemes can coexist when even plectonemes spanning the whole duplex are not sufficient to absorb all the ΔLk.

1.4.2 Fluorescence microscopy

While MT are good at measuring global properties of DNA, they do not, on their own, provide access to local details. In this section we describe fluorescence microscopy, which does allow such local access. With fluorescence microscopy, selected parts of a molecular system are made to shine, thereby effectively highlighting them. Fluorescence microscopy techniques can be used to image the evolution in time of

the conformation of a single DNA duplex, and they can be combined with MT to study how the local details of a duplex are influenced by force and torque. Several intercalating dyes (composed by aromatic rings that preferentially stack between adjacent bases along the same strand) have been investigated (Biebricher et al. 2015), as well as different DNA phases, including overstretched DNA (Gross et al. 2011; King et al. 2013) and the dynamic behaviour of supercoiled DNA (Loenhout et al. 2012). Other studies include protein-DNA interactions, such as the dynamics of recombinant protein A (RecA) on DNA (Forget and Kowalczykowski 2012) as well as the bacterial replication molecular machinery (Oijen and Loparo 2010; Hamdan et al. 2009). Some of these intercalating dyes induce positive supercoiling, an effect which can be used to change the structural properties of large DNA constructs (Zadegan et al. 2017). Recent work (Ganji et al. 2016; Kim, Ganji, et al. 2017) has also used intercalators to induce supercoiling in DNA duplexes with bond ends fixed on a coverslip, effectively mimicking the effects of MT without using beads, though without explicit control of torque and force.

1.5 Modelling the behaviour of DNA duplexes under torsion and twist

Several theoretical and computational models have been developed to understand the behaviour of DNA when subject to external torques and forces, helping to interpret and sometimes complementing the ever growing number of experimental results. Processes involving DNA span a remarkable range of time and length scales, from the femtoseconds of the hydrogen-bond vibration times to the years of life of a single cell, and from the subnanometer-scale to the two meters of DNA wrapped-up inside each of our cells. Treating all of these scales with the same level of detail is currently impossible. Instead, care must be used when choosing a model that captures the aspects of DNA relevant to the question at hand. This section provides a broad review of some of them, while a few specific models will be discussed in more detail in the rest of the chapter.

1.5.1 Atomistic models

Among the theoretical models discussed here, the ones with the higher spatial resolution are atomistic models such as the force fields AMBER (Cornell et al. 1995) and CHARMM (Brooks et al. 2009). These simulations encode the position of individual atoms, and simulate their interaction with computational techniques such molecular dynamics and the Monte Carlo method. They can be used to investigate the chemical details of the different nucleotides, their interactions, and the impact that sequence can have on the overall duplex properties like in the large study performed by the Ascona B-DNA Consortium (Lavery, Zakrzewska, Beveridge, et al. 2010). Other work includes investigations in DNA elasticity and chirality, as well as the detailed interactions between the duplex and the aqueous solvent (Noy and Golestanian 2010; Noy and Golestanian 2012). The main problem with atomistic simulations is their large computational cost, though the exponential increase in available computing power and recent advances in computational techniques (Kannan and Zacharias 2009; Pérez et al. 2012) keep extending the reach of such methods: these systems are typically only simulated for up to a few microseconds (Maffeo et al. 2014), such that only fast modes can be properly sampled.

In particular, atomistic models have been used to study the local deformation induced by external torques (Randall et al. 2009; Wereszczynski and Andricioaei 2006) and linear stretching forces (Severin et al. 2011; Harris et al. 2005; Piana 2005) in B-DNA of 12 base-pairs (bp). These are also used to simulate large DNA structures (Yoo et al. 2017), DNA-protein complexes (Knips and Zacharias 2017), and damaged DNA (Knips and Zacharias 2017). Despite recent advances, physical processes involving large deformations of long DNA systems are hard to study with atomistic models. Continuum models and coarse grained models, described in section 1.5.2 and section 1.5.3, are better suited for this task.

1.5.2 Continuum models

At the opposite end of the resolution spectrum from atomistic models, continuum models neglect the microscopic structure of nucleic acids. Instead, they often aim to capture the behaviour of DNA in the limit of length scales much larger than the

bending persistence length. Because of their low level of detail, continuum models are often computationally very efficient. A very well known continuum model is the worm-like chain (WLC) of (Kratky and Porod 1949), depicting DNA as a flexible, isotropic, homogeneous rod. Later on, the model was expanded in order to keep up with the results provided by experiments, including the effects of the stretching (Marko and Siggia 1995) and torsional (Moroz and Nelson 1997) elasticity of DNA. When excluded volume effects can be neglected, or treated in an *ad hoc* fashion, the models remain analytically tractable. Several approximations have been developed for extension and torsional behaviours (Marko and Siggia 1995; Bouchiat and Mézard 1998; Odijk 1995), allowing for direct comparison with experiments. Marko, Neukirch and collaborators have successfully applied this class of models to describe several different aspects of DNA in MT (Marko 2007; Marko and Neukirch 2012; Marko and Neukirch 2013; Neukirch and Marko 2011). Some continuum models also take into account the features of the solvent, such as the influence of salt concentration to the electrostatics of DNA (Argudo and Purohit 2012; Neukirch and Marko 2011).

Continuum models have provided valuable insight into the behaviour of long DNA duplexes. They can also be extended to take into account the chiral nature of the DNA helix, but each added degree of realism makes the model less tractable. These effect include twist-stretch coupling (Marko 1997), in which DNA becomes longer when overwound, as well as the interconversion between B-DNA and other helical phases (Marko and Neukirch 2013), sequence-dependent stiffness (Goyal and Perkins 2008), and twist-bend coupling (Skoruppa et al. 2017). They remain hampered by the fact that it is hard to provide a realistic description of strand helical geometry, which is important when the details of contact interactions with cannot be overlooked.

1.5.3 Coarse-grained models

Coarse-grained models attempt to compromise between the chemical detail and the versatility of atomistic models, and the practicality of continuum models. The many degrees of freedom in atomistic model are summarised by a reduced number of coordinates, which is still able to capture some of the microscopic features of a system. In practice, the process by which this summary should occur depends

strongly on the features characterising the system to be modelled; even after the focus of the model has been set, defining the model is more an art than a science.

Several models have been proposed with the specific goal of predicting the probability of base-pair denaturation (Metzler et al. 2009). Some of these models carefully choose one-dimensional coordinates to model the state of the nucleotides in the strand, and apply statistical physics methods to compute their equilibrium properties. The model by (Poland and Scheraga 1966) defines the melting transition of polynucleotides at the cross-over between the entropy gain due to denaturation and the energetic cost of breaking a base-pair. The model has been studied extensively, especially in the context of the critical exponents of the melting transitions (Metzler et al. 2009; Kafri et al. 2000; Carlon et al. 2002; Kabakçioğlu et al. 2009) and the probability of occurrence of each run of consecutive denatured base-pairs, commonly termed *bubbles* (Ambjörnsson et al. 2007; Bar et al. 2007). The model developed by SantaLucia and coworkers (SantaLucia and Hicks 2004) is probably the one most frequently used to predict the melting temperature of an oligonucleotide, and is discretised at the level of the base-pair-step, as opposed to the base-pair one like the model by Poland and Sheraga. The SIDD model of Benham and his collaborators (Benham 1992; Bauer and Benham 1993; Fye and Benham 1999) aims to predict the denaturation probability of a linear or circular DNA duplex in the presence of supercoiling, and will be discussed more detail in section 2.5. Another model worth mentioning is the Peyrard-Bishop-Daxois model, which represents base-pair denaturation using a continuous coordinate (Peyrard and Bishop 1989; Dauxois 1991). Other models choose to keep a three-dimensional representation of the nucleotides, explicitly representing their coordinates in the cartesian space. This significantly increases the complexity of the modelling approach, but can sometimes be more versatile than one dimensional models. Some of these models are the 3SPN model of Knotts et al. 2007, the recent TIS-DNA model of (Chakraborty et al. 2018), and the oxDNA model, used in much of the thesis and described in the following section.

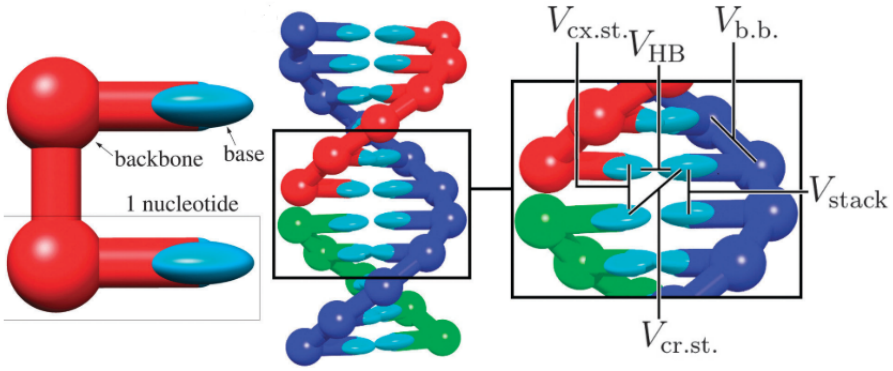

Figure 1.5: Left: close up of two nucleotides in the oxDNA model. Right: a nicked duplex in oxDNA, with different strands coloured in red, blue and green, with a close-up showing the interaction between the backbone interaction site and the base interaction sites (one for the stacking potential and one for the hydrogen bonding potential). From (Doye et al. 2013).

1.6 OxDNA

Most of the simulations in this thesis have been performed with the oxDNA model (Ouldridge, Louis, and Doye 2011; Snodin et al. 2015), introduced in this section. In brief, the oxDNA model is a coarse-grained nucleotide-level model of DNA. Each nucleotide is represented as a rigid body with three interaction sites: the backbone interaction site, the stacking interaction side, and the hydrogen bonding interaction site. A schematic depiction of the model is found in fig. 1.5.

The parameters of oxDNA are fixed using a top-down approach: first, a set of phenomenologically motivated interactions are included in the interaction potential with given functional forms and several free parameters. Then, the functional parameters are set in order that to satisfyingly reproduce the experimental values of selected structural, thermodynamic, and elastic properties of DNA. For computational simplicity, all the interaction terms are pairwise.

In this thesis, we used the most recent version of the oxDNA model, oxDNA2 (Snodin et al. 2015; Snodin 2015), with the interaction potential

$$V_{\text{oxDNA2}} = \sum_{\text{nearest neighbours}} (V_{\text{backbone}} + V_{\text{stack}} + V_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross stack}} + V_{\text{exc}} + V_{\text{coax stack}} + V_{\text{DH}}) \,. \quad (1.3)$$

Here the nearest-neighbour terms V_{backbone}, V_{stack}, V_{exc} capture backbone connectivity, base-pair stacking, and nucleotide excluded volume interactions respectively, and

affect only the pairs formed by nucleotides that are adjacent on the same strand. The other terms, affecting all the nucleotide pairs regardless of strand topology, represent hydrogen bonding (V_{HB}), cross stacking ($V_{\mathrm{cross\ stack}}$, i.e. stacking between non-adjacent bases), excluded volume (V_{exc}), coaxial stacking ($V_{\mathrm{coax\ stack}}$, i.e. stacking between adjacent bases on different strands, as across a nick) and electrostatic interactions in the Debye-Huckel theory (V_{DH}). We refer to chapter 2 of (Ouldridge 2011) and appendix A of (Snodin 2015) for the detailed forms of these interaction potentials.

The model can capture effects due to sequence-dependence in the hydrogen-bonding and stacking interaction, solvent salt concentration, and system temperature. DNA features that the model does not account for include sequence-dependent elasticity, hydrodynamic effects, and the details of the chemical interactions between the DNA and the molecules of the aqueous solvent, most notably water and ions. Despite these shortcomings, the model's domain of applicability is vast: oxDNA has been used to simulate a plethora of DNA structures (Doye et al. 2013; Ouldridge 2014), including nanotechnological devices and processes as well as *in vitro* experiments of DNA structural properties, minicircles and more. All the oxDNA simulations presented in this thesis have been carried out with the publicly available oxDNA code (Ouldridge, Louis, Doye, et al. 2014), using mostly the GPU molecular dynamics (MD) backend (Rovigatti et al. 2015) for MT experiments and the CPU backend for the duplex melting simulations of section 4.3. An implementation of the oxDNA model in the popular molecular simulation program LAMMPS is also available (Henrich et al. 2018).

1.7 OxDNA simulations of MT experiments and plectoneme distribution

The oxDNA model has been shown to provide semi-quantitative agreement with MT experiments Matek 2014; Matek et al. 2015: in particular, the model correctly reproduces the duplex extension and torque at varying σ and F and the supercoiled DNA phase diagram, as shown in fig. 1.6. Most notably, oxDNA simulations showed a previously undetected phase at the boundary of each of the L/P-DNA phases and the supercoiled B-DNA phase: in this phase (called the tip-bubble phase), a

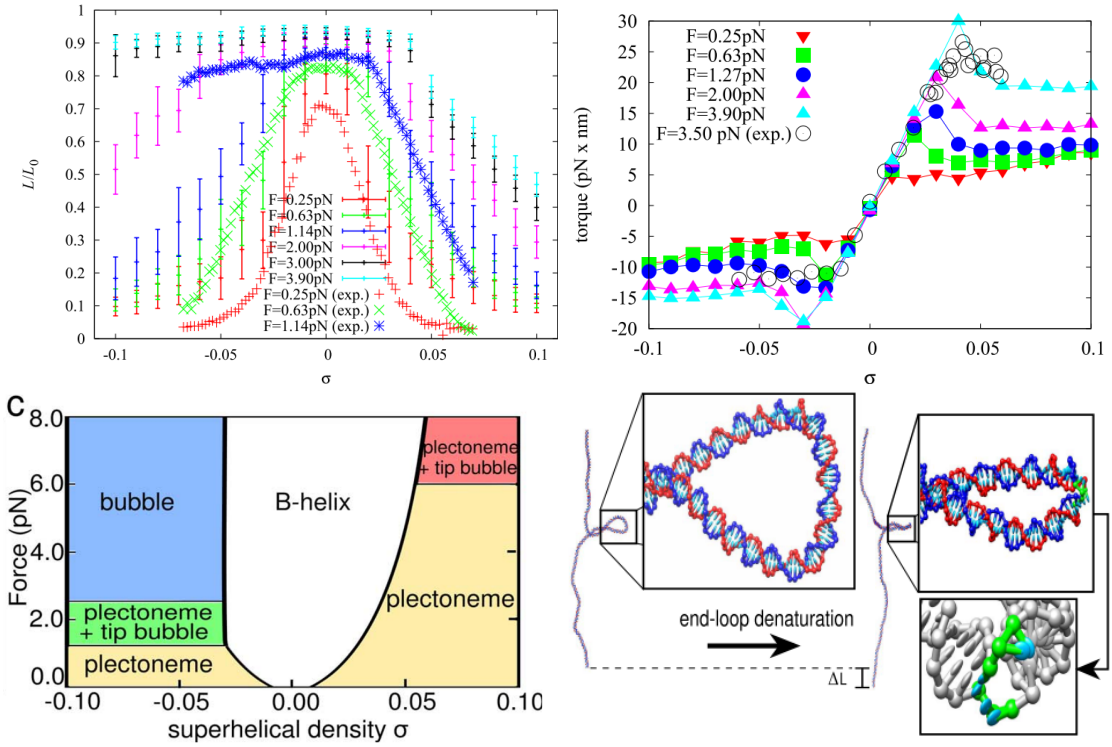

Figure 1.6: Top: comparison of duplex relative extension L/L_0 and torque in oxDNA simulations and experiments from (Salerno et al. 2012) and (Janssen et al. 2012) respectively. Bottom left: phase diagram of a 600 bp oxDNA duplex in simulated MT experiments. Notice the presence of the tip-bubble phase for both positive and negative superhelical density. Bottom right: comparison of plectoneme loops with and without a tip-bubble observed during oxDNA simulations. All figures reproduced from (Matek et al. 2015).

bubble co-localises on the tip of the plectoneme loop, lowering the energy cost of loop formation and significantly slowing down plectoneme diffusion.

The same work also observed that the probability of having a plectoneme loop on a particular duplex segment can significantly depend on the segment sequence. This effect has recently been observed in experiments (Kim, Ganji, et al. 2017), where the DNA was intercalated with sytox orange to induce positive supercoiling and fluorescence. This allowed plectonemes to be detected as peaks in DNA luminosity, as shown in fig. 1.7.

1.8 Thesis outline

The focus of this thesis is the behaviour of DNA when subjected to external torque and pulling force. The response of DNA to such external conditions is important not only for its virtually ubiquitous presence in biology, but for many

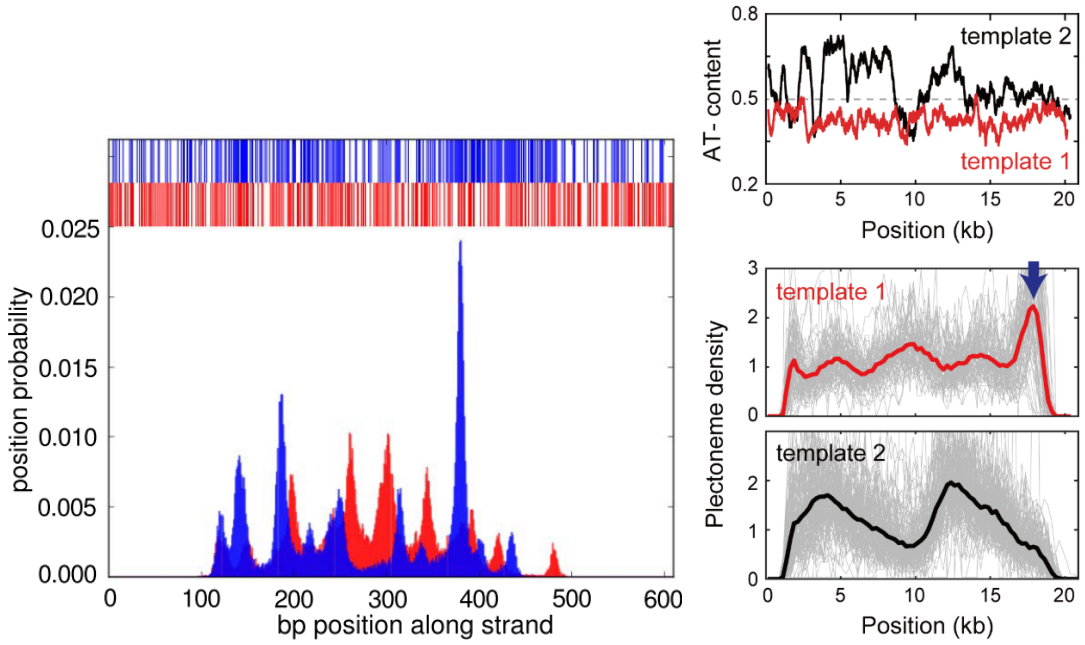

Figure 1.7: Both oxDNA simulations and experiments have observed that sequence can have a significant impact on plectoneme location. Left: the peaks in the plectoneme position distribution in 600 bp oxDNA simulations of a random sequence two sequences (red and blue respectively) correlate with AT base-pair density, shown by colouring AT base-pairs above the histograms. GC base-pairs are left white. From (Matek et al. 2015). Right: the effect is also observed in experiments, as can be seen by comparing the AT-contenst binned to 300 bp in two different sequences (red and black respectively) with the plectoneme densities of 70 MT experiments (grey thin lines) and their ensemble average (coloured). From (Kim, Ganji, et al. 2017).

other reasons as well; it can be investigated to test basic theories of statistical physics, from nucleation theory (Daniels and Sethna 2011) to non-equilibrium statistical mechanics (Mossa et al. 2009). As discussed in the previous sections, the properties of a DNA duplex in MT experiments have been thoroughly investigated in experiments and simulations, but our understanding of the impact that DNA inhomogeneities can have on supercoiled structure is still very limited. Yet, DNA is far from homogeneous: base-pair steps of different chemical identity have different structural, thermodynamic, and mechanical properties. These inhomogeneities are particularly pronounced in several biologically-important DNA loci, including methylated DNA, DNA damaged by UV light and metabolic processes, promoter sequences (which tend to have a particularly low content of GC base-pairs), and more. Understanding the interplay between DNA inhomogeneities and supercoiled structure could therefore unlock a deeper understanding of how DNA structure is used to perform biological

functions, such as gene regulation, DNA damage repair, and more.

In this thesis, we develop a theoretical model to quantify loop energetics of anomalous sequences, i.e. DNA segments with flexibility, bending or both significantly higher than the rest of the duplex. The model is then applied to investigate the impact that such an anomalous sequence has on a simple supercoiled system, such as a DNA duplex in a magnetic tweezers experiment, and tested in simulations of two distinct coarse-grained models, oxDNA and TEP, both extended as needed.

We also use the oxDNA model to investigate the properties of the denatured L-DNA and P-DNA phases, as well as their boundary with B-DNA phases. While oxDNA has been able to reproduce some important features of molecular tweezers experiments, such as the hat-curves observed in the plots of duplex extension as a function of superhelical density σ and the torque-response curves, using oxDNA to study the denatured phases means pushing oxDNA beyond its original parametrisation domain. Such a study is therefore a good test of the limits of the model.

Most importantly, if oxDNA can reproduce experimental measures of L/P-DNA well enough, then we can use the simulations to provide a nucleotide-level description of the supercoiled structures assumed by DNA in these regimes, something that to the best of our knowledge no other theoretical or experimental method has been able to provide so far.

The detailed thesis outline is the following:

- Chapter 1 (this one) introduces the basic structural and biological features of DNA, surveys some of the experimental techniques referenced in the rest of the thesis, reviews the oxDNA model and its applications to MT experiments, and contains the thesis outline.

- Chapter 2 reviews the theoretical methods commonly used to investigate the properties of DNA, including some of the theoretical models used in the thesis and the computational coarse-grained models at the base of this body of work, such as the Twistable Elastic Polymer (TEP) model.

- Chapter 3 develops a theoretical model of loop energetics to take into account structural inhomogeneities in the DNA duplex, and builds upon it to compute

the impact of such inhomogeneities to the buckling transition and to the pinning of plectonemes. Theory is then compared with TEP simulations.

- Chapter 4 discusses the application of the theory presented in the previous chapter to thymine dimers, the type of DNA damage most commonly induced in cells by exposure to UV light. The theory is tested in simulations of an extension of the oxDNA model that allows for the simulation of thymine dimers. A protocol to easily generalise this modelling approach to other inhomogeneities is also presented.

- Chapter 5 presents the results of simulated molecular tweezers experiments in oxDNA where the unwinding and pulling of the DNA are so strong to break base-pairs. Simulations show that the resulting DNA phase, known as L-DNA in the literature, preferentially relaxes writhe not in the form of plectonemes, as regular DNA does, but as solenoids.

- Chapter 6 investigates overwound DNA, measuring extension and torque over a wide range of forces at positive supercoiling. OxDNA is shown to provide good agreement with experiments and previous models in terms of the boundaries between the pre-buckling, post-buckling, and melted phases. The P-DNA phase, a right-handed analogue of the L-DNA phase, is also investigated and found to prefer solenoids to plectonemes.

- Finally, chapter 7 reports the conclusions, summarising the thesis and discussing possible directions for future work.

In which all the useful tools, artifacts, and secret techniques relevant to the protagonists of this story are presented. After all, what would Artemis be without her bow, or private Joker without his rifle, or Mario Magnotta without his washing machine?

2 Models and methods

Contents

This chapter describes some of the computational and theoretical methods used in the thesis. In particular, it reviews the TEP model, the details of the simulation methodologies, and some of the theoretical foundations upon which the main work is built.

2.1 The twistable elastic polymer (TEP) model

Simulations presented in chapter 3 model a DNA duplex with the twistable elastic polymer (TEP) model, developed by (Brackley, Morozov, et al. 2014), implemented in the oxDNA simulation program (Randisi and Rovigatti, unpublished). The model is a simple discretisation of the energy of a twistable, elastic rod with a given bending stiffness k_b, twisting stiffness k_t, and radius a, neglecting any other structural detail of the duplex. By setting a k_t=0 and in the limit of a=0, one recovers the worm-like chain (WLC) model which will be presented in section 3.1.

The elementary unit of the model is a spherical rigid body depicting a polymer segment, each with its reference frame $(\hat{\mathbf{u}}_i, \hat{\mathbf{v}}_i, \hat{\mathbf{f}}_i)$, and the interaction potential models bending, twisting, and excluded volume interactions

$$V_{\mathrm{TEP}} = V_{\mathrm{bend}} + V_{\mathrm{twist}} + V_{\mathrm{exc}} + V_{\mathrm{FENE}} + V_{\mathrm{align}} \tag{2.1}$$

where

$$V_{\mathrm{bend}} = \frac{k_b}{a} \sum_i (1 - \hat{\mathbf{u}}_{i+1} \cdot \hat{\mathbf{u}}_i) \tag{2.2}$$

$$V_{\mathrm{twist}} = \frac{k_t}{2a} \sum_i (1 + \hat{\mathbf{u}}_{i+1} \cdot \hat{\mathbf{u}}_i - \hat{\mathbf{v}}_{i+1} \cdot \hat{\mathbf{v}}_i - \hat{\mathbf{f}}_{i+1} \cdot \hat{\mathbf{f}}_i)\,. \tag{2.3}$$

The excluded volume is treated with a Weeks-Chandler-Andersen potential

$$V_{\mathrm{exc}} = \sum_{i \neq j} 4k_B T \left[\left(\frac{a}{r_{ij}} \right)^{12} - \left(\frac{a}{r_{ij}} \right)^{6} + \frac{1}{4} \right] \tilde{\theta}(r_{ij} - 2^{1/6} a)\,, \tag{2.4}$$

with $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and the step function $\tilde{\theta}$, the chain connectivity is preserved with finitely extensible nonlinear elastic (FENE) springs

$$V_{\mathrm{FENE}} = -\frac{K_{\mathrm{FENE}} R_0^2}{2} \sum_i \ln \left[1 - \left(\frac{|\mathbf{r}_i - \mathbf{r}_{i+1}}{R_0} | \right)^2 \right]\,, \tag{2.5}$$

and the last interaction term makes sure that the $\hat{\mathbf{u}}_i$ is preferentially aligned with the tangent vector $\mathbf{t}_i = (\mathbf{r}_{i+1} - \mathbf{r}_i)/r_{ij}$

$$V_{\mathrm{align}} = k_a \sum_i (1 - \hat{\mathbf{u}}_i \cdot \mathbf{t}_i) \tag{2.6}$$

In order to model DNA, the model parameters are set to $a = 2.5$ nm $= 7.4$ bp, $k_b = 45$ nm $k_B T$, $k_t = 100$ nm $k_B T$ and $k_a = 90 k_B T$.

The model has been previously applied to study several different DNA systems, including *in vivo* chromatin reconstruction experiments (Brackley, Allan, et al. 2014) and understanding facilitated diffusion (Brackley, Cates, et al. 2012; Brackley, Cates, et al. 2013). With a single rigid body per fifteen nucleotides and only three free parameters, TEP is a significantly simpler, coarser, and less versatile model than oxDNA. However, because of this simplicity, it can easily simulate much longer duplexes than oxDNA. Furthermore, its k_b and k_t directly control its persistence

lengths, making it reasonably straightforward to simulate a inhomogeneous polymers (as in chapter 3) or changing the persistence lengths to model systems that significantly differ from average B-DNA.

In chapter 3, we will be using a modified version of the TEP model to model DNA duplexes containing anomalous sequences, meaning sequences that either present a substantial bending angle $\theta_0 \neq 0$ or a bending modulus $B_1 < B$. Floppy sequences can be modelled by changing the curvature of the bending potential from k_b to $k_b B_1/B$. Bent sequences can instead be obtained by rotating the reference frame on a particle around a vector perpendicular to the longitudinal vector $\mathbf{u}$: therefore, the modified bending potential is

$$V'_{\text{bend}} = \frac{k_b}{a} \sum_i \frac{B_{1i}}{B} \left(1 - \mathbf{R}(\mathbf{v}_{i+1}, \theta_{0,i}) \mathbf{u}_{i+1} \cdot \mathbf{u}_i\right) \tag{2.7}$$

where B_{1i}/B is the bending stiffness of the i-th TEP segment relative to that of homogeneous DNA, $\mathbf{u}_i$ is the longitudinal vector of its reference frame, a is the length of the segment, and $\mathbf{R}(\mathbf{v}_{i+1}, \theta_{0,i})$ is the matrix that performs a rotation of $\theta_{0,i}$ around $\mathbf{v}_{i+1}$. The behaviour of the i-th duplex segment then differs from that of the standard TEP model only when $\theta_{0,i} \neq 0$, $B_{1,i} \neq B_1$, or both.

2.2 Simulation techniques

The oxDNA model, introduced in section 1.6, is used throughout the thesis to model DNA. In particular, MT experiments are simulated using the molecular dynamics backend of the oxDNA simulation program, as discussed in detail in (Machinek et al. 2014). In these simulations, Newton's equations of motion stemming from the potential eqn. 1.3 are integrated with the velocity Verlet algorithm (Swope et al. 1982), obtaining a system trajectory through the phase space. The simulations in this thesis use an integration time step $\mathrm{d}t = 6.6$ fs $= 0.002$ simulation units time. As oxDNA does not explicitly model the solvent, direct integration of the equations of motion would lead to nonphysical ballistic motion, affecting both dynamic and static properties of our system. Instead, the system must be coupled to the solvent, acting as a heat bath, by using a thermostat to generate a more realistic diffusive dynamics.

In this thesis, we use a Langevin-like dynamics thermostat (Allen and Tildesley 1991), adding noise and damping particle velocities to simulate the impact of the solvent.

In all simulated MT experiments we take a diffusion coefficient $D = 2.5$ simulation units. This diffusion coefficient is set to deliberately accelerate diffusive dynamics to be about 2 orders of magnitude faster than the measured diffusion of individual nucleotides (Lapham et al. 1997). This enables the simulation to access longer time-scales, thereby accelerating the convergence of the average values of quantities measured during simulations (Russo et al. 2009).

The implicit solvent and the increased diffusion makes it somewhat non-trivial to map time-scales in oxDNA to those observed in experiments. Using the simulation time unit 3.03 ps, obtained by combining simulation unit energy, mass, and length, as the relevant time-scale takes into account the microscopic motion of the nucleotides subject to the potential, but not the sped-up diffusion, and therefore it is likely to grossly underestimate the simulated time-scale. Likewise, just comparing the diffusion coefficients of individual nucleotides is likely to overestimate the accessed times, since it neglects interactions between different nucleotides. Another way of equating the time-steps is to measure a diffusion coefficient in simulation, and match this to experiment, as recently done for 2 kbp plasmids by (Coronel et al. 2018). This leads to an equivalent simulation time of $\sim$70 ps per simulation time unit (equivalent to 0.14 ps per simulation time step), about a factor of 23 greater than the microscopic interaction time of 3.03 ps, but smaller than the diffusive time-scale of single nucleotides by a factor of five. Under the assumptions that duplexes in MT experiments have a similar behaviour to that of supercoiled plasmids, their relevant simulation times should also be similar, so we will use this time unit in the rest of the thesis when dealing with simulated MT experiments. However, care should always be taken in interpreting these overall times.

2.3 Supercoiled DNA

In this section we want to describe the basic terminology involved in supercoiled DNA. As discussed in section 1.4.1.4, DNA responds to external torques in a number of ways, such as forming supercoiled structures, tightening or loosening the double helix,

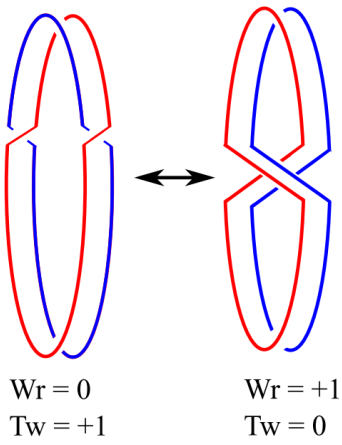

Figure 2.1: Interconversion between twist Tw and writhe Wr in a closed circular ribbon. Both configurations have a linking number Lk=1. From (Matek 2014)

or breaking base-pairs. This is due to the doubly-linked nature of DNA duplexes: the linking number Lk is the number of turns stored in the duplex, and it is a topological invariant when the strands are not allowed to cross each other and their ends are constrained, such as if they are joined to form a circular molecule or tethered to an impenetrable, non-rotating plane. If we consider two strands of coordinates described by the curves $\mathbf{r}_1(s)$ and $\mathbf{r}_2(s)$, with the curvilinear coordinate s, we can write

$$\mathrm{Lk} = \frac{1}{4\pi} \oint_0^L ds \oint_0^L ds' [\mathbf{t}_1(s) \times \mathbf{t}_2(s')] \cdot \frac{\mathbf{r}_1(s) - \mathbf{r}_2(s')}{|\mathbf{r}_1(s) - \mathbf{r}_2(s')|^3} \,. \tag{2.8}$$

where $\mathbf{t}_1(s) = d\mathbf{r}_1(s)/ds$ and $\mathbf{t}_2(s') = d\mathbf{r}_2(s')/ds$. The duplex can accommodate the Lk by either twisting the strands around each other, or by writhing the molecular axis in the 3D space. Each of these quantities has an associated topological number, respectively called twist Tw and writhe Wr, whose sum is the conserved Lk according to the famous Fuller-White-Călugăreanu relation (Kamien 2002) (see fig. 2.1)

$$\mathrm{Lk} = \mathrm{Tw} + \mathrm{Wr} \,. \tag{2.9}$$

If we define the versor connecting the single strands to each other $\hat{\mathbf{u}} = [\mathbf{r}_1(s) - \mathbf{r}_2(s)]/|\mathbf{r}_1(s) - \mathbf{r}_2(s)|$, the centre line of the double strand $\mathbf{r}(s) = [\mathbf{r}_1(s) - \mathbf{r}_2(s)]/2$, and its tangent vector $\mathbf{t} = d\mathbf{r}(s)/ds$, then (Berger and Prior 2006)

$$\mathrm{Tw} = \frac{1}{2\pi} \oint_0^L ds \mathbf{t}(s) \cdot \left[\hat{\mathbf{u}} \times \frac{d\hat{\mathbf{u}}(s)}{ds} \right] \tag{2.10}$$

$$\mathrm{Wr} = \frac{1}{4\pi} \oint_0^L ds \oint_0^L ds' [\mathbf{t}(s) \times \mathbf{t}(s')] \cdot \frac{\mathbf{r}(s) - \mathbf{r}(s')}{|\mathbf{r}(s) - \mathbf{r}(s')|^3} \,. \tag{2.11}$$

2.4 OxDNA magnetic tweezers protocol

Simulating molecular tweezers (MT) experiments with the oxDNA model is the focus of most of this thesis. In this section, we describe the different simulation protocols, pointing out differences as needed. The protocols are similar to the protocol used in previous work by (Matek et al. 2015).

In all simulations, an oxDNA duplex with a contour length $L = 600$ bp was simulated at an effective NaCl concentration of 100 mM, except for some simulations of chapter 4 where 150 mM was used. First, we generated a double-helical configuration of 624 bp with a Python script. 12 bp at each end of the duplex were used as handles, in order to constrain the linking number of the duplex, its spatial position, and to apply the pulling force F. The straight configuration was overwound or underwound as needed to reach the desired σ.

On one end of the duplex, referred to as the *surface* end, the nucleotides in the handle were kept in a fixed position by applying harmonic traps, to simulate the presence of the fixed surface on which to tether the duplex: each nucleotide was subject to an external potential

$$V_{\text{trap}}(\mathbf{r}_n; \mathbf{r}_{n,0}) = \frac{1}{2} \sum_{i=1}^{3} k^i_{\text{trap}} (r^i_n - r^i_{n,0})^2 \tag{2.12}$$

where $\mathbf{r}_n = (r^1_n, r^2_n, r^3_n)$ is the centre-of-mass position of the n-th nucleotide in the harmonic trap, and the position of the trap $\mathbf{r}_{n,0} = (r^1_{n,0}, r^2_{n,0}, r^3_{n,0})$ is set to the initial centre-of-mass position of the nucleotide. The stiffnesses k^i_{tarp} are set for all nucleotides in the handle to 571 N/m, or 10 simulation units force constant.

To prevent any part of the strands from going over the handle on the surface end of the duplex, thereby changing the linking number of the strands, a repulsion plane external potential was added to all nucleotides not in the handle

$$V_{\text{plane}}(\mathbf{r}_k; \mathbf{R}) = \frac{1}{2} k^{\text{plane}} (\mathbf{r}_k - \mathbf{R}) \cdot \hat{\mathbf{o}})^2 \tilde{\theta}(-(\mathbf{r}_k - \mathbf{R}) \cdot \hat{\mathbf{o}}) \tag{2.13}$$

with $\mathbf{R}$ set to the position of the nucleotide in the first strand and in the handle closest to the duplex centre, $\mathbf{r}_k$ is the position of the k-th nucleotide in the body of the duplex, the step function $\tilde{\theta}$, and $\hat{\mathbf{o}}$ a versor perpendicular to the repulsion plane, taken in this context to be collinear to the helical axis of the molecule. The k^{plane}

stiffness was also set to be 571 N/m. In order to prevent nonphysical behaviour of the duplex, the plane is implemented to satisfy Newton's third law, so that when the plane acts on a nucleotide outside of the handle, the nucleotide applies a force equal in magnitude and opposite in direction to the plane, and therefore to the relative reference nucleotide. This was observed to be especially necessary to prevent duplex collapse when simulating L-DNA, as in chapter 5.

The external potentials applied on the other handle of the duplex are similar to the ones described above, with the following differences. First, this end of the duplex is tethered to the freely-moving optical/magnetic bead, and is therefore referred to as the *bead* end of the duplex. The optical traps constrain the nucleotides only in the $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ coordinates, lying in the plane perpendicular to the helical axis, but leave the coordinate parallel to the collinear axis $\hat{\mathbf{x}}_3$ free to move. This effectively sets $k^3_{\text{trap}} = 0$ in V_{trap}. Second, a constant external force $F/2$ is applied to each of the outermost nucleotides on the handle, to model the presence of an external pulling force totalling F. Third, during underwinding and overwinding simulations, the harmonic traps[1] were also moving in circles, as described below.

The initial straight configuration has an helical repeat of 10.03 bp/turn, corresponding to a superhelical density $\sigma = 0.05$ given the equilibrium helical repeat d=10.55 bp/turn in oxDNA2 (Snodin et al. 2015; Snodin 2015). To generate systems at a different supercoiling density, simulations were run where the reference positions of the harmonic traps on the bead end were moving, circling the helical axis, under a pulling force F. This added or removed turns to the duplex, effectively changing σ. A rotation period of 10^7 simulation steps per turn, equivalent to 1.4 μs, was used for simulations at $\sigma < 0.2$, reported in chapter 4 and the first part of chapter 6. When investigating L-DNA and P-DNA, in chapter 5 and the second part of chapter 6 respectively, faster rotations with a period of 10^6 simulation steps (140 ns) were used for convenience. As a rule of thumb, the period should be slow enough not to cause the formation of denaturation bubbles or plectonemes beyond those that would be observed at the target F and σ, since these structures can take a long time to disappear during the equilibration of simulations. Then, configurations at the desired

[1] here and in the following, the movement of a trap is the movement of its reference position $\mathbf{r}_{n,0}$

σ and at the same pulling force F were extracted at the appropriate simulation time, and simulated with fixed harmonic traps also on the bead end. Unwinding simulations were repeated in at least three independent replicas, and simulations at fixed σ were ran starting from configurations taken from each of the three replicas. At times, when under/overwinding simulations were not available at a given force for some σ, simulations were started from the last configuration available from simulations at the same σ and at the closest F available, and preferentially in the same phase (post-buckling, pre-buckling, torsionally melted): this minimises the equilibration time, which can be particularly long when changing phase to/from the post-buckling phase, or when the length of the plectoneme tail (if present) changes significantly.

2.5 The Fye-Benham model for Stress Induced Duplex Destabilisation (SIDD)

When a duplex is sufficiently twisted, it will break some base-pairs in order to form denaturation bubbles. Breaking the base-pairs comes with an energy cost, but bubbles have a lower twist stiffness than base-paired DNA, so that when the excess linking number is high enough, bubble formation becomes energetically favourable. While this process can be studied with oxDNA, it is also useful to consider simpler statistical models that are effectively fit to experiment, and can identify the location and extent of SIDD sites in DNA under torsion. Perhaps the best known model is that by (Fye and Benham 1999), which predicts the location and extent of Stress Induced Duplex Destabilisation (SIDD) sites in torsionally constrained underwound DNA, either circular or linear, in the absence of external forces. It has been extensively applied to different DNA systems, from plasmids (Fye and Benham 1999) to minicircles (Sutthibutpong et al. 2016), with quantitative agreement with experimental results. The model is also implemented in a web-server (Bi and Benham 2004), which makes it particularly accessible. A program implementing the model is also available for download. The model will be extended in section 4.6 and applied to to investigate the impact that the location of a thymine dimer can have on plasmid denaturation.

In the model, the configuration of a DNA duplex is stored with the binary variables n_j, each set to 1 if the j-th base-pair is denatured and to 0 otherwise. The ΔLk turns that are subtracted to a topologically constrained duplex can then be stored by opening and twisting denaturation bubbles, or by twisting and writhing the rest of the base-paired duplex. Unlike formed base-pairs, denaturation bubbles do not force any helicity, such that if n is the total number of denatured base-pairs then the linking number difference stored by denaturation will be $-n/A$, with A=10.4 bp/turn. Furthermore, the unpaired, floppy single-strands in a bubble are expected to twist back on themselves, storing $\mathcal{T}$ additional turns. Therefore, we write

$$\Delta\mathrm{Lk} = -\frac{n}{A} + \mathcal{T} + \Delta\mathrm{Lk_r}\,, \tag{2.14}$$

with $\Delta\mathrm{Lk_r}$ the residual linking number difference, stored in the base-paired DNA as twist and writhe. Each of this terms has its respective contribution to the system Hamiltonian

$$H = H_c + H_t + H_r\,. \tag{2.15}$$

The H_c term, taking into account the chemical energy of base-pair denaturation, is

$$H_c = a\sum_{j=1}^{N} n_j(1-n_{j+1}) + \sum_{j=1}^{N} b_j n_j = \sum_{j=1}^{N}[(a+b_j)n_j - a n_j n_{j+1}] \tag{2.16}$$

with the first summation counting the number of denatured bubbles in the duplex, and the second summation taking into account the energy needed to open all the denatured base-pairs. The nucleation free energy required to initiate a bubble a has been measured to vary between 10 and 12 kcal/mol, depending on solvent conditions (Benham and Bi 2004). The denaturation free energies depend on the chemical nature of the base-pair, and are derived from the formula

$$b_j = \Delta H_j \left(1 - \frac{T}{T_{m_j}}\right) \tag{2.17}$$

with the enthalpy variations $\Delta H_{\mathrm{AT}} = 7.2464$ kcal/mol and $\Delta H_{\mathrm{GC}} = 9.0172$ kcal/mol and the the base-pair melting temperatures

$$T_{m_{\mathrm{AT}}} = 354.65 + 16.6\log_{10}(s)\ \mathrm{K} \tag{2.18}$$

$$T_{m_{\mathrm{GC}}} = 395.65 + 16.6\log_{10}(s)\ \mathrm{K} \tag{2.19}$$

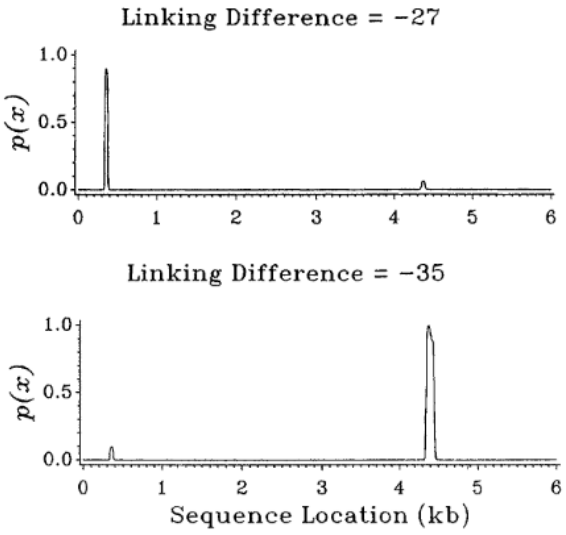

Figure 2.2: Denaturation profile of a 6 kb region containing the hen histone H5 genes at different linking number differences ΔLk, from (Benham and Bi 2004). As ΔLk goes from -27 to -35, the dominant denaturation bubble moves from position 400 to position 4400.

where s is the monovalent cation molar concentration of the solution.

If τ_j is the twisting angle within the j-th base-pair-step, the amount of twist relaxed by the denatured regions of the duplex is given by

$$\mathcal{T} = \sum_{j=1}^{N} \frac{n_j \tau_j}{2\pi} \tag{2.20}$$

and the energy associated with it is simply

$$H_t = \frac{C_{\rm ss}}{2} \sum_{j=1}^{N} \tau_j^2 \,. \tag{2.21}$$

with $C_{\rm ss} = 9.3$ pN nm^2 the torsional modulus of denatured DNA(Benham 1992; Bauer and Benham 1993). Finally, the rest of the twist will be relaxed partly by the base-paired duplex with twist and writhe, at the energy cost

$$H_r = \frac{C \Delta {\rm Lk}_{\rm r}^2}{2} = \frac{C}{2} \left(\Delta {\rm Lk} + \frac{n}{A} - \mathcal{T} \right)^2 \,. \tag{2.22}$$

The total energy is therefore

$$H = \sum_{j=1}^{N} \frac{C_{\rm ss} n_j \tau_j^2}{2} + \frac{C}{2} \left(\Delta {\rm Lk} + \frac{n}{A} - \sum_{j=1}^{N} \frac{n_j \tau_j}{2\pi} \right)^2 + \sum_{j=1}^{N} [(a + b_j) n_j - a n_j n_{j+1}] \,. \tag{2.23}$$

The state of the duplex is therefore uniquely determined by the τ_j and n_j, and the probability associated to each such state can be computed in the canonical ensemble as $\exp(-\beta H)$. The model can then be used to estimate several quantities, such as the average number of denaturation bubbles at given solvent conditions, or the probability of denaturation for each base-pair. The denaturation response is sometimes non-trivial: for example, as shown in fig. 2.2, the preferred denaturation site can vary with ΔLk.

Have you ever been in a crowded room with fancy dressed people, something slightly off in your outfit, hoping that no-one would notice it? Maybe you had mismatched socks or a ripped stocking. Well, maybe you were lucky enough not to be called on it, but DNA sequences are often not so lucky: plectonemes tend to be pretty vocal when something is odd. And that might be a good thing...

3

Statistical Mechanics of plectonemes in inhomogeneous DNA

Contents

Several theoretical models have been proposed to model the behaviour of DNA in magnetic tweezers (MT) experiments, introduced in chapter 1. Loop energetics can be well described by a simple circular model, which despite its simplicity provides good agreement with experiments. This can then be used in more general models of MT experiments, such as the model of the buckling transition by (Brutzer et al. 2010). Most of these approaches treat DNA as a homogeneous polymer or curve, thus neglecting the impact that sequence-dependent stiffness or local bending can have on loop energetics, with some notable exceptions (Kim, Ganji, et al. 2017; Goyal and Perkins 2008). However, many biologically relevant DNA loci (Kahn 2014) (including mismatches (Dittmore and Neuman 2018), modified cytosine (Ngo et al. 2016), A-tracts (Dršata et al. 2014) and more) are significantly less stiff than average

DNA. Similarly, significant bending is a features of DNA photoproducts (Park, Lee, et al. 2007) and TATA boxes (Davis et al. 1999), while promoter sequences can have both lowered stiffness and a distinct bending (Kanhere and Bansal 2005). The mere presence of these anomalies might significantly alter the supercoiled structure of DNA, and this alteration could be exploited to regulate biological activities such as DNA damage repair, gene regulation, and more.

In this chapter, we present a model to predict the impact that these anomalous sequences can have on loop energetics, and apply it to study plectoneme pinning, a phenomenon observed in simulations (Matek 2014) and experiments (Brutzer et al. 2010; Dittmore, Brahmachari, et al. 2017; Kim, Ganji, et al. 2017), as well as on the buckling transition, building on the model of (Brutzer et al. 2010). First, section 3.1 introduces the basics of homogeneous loop energetics, mainly focusing on the worm-like chain (WLC) model. Then, a simple model of the buckling transition for a homogeneous circular loop is described in section 3.2. Section 3.3 presents a general framework for the impact that anomalous sequences can have on the energetics of a loop, applied to the special case of piecewise-circular loops in section 3.4. The application of the model to plectoneme pinning is then presented in section 3.5 and confronted with results of a modified version of the TEP model of (Brackley, Morozov, et al. 2014). Section 3.6 treats the impact of an anomalous sequence on the buckling transition. Finally, section 3.7 presents the conclusions.

3.1 Loop energetics in the WLC model

The energetics of plectoneme loop formation are a central theme in both this chapter and chapter 4. In this section we review this topic as treated in the framework of the WLC model, with particular regard to the circular loop approximation.

In the WLC model, DNA is treated as a homogeneous, flexible polymer of contour length L with a bending modulus B, without any further internal structure. Despite its crudeness, the approximation has been successfully applied in numerous contexts where structural details can be neglected, mostly when the contour length is much larger than the DNA bending persistence length $p = B/k_BT$, with the Boltzmann constant k_B and the temperature T (Strick, Dessinges, et al. 2003; Marantan

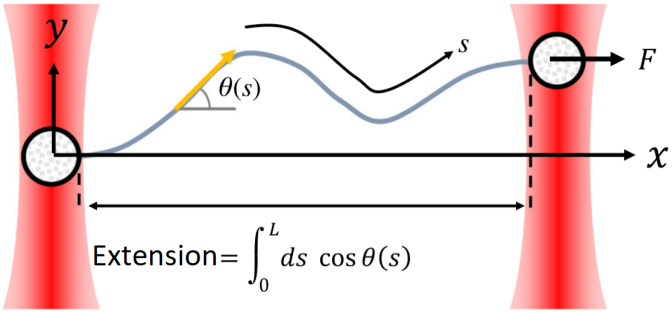

Figure 3.1: A DNA duplex in an optical tweezers experiment modelled as a worm-like chain (WLC), adapted from (Marantan and Mahadevan 2018).

and Mahadevan 2018). If we consider the duplex to be in a MT experiment, with one fixed in space and the other free to respond to a force F, then the Hamiltonian of the system in fig. 3.1 is

$$H = \int_{-L/2}^{L/2} \left[\frac{B}{2} \dot{\theta}^2 - F \cos\theta \right] ds \,, \tag{3.1}$$

with θ the angle between the direction of the force and the infinitesimal DNA segment ds, and with $\dot{\theta} = d\theta/ds$.

The Euler-Lagrange equation corresponding to the Hamiltonian 3.1 is

$$\ddot{\theta} = \lambda \sin\theta \quad \text{with} \quad \lambda = \sqrt{\frac{B}{F}} \tag{3.2}$$

which, separating variables, gives

$$s - s_0 = \lambda \int_{\theta(s_0)}^{\theta(s)} \frac{d\theta'}{\sqrt{2(C - \cos\theta')}} \tag{3.3}$$

for some constant C and values of the curvilinear coordinate s and s_0. The right hand side is the well studied incomplete elliptic integral of the first kind, which does not in general have a solution that can be expressed in terms of elementary functions.

If the DNA forms a loop, as when twisted to a post-buckling state, we need to look for a solution with boundary conditions $\theta(-L/2) = 0$ and $\theta(L/2) = 2\pi$, so that it must be $C = 1$. With some algebra, one can show (Schiessel 2014) that the solution is given by the parametric functions

$$z(s) = s - 2\lambda \tanh\left(\frac{s}{\lambda}\right) \tag{3.4}$$

$$x(s) = 2\lambda \left(1 - \frac{1}{\cosh\left(\frac{s}{\lambda}\right)}\right) \tag{3.5}$$

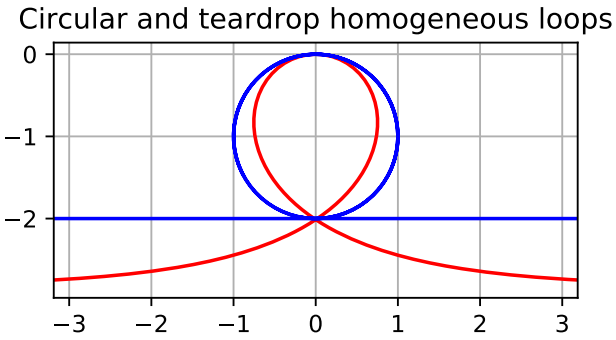

Figure 3.2: Comparison of circular and teardrop-shaped loop for the same values of F and B. The loops have similar shape and size. Both loop conformations are independent of σ, and depend on F only in scale.

This yields a teardrop-like loop, with the associated energy

$$E_{\text{loop}} = k\sqrt{BF} \tag{3.6}$$

with $k \approx 8$. Instead, the circular loop of minimum energy has $k = \sqrt{8\pi^2} \approx 8.8$ and radius $R = \sqrt{B/2F}$. As shown in fig. 3.2, the loops have similar sizes and shapes.

3.2 Buckling transition for homogeneous DNA

As discussed in section 1.4.1.4, a DNA duplex in a MT experiment forms a plectoneme loop at the buckling supercoiling density σ_b^p. As σ increases above σ_b^p, the plectoneme grows, forming a tail. The energetics of this transition has been analysed by (Brutzer et al. 2010), which introduce a phenomenological model able to deduce the location of σ_b^p from the chain length, the salt concentration, and other experimental features. Within the same framework, the chain extension L, the post-buckling torque Γ, and the extension post-buckling slope can also be estimated. This model will also be used in chapters 4, and 6.

In the pre-buckling state, the free energy of a duplex of length L_0 with a helical repeat d storing N complete turns is that of a torsional spring (Moroz and Nelson 1997; Marko 2007)

$$E_{\text{pre}} = \frac{1}{2}\frac{C_s L_0}{d^2}(2\pi)^2\sigma^2 = \frac{1}{2}\frac{C_s}{L_0}(2\pi)^2 N^2\,. \tag{3.7}$$

The duplex will writhe to accommodate part of the excess linking number, but this writhing can be approximated by defining an effective torsional modulus (Moroz and Nelson 1997)

$$C_s = C\left(1 - \frac{C}{4pk_BT}\sqrt{\frac{k_BT}{pF}}\right) \tag{3.8}$$

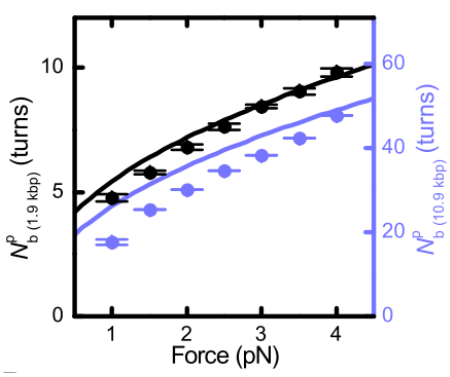

Figure 3.3: Excess linking number at the buckling transition N_b^p as predicted by eqn. 3.11 (line) and as measured in experiments (points) on duplexes of 1.9 kbp (black) and 10.9 kbp (blue), both in 320 mM Na^+. From (Brutzer et al. 2010).

Upon buckling, the duplex relaxes $N_{\text{plect}} \geq 1$ excess turns in the plectoneme, specifically one excess turn in the plectoneme loop at an energy cost E_{loop}, and potentially $N_{\text{plect}} - 1$ turns in the plectoneme tail, each at an energy cost E_{tail} per loop. The energy of the post-buckling state is therefore

$$E_{\text{post}} = E_{\text{loop}} + E_{\text{tail}}(N_{\text{plect}} - 1) + \frac{1}{2}\frac{C_s}{L_0}(2\pi)^2(N - N_{\text{plect}})^2\,, \tag{3.9}$$

and minimising this energy with respect to N_{plect} yields the minimum post-buckling energy

$$E_{\text{post, 0}} = E_{\text{loop}} + E_{\text{tail}}\left(N - \frac{1}{2}\frac{E_{\text{tail}}}{\frac{C_s}{L_0}(2\pi)^2} - 1\right)\,. \tag{3.10}$$

At the buckling transition we have $E_{\text{post, 0}} = E_{\text{pre}}$, so that comparing equations 3.10 and 3.7 yields the number of turns N_b^p and superhelical density σ_b^p at the buckling transition

$$N_b^p = \frac{E_{\text{tail}}}{\frac{C_s}{L_0}(2\pi)^2} + \sqrt{\frac{2(E_{\text{loop}} - E_{\text{tail}})}{\frac{C_s}{L_0}(2\pi^2)}} \tag{3.11}$$

$$\sigma_b^p = N_b^p\frac{d}{L_0} = \frac{E_{\text{tail}}}{\frac{C_s}{d}(2\pi)^2} + \sqrt{\frac{2d^2(E_{\text{loop}} - E_{\text{tail}})}{C_s L_0(2\pi)^2}}\,. \tag{3.12}$$

Equation 3.11 has been shown (see fig. 3.3) to provide good agreement with experiments for F between 1 pN and 4 pN, in duplexes of length 1.9 kbp and 10.9 kbp, at 320 mM Na^+, when using the circular loop energy corrected for writhe fluctuations

$$E_{\text{loop}} = 2\pi\sqrt{2BFz}\,, \tag{3.13}$$

with the relative zero-torque extension

$$\tilde{z} = 1 - \frac{1}{2}\sqrt{\frac{pF}{k_BT} - \frac{1}{32}}\,, \tag{3.14}$$

and the plectoneme tail free energy (Marko 2007)

$$E_{\text{tail}} = 2\pi\Gamma = 2\pi\sqrt{\frac{2P(F - \sqrt{\frac{k_B T F}{p}})}{1 - \frac{P}{C_s}}}\,. \tag{3.15}$$

The plectonemic twist stiffness P has been measured to be 31, 28, 24, 21 k_BT nm for 30, 60, 170, 320 mM $[Na^+]$ (Brutzer et al. 2010) respectively. The simulations in this thesis are all carried out with a salt concentration of 100 or 150 mM, so we use the value $P = 26k_BT$ nm when analysing our simulations. Other values of P within the 28-24 k_BT nm range have also been tested, but the differences in the estimates thus obtained is much smaller than the simulation errors.

The good agreement with experiment from this relatively simple model, based on a circular loop approximation, is encouraging, as it suggests that we can use it to compare to oxDNA simulations of homogeneous DNA, and to extend it to cases of inhomogeneous DNA.

3.3 Inhomogeneous loop energetics

The formation energy of a loop is a key parameter in the model of the buckling transition by (Brutzer et al. 2010) (see section 3.2). We now extend the model of section 3.1 to take into account the presence within the loop of an anomalous sequence of length L_1. In the following, a sequence will be referred to as *anomalous* if it is characterised by either:

- A substantial natural bending angle $\theta_0 \neq 0$ between the two ends of the sequence in the relaxed state. Such a sequence will be said to be *naturally bent*, or simply *bent*.

- An anomalously low bending stiffness $B_1 < B$. Such a sequence will be said to be *floppy*.

- Both a natural bending angle $\theta_0 \neq 0$ and a bending stiffness $B_1 < B$. Such a sequence will be said to be both *bent* and *floppy*.

The energetics discussed in the rest of the section are independent of the origin of these effects, and can be applied to model effects including the floppiness induced by the presence of mismatches, or TATA boxes, or the bent angle induced by sequence effects (Nguyen et al. 2015), or to the presence of DNA damages such as thymine dimers, which we will see in section 4.4.2. In this section, we show that they can also be obtained through general scaling arguments, which apply for the most part to arbitrary loop functional forms. A different derivation, applied to the case of a piecewise-circular loop, is presented in the next section.

First let's consider the Euler-Lagrange equation 3.2. Multiplying both sides by $2\dot{\theta}$, integrating over s and taking the square root yields

$$\dot{\theta} = \lambda^{-1}\sqrt{2(C - \cos\theta)}\,, \tag{3.16}$$

with $C = 1$ as discussed in section 3.1. Plugging this in the Hamiltonian eqn. 3.1

$$H = \int_{-L/2}^{L/2} \left[\frac{B}{2}\dot{\theta}^2 - F\cos\theta\right] ds \tag{3.17}$$

$$= \int_{-L/2}^{L/2} \left[F(1 - 2\cos\theta)\right] ds \tag{3.18}$$

shows that the energy density of the minimum energy configuration is linear in the pulling force. The energy of an unbuckled configuration is obtained by taking the solution $\theta(s) = 0$

$$H_{straight} = \int_{-L/2}^{L/2} -F ds = -FL \tag{3.19}$$

so that the energy cost of looping the DNA is

$$E_{\text{loop}} = H - H_{straight} = 2F\Delta L_{loop} \tag{3.20}$$

where ΔL_{loop} is the difference in the end-to-end distance between the pre-buckling and post-buckling states. The factor of 2 stems from the fact that in order to create a loop one should not only do work against the pulling force F, but also bend the DNA. The key message of eqn. 3.20 is that the two energy costs turn out to be equal.

This is a general result that applies not only for the teardrop solution, but for any other functional form where the characteristic length scale λ of the loop is a

parameter that can be varied to minimise the energy, whether or not the bending is harmonic or the rest bending angle is zero.

We can rewrite the loop energy in terms of the loop length scale λ, an energy density $\mathcal{E}$, and geometric prefactors k and m

$$\tilde{E}_{\text{loop}} = k\,\lambda\,\mathcal{E} \tag{3.21}$$

$$\mathcal{E} = F + m\frac{B}{\lambda^2}\,. \tag{3.22}$$

For a circular loop, $\lambda = R$, $k = 2\pi$, $m = 1/2$. Notice that eqns. 3.21 and 3.22 are merely dictated by generic dimensionality arguments, so will hold for any scalable shape. To minimise the energy in eqn. 3.21 we set

$$\frac{\partial \tilde{E}_{\text{loop}}}{\partial \lambda} = k\left(\mathcal{E} + \lambda\frac{\partial \mathcal{E}}{\partial \lambda}\right) = 0 \tag{3.23}$$

which with some algebra leads to the minimum energy and the relative length scale

$$\lambda = \sqrt{\frac{mB}{F}} \tag{3.24}$$

$$\mathcal{E} = 2F \tag{3.25}$$

$$E_{\text{loop}} = k\sqrt{\frac{mB}{F}}2F = 2k\sqrt{mBF}\,. \tag{3.26}$$

Eqn. 3.24 is kind of obvious, since it tells us that the relevant length scale of the system can be obtained by properly combining B and F. In contrast, eqn. 3.25 shows that the energy density is independent on the features of the chain, and eqn. 3.26 how the scaling in force and stiffness is true regardless of the loop functional form.

In the postbuckled state, the angle θ covers an angle of 2π. Therefore, we can write the loop energy as:

$$E_{\text{loop}} = \int_0^{2\pi} \frac{\partial E_{\text{loop}}}{\partial \theta} d\theta \tag{3.27}$$

and we can immediately deduce that, if a segment in the loop has natural rest angle of θ_0, there is no energy cost associated to it: we only need to pay bending energy to cover the remaining $2\pi - \theta_0$ radians. If the loop geometry has the most bent region at its tip (or is equally bent everywhere) and has vertical symmetry, then we can write

$$E_{1,\text{loop}} = 2\int_{\pi/2+\theta_0}^{2\pi} \frac{\partial E_{1,\text{loop}}}{\partial \theta} d\theta = f(\theta_0) E_{\text{loop}} \tag{3.28}$$

and, for a circular loop where the curvature is constant, we get

$$\Delta E(\theta_0) = \frac{\theta_0}{2\pi}\mathcal{E}R = \frac{\theta_0}{2\pi}E_{\text{loop}}\,. \tag{3.29}$$

With similar considerations, we can also get an expression for the loop energy gain for a floppy sequence. The energy density is the same everywhere, but the bending angle gained per unit energy is inversely proportional to the bending stiffness B. Usually, a circular loop would take a length $k\lambda$, but over a segment of length L_1 and stiffness B_1 the gain in angle is actually B/B_1 as high as elsewhere. This means that the contour length is shortened by $L_1(B/B_1 - 1)$, and we have

$$E_{1,\text{loop}}(B_1) = \mathcal{E}\left[k\lambda - L_1\left(\frac{B}{B_1} - 1\right)\right] \tag{3.30}$$

so that the loop energy gain is

$$\Delta E(B_1) = \mathcal{E}L_1\left(\frac{B}{B_1} - 1\right) = 2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right)\,. \tag{3.31}$$

Comparing eqns. 3.31 and 3.29, where we have highlighted the dependence on the energy linear density $\mathcal{E}$, shows why bent sequences and mismatches present a loop energy gain with a different dependence on F: while the energy gain due to a bent sequence is proportional to R, the energy gain due to a mismatch isn't. As $R \propto F^{1/2}$, this leads to the difference.

3.4 The inhomogeneous loop in the circular approximation

The previous section uses general scaling arguments to obtain the difference between the energies of an homogeneous and an anomalous loop. Here, we use energy minimisation to obtain the same formulas (eqns. 3.29 and 3.31) for the specific case of a piecewise-circular loop. The theory will then be used in the following sections to show the impact that this has on plectoneme pinning and the buckling transition.

The starting point is the circular loop energy model discussed in section 3.1. Recall that the Hamiltonian of a DNA duplex of length L pulled with a force F is

$$H = \int_{-L/2}^{L/2}\left[\frac{B}{2}\dot{\theta}^2 - F\cos\theta\right]ds\,, \tag{3.32}$$

with B the DNA bending modulus and θ the angle between the direction of the force and the infinitesimal DNA segment ds. The characteristic teardrop-like shape of fig. 3.2 can be approximated by a circular loop joined to a straight line. In this case, the energy is

$$\tilde{E}_{\text{loop}} = 2\pi R \left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2} \right) . \tag{3.33}$$

with $\tilde{z}$ the relative DNA extension. Minimising this with respect to R yields the following values for the loop radius and the energy:

$$R = \sqrt{\frac{B}{2F\tilde{z}}} \tag{3.34}$$

$$E_{\text{loop}} = 2\pi\sqrt{2BF\tilde{z}} . \tag{3.35}$$

The interplay between the decreasing loop length $2\pi R$ and the increasing energy density (between parentheses in eqn. 3.33) causes E_{loop} to be ultimately proportional to $\sqrt{F}$, rather than scaling linearly with F. If we take the usual value (Moroz and Nelson 1997) $B = 45 k_B T$ nm, then R varies from 15.4nm (a circle of 291 bp) for $F = 0.5$ pN to 5.01 nm (a circle of 95 bp) for $F = 4$ pN.

As a key simplifying approximation, we assume that a loop containing a weak DNA segment of length L_1 will take on a piece-wise circular configuration, with two different curvature radii R and R_1 in the homogeneous and anomalous DNA parts respectively. Multiple anomalous sequences can be present in the same loop, and the total energy difference will be the sum of the energy differences induced by each anomalous sequence, assuming the loop is long enough to accommodate all of them. This stems from the locality of the Hamiltonian eqn. 3.32. Let us first generalise the circular loop model for a floppy sequence. The total contour length of the loop will now be $L_1 + L_2$, with $L_2 = (2\pi - \alpha)R$ the length of regular DNA within the loop, and $\alpha = L_1/R$ the angle covered by the anomalous sequence. The total energy of the loop will now be

$$\begin{aligned} \tilde{E}_{1,\text{loop}} =& (2\pi - \alpha) R \left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2} \right) + L_1 \left(F\tilde{z} + \frac{1}{2}\frac{B_1}{R_1^2} \right) \\ =& (2\pi - \alpha) R \left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2} \right) + L_1 \left(F\tilde{z} + \frac{1}{2}\frac{B_1 \alpha^2}{L_1^2} \right) . \end{aligned} \tag{3.36}$$

Minimising eqn. 3.36 with respect to R still yields the same $R = \sqrt{B/2F\tilde{z}}$, and minimising it with respect to α gives

$$R_1 = \frac{L_1}{\alpha} = R\frac{B_1}{B}\,.$$

Therefore, the minimum loop energy is

$$E_{1,\text{loop}} = \left[2\pi R - L_1\left(\frac{B}{B_1} - 1\right)\right]\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) \tag{3.37}$$

$$= E_{\text{loop}}\left[1 - \frac{L_1}{2\pi R}\left(\frac{B}{B_1} - 1\right)\right] \tag{3.38}$$

$$= 2\pi\sqrt{2BF\tilde{z}} - 2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right)\,, \tag{3.39}$$

so that the loop energy difference due to the weakened sequence is given by

$$\Delta E(B_1) = E_{\text{loop}} - E_{1,\text{loop}} = 2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right)\,, \tag{3.40}$$

in agreement with eqn. 3.31. The total contour length given by $L_2 = (2\pi - \alpha)R + L_1 = 2\pi R - L_1(B/B_1 - 1)$ is equal to that of a regular loop when $B = B_1$, but can decreases significantly as B_1 decreases. For example, for $B_1 = 0.35B$ (equivalent to a 10 bp mismatch, as seen in section 3.5), the circumference of the inhomogeneous loop varies from 92.3 nm (277 bp) for $F = 0.5$ pN to 27.0 nm (81 bp)for $F = 4$ pN, as opposed to 96.8 nm (293 bp) and 31.6 nm (95 bp) for the homogeneous case. The energy gain is linear in F since, while the energy density is unchanged, the loop is shortened by a length $L_1(B/B_1 - 1)$.

Similarly, if the weak sequence has the same bending stiffness as the rest of the chain, but is naturally bent at an angle $\theta_0 \neq 0$, then eqn. 3.36 becomes

$$\tilde{E}_{1,\text{loop}} = (2\pi - \alpha)R\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) + L_1\left(F\tilde{z} + \frac{1}{2}\frac{B(\alpha - \theta_0)^2}{L_1^2}\right)\,. \tag{3.41}$$

and minimising this yields the same value of R and

$$\alpha = \frac{L_1}{R} + \theta_0\,. \tag{3.42}$$

The minimum loop energy is now

$$\begin{aligned} E_{1,\text{loop}} =& \left(2\pi - \frac{L_1}{R} - \theta_0\right)R\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) + L_1\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) \\ =& (2\pi - \theta_0)R\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) \\ =& (2\pi - \theta_0)\sqrt{2BF\tilde{z}} \end{aligned} \tag{3.43}$$

so that the loop energy difference is

$$\Delta E(\theta_0) = E_{\text{loop}} - E_{1,\text{loop}} = \theta_0\sqrt{2BF\tilde{z}} = \frac{\theta_0}{2\pi}E_{\text{loop}}, \tag{3.44}$$

just as shown in eqn. 3.29.

The total contour length given by $L_2 = (2\pi - \alpha)R + L_1 = 2\pi R - \theta_0 R$. Sketches of the loop conformation that minimises the energy $E_{1,\text{loop}}$ in the two cases according to the theory are shown in fig. 3.4. The energy difference turns out to be very simple

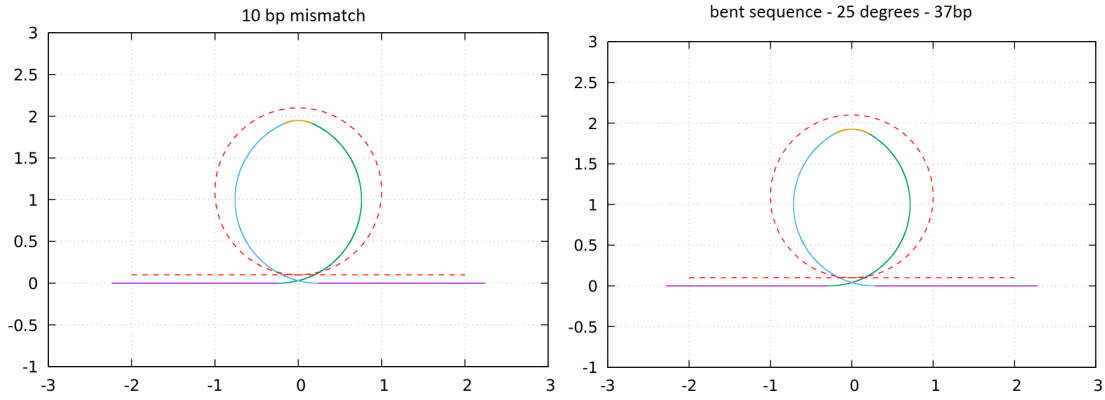

Figure 3.4: Sketches of a circular loop of homogeneous DNA (dashed red), a loop containing a 10-bp mismatch (left) and a loop containing a $\theta_0 = 35°$ naturally bent sequence with $L_1 = 37$bp (right), both at $F = 1$ pN. The anomalous sequence (yellow) is at the top of both loops, with homogeneous DNA completing the loop on both of its sides (teal and green) and straight DNA (purple). The contour length of all three DNA systems (loop + straight DNA) is the same. These shapes are independent of σ.

and elegant: this follows fundamentally from the quadratic nature of the elastic Hamiltonian, see section 3.3. It does not depend on L_1 (as long as $L_1 < L_2$), but only on the total angle. This result should also hold if there is more than one bent region in the loop. For example, the energy difference for two sequences bent in the same direction within the loop will have their angles add linearly in the expression above. Likewise, sequences with natural bending in opposing directions are going to partially neutralise each other's effect, provided they share a loop. Notice that, while the energy density does not change, the contour length changes proportionally with R.

Finally, if a sequence has both a non-zero rest angle θ_0 and a stiffness $B_1 < B$, the energy can be written as a generalisation of eqn. 3.36 and eqn. 3.41

$$\tilde{E}_{1,\text{loop}} = (2\pi - \alpha)R\left(F\tilde{z} + \frac{1}{2}\frac{B}{R^2}\right) + L_1\left(F\tilde{z} + \frac{1}{2}\frac{B_1(\alpha - \theta_0)^2}{L_1^2}\right), \tag{3.45}$$

minimising it gives $R = \sqrt{B/2F\tilde{z}}$ yet again and

$$\alpha = \frac{L_1}{R}\frac{B_1}{B} + \theta_0 \,, \tag{3.46}$$

which leads to

$$\begin{aligned} E_{1,\mathrm{loop}}(B_1, \theta_0) &= E_{\mathrm{loop}} \left[1 - \frac{L_1}{2\pi R} \left(\frac{B}{B_1} - 1 \right) \right] \\ &= (2\pi - \theta_0)\sqrt{2BF\tilde{z}} - 2F\tilde{z}L_1 \left(\frac{B}{B_1} - 1 \right) , \end{aligned} \tag{3.47}$$

Therefore, the energy difference $\Delta E = E_{\mathrm{loop}} - E_{1,\mathrm{loop}}$ is simply the sum of the two contributions

$$\Delta E(B_1, \theta_0) = \theta_0\sqrt{2BF\tilde{z}} + 2F\tilde{z}L_1 \left(\frac{B}{B_1} - 1 \right) . \tag{3.48}$$

Notice that $E(\theta_0)$, $E(B_1)$, and $E(\theta_0, B_1)$ have a different functional dependence on F. The force has a stronger effect for mismatches and other inhomogeneities that change the bending constant than it does for intrinsic curvature changes. The fundamental reason for this difference is that the energy density of the loop depends only on the magnitude of the pulling force F, and the length of the loop gets shortened by a factor independent from R when the weak segment is naturally bent, and by a factor proportional to R when the weak segment contains a mismatch (i.e. is more flexible), as discussed in the next section.

3.5 Plectoneme pinning

The simple theory presented in section 3.4 shows how an anomalous sequence influences the loop energy. In this section we will discuss how the change in loop energy causes plectonemes to preferentially locate on anomalous sequences, for values of σ and F where plectonemes are observed.

A plectoneme with an anomalous sequence within the loop will have an energy cost lower by ΔE than that of a plectoneme without an anomalous sequence within the loop. We define the plectoneme pinning coefficient Q as the ratio of the probabilities of having a plectoneme centred around an anomalous sequence and on a sequence of the same size, but far from any anomalous sequences. This can be written as

$$Q = \frac{\mathrm{Prob}(x \in L_1)}{\mathrm{Prob}(x \notin L_1)} = \exp(-\beta\Delta E) \,. \tag{3.49}$$

Studying plectoneme pinning in duplex is therefore a way to extract information about ΔE. We can also use simulations of the pinning as a way of testing the theory presented in section 3.4. The plectoneme pinning coefficients associated with the loop energy gains $\Delta E(\theta_0)$, $\Delta E(B_1)$ and $\Delta E(B_1, \theta_0)$ can be obtained respectively by plugging eqns. (3.40), (3.44) and (3.48)check it worked properly into eqn. 3.49).

$$Q(\theta_0) = \exp\left[\beta\theta_0\sqrt{2BF\tilde{z}}\right] \tag{3.50}$$

$$Q(B_1) = \exp\left[\beta 2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right)\right] \tag{3.51}$$

$$Q(B_1, \theta_0) = \exp\left[\beta 2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right) - \beta\theta_0\sqrt{2BF\tilde{z}}\right] . \tag{3.52}$$

In order to test the theory, we need to measure the plectoneme pinning distributions at different F and for several different B_1 and θ_0. Since obtaining these distributions from oxDNA requires long simulations, we choose instead to treat anomalous sequences with the extended TEP model of section 2.1.

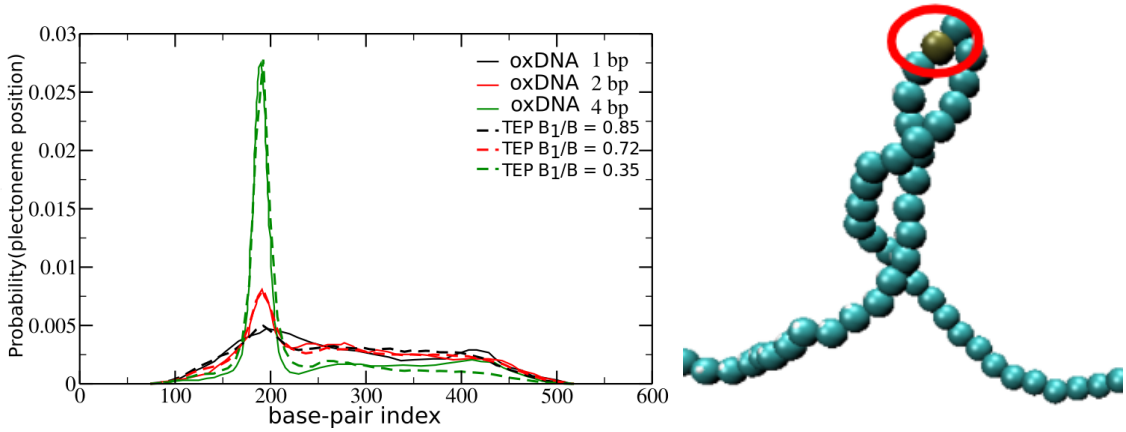

Figure 3.5: Plectoneme position probability distributions for 1 bp (black), 2 bp (red) and 4 bp (green) consecutive mismatches in oxDNA, in a 600 bp duplex with a pulling force F=2 pN at 100 mM [NaCl] and T = 300 K under $\sigma = -0.07$ (continuous) and TEP simulations for the same system containing matching floppy segments (dashed). Adapted from (Lorenzo Rovigatti, unpublished). Right: detail of a simulation snapshot of a plectoneme in the TEP model. The red circle shows the location of the segment with $B_1/B = 0.35$ (dark).

In order to make our theory more biologically relevant, we wish to test it for the values of B_1 relative to mismatches of varying length. To determine them, we first run oxDNA simulations on a 600bp duplex at biological conditions, then we run the same system again with the TEP model with a single bending term with $B_1/B \neq 1$,

in order to determine how to model each number of consecutively mismatched base-pairs[1]. The oxDNA simulations were performed by Christian Matek, while the TEP simulations and fit to the oxDNA distributions were performed by Lorenzo Rovigatti. Their results are presented in fig. 3.5 and table 3.1.

#bp	B_1/B	$B/B_1 - 1$
1	0.85	0.176
2	0.72	0.389
4	0.53	0.887
10	0.35	1.857

Table 3.1: Ratios B_1/B and coefficients $B/B_1 - 1$ for different mismatches, as measured from oxDNA simulations at $T = 300$ K and [NaCl]=150 mM. These values have been shown to reproduce the plectoneme pinning coefficients obtained by oxDNA when applied to a single TEP discrete segment (see fig. 3.5).

With the TEP model, Lorenzo Rovigatti also simulated a 2 kbp long strand with a force F =1, 3 pN and a superhelical density σ =0.03, 0.045 excess twists per turn respectively, for at least 10 days (more than 4×10^9 time steps, corresponding to hundreds of seconds if converted using the model's internal units of measurements). Note that to first order, the pinning coefficient Q does not depend on the length of the DNA (although it needs to be long enough so that end effects are not important). The comparison of theory and simulation is reported in figs. 3.6 and 3.7. Our theory shows encouraging agreement with simulations of bent sequences with different values of θ_0 and L_1, especially for F=3 pN, while the agreement is semi-quantitative for F=1 pN. In particular, the plots show that $Q(\theta_0)$ is indeed independent on L_1 within numerical errors, as predicted by the theory. Note that, as evident from fig. 3.7 a 10 bp mismatch and a sequence with $\theta_0 = 25°$ have very similar Q at 1 pN, but a different one at 3 pN, because the energy differences scale differently with F. While the theory and pinning coefficients agree well for bent sequences, they agree less well for sequences with a $B_1 < B$, made to represent mismatches. In particular at higher forces the linear dependence of ΔE on F for mismatches should lead to

[1]This can be done under the assumption that the effect of the mismatch can be captured well enough by a harmonic potential. The mismatch is not expected to generate a $\theta_0 \neq 0$, but the bending potential might in principle have a non-quadratic dependence on the bending angle. The quadratic assumption is good enough to produce a reasonable agreement between theory and simulation, as shown in the rest of the section

larger increases in the pinning coefficient than we observe. This discrepancy could be either because of limits in the theory or lack of resolution in simulation.

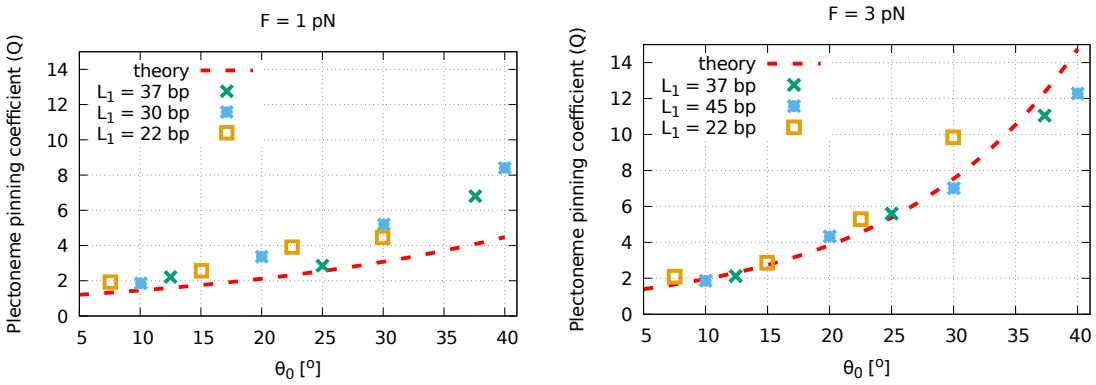

Figure 3.6: Plectoneme pinning coefficients Q as a function of the rest bending angle θ_0 for $F = 1$ pN (left) and $F = 3$ pN (right). Data points are obtained with simulations of the TEP model with the rest angle θ_0 equally spread among 3, 4, and 5 DNA segments (respectively equal to a length of the naturally bent sequence $L_1 = 22$, 30, 37 bp). The theoretical prediction (red line) is independent of L_1.

3.6 Buckling transition and anomalous sequences

The Brutzer model, discussed in section 3.2, establishes a relationship between the energy cost of forming a circular loop E_{loop}, the cost of storing one unit of writhe in the plectoneme tail E_{tail}, and the superhelical density at the buckling transition σ_b^p

$$\sigma_b^p = \frac{E_{\text{tail}}}{\frac{C_s}{d}(2\pi)^2} + \sqrt{\frac{2d^2(E_{\text{loop}} - E_{\text{tail}})}{C_s L_0 (2\pi)^2}} \tag{3.53}$$

where C_s is the Morotz-Nelson effective twist stiffness, L_0 is the duplex length, and d=10.4 bp is the number of base-pairs per helical turn. If the loop harbours an anomalous sequence, the loop energy will contain the correction given by eqn. 3.48, so that σ_b^p will be lowered by

$$\Delta\sigma_b^p = \sqrt{\frac{2d^2}{C_s L_0 (2\pi)^2}} \left(\sqrt{E_{1,\text{loop}}(B_1 = B, \theta_0 = 0) - E_{\text{tail}}} - \sqrt{E_{1,\text{loop}}(B_1, \theta_0) - E_{\text{tail}}} \right) \tag{3.54}$$

A plot of σ_b^p for different systems is shown in fig. 3.8. There is therefore a thin region of the phase diagram where plectonemes are energetically favoured only on an anomalous sequence.

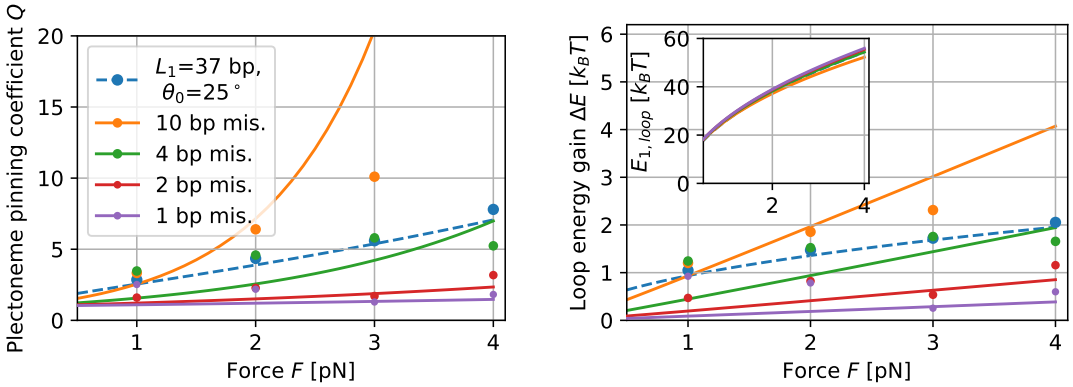

Figure 3.7: Plectoneme pinning coefficients Q (left) and loop energy gains ΔE (right) for a system with a L_1=37 bp and θ_0 =25° (blue dashed line) and a system with a single 10 bp (yellow), 4 bp (green), 2 bp (red) and 1 bp (violet) mismatch, as obtained with theory (solid lines) and TEP model simulations (points). While a highly bent sequence pins plectonemes more than a mismatch at low forces, the mismatch causes stronger pinning at higher forces. The inset shows the absolute loop energy $E_{1,\mathrm{loop}}$ for each system, as a function of force.

For example, for $\sigma = 0.045$ in a 2 kbp duplex, plectonemes in typical B-DNA would disappear for $F > 2.5$ pN, but for a duplex containing a 10 bp mismatch not till $F > 2.75$ pN, giving a range of about $\Delta F = 0.25$ pN where the plectoneme is primarily localised on the defect. Plectonemes that do not contain the anomalous sequence will only be present as statistical fluctuations, and their probability will be much lower. This effect can also be described with an effective plectoneme pinning coefficient Q, as done in the previous section, but its origin is significantly different. Previously, plectonemes containing and not containing an anomalous sequence are energetically more favoured than the B-DNA state. Now, only plectonemes on the anomalous sequence are favoured. While we have not investigated this with TEP simulation, this effect is explored with oxDNA in section 4.4.5.

3.7 Conclusions

This chapter presents a model of loop energetics in the presence of sequences that are significantly bent, floppy, or both (sections 3.3 and 3.4), uses it to predict plectoneme pinning on such sequences (section 3.5), and applies it to extend the model of the buckling transition (section 3.6) developed by (Brutzer et al. 2010) and discussed in section 3.2.

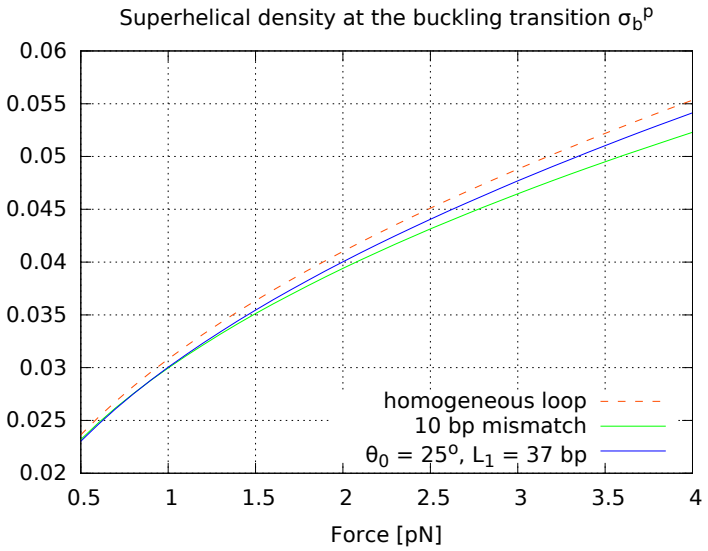

Figure 3.8: Supercoiling density at the buckling transition σ_b^p for a homogeneous loop (red dashed), one with a 10 bp mismatch (green), and a sequence with a rest angle $\theta_0 = 25°$ and length L_1=37 bp, in a 2 kbp DNA duplex at 150 mM [NaCl]. The theory predicts a measurable difference in σ_b^p depending on the structure of the loop.

Our theory of loop energetics predicts a number of non-trivial features:

- The presence of a bent sequence generates a difference in loop energy ΔE proportional to the square root of the pulling force F, while a floppy sequence is associated with an energy gain linear in F.

- ΔE is independent of the length of the anomalous sequence L_1 if the sequence is bent, while it is linear in L_1 if the sequence is floppy or both bent and floppy.

- Sequences that are both floppy and bent are expected to have a ΔE given by the sum of the contributions due to bending and floppiness, that can therefore be considered independent. We have not tested this in TEP, but the pinning observed for duplexes containing a thymine dimer (TD) presented in chapter 4 supports this statement.

We have applied this theory to study the impact of an anomalous sequence on plectoneme pinning and on the buckling transition. Our theory of plectoneme pinning is quite simple, but it works remarkably well compared to TEP simulations, suggesting that we have captured the dominant physics of the system. The study of how inhomogeneities and supercoiling interact is in its infancy. That means that

much work is still to be done. An obvious extension of the model would be to take into account the local bending angle of the whole sequence, abandoning the approximation that the anomalous sequence is embedded in an homogeneous polymer, extending this model in the direction of very recent work by (Kim, Ganji, et al. 2017).

Inhomogeneities can significantly pin plectonemes, as several studies have shown in the past decade (Brutzer et al. 2010; Matek 2014; Dittmore, Brahmachari, et al. 2017; Kim, Ganji, et al. 2017). This might have important biological implications, including influencing the position of nucleosomes in chromatin and facilitating the recruiting of RNA polymerase at a promoter site (Kim, Ganji, et al. 2017). More research is needed to assess which of these potential implications are actually realised.

Where it is revealed that when you're just chilling in the sun, without a care in the world, your skin cells are frantically trying not to succumb to the ultraviolet havoc that is being wrecked on their precious DNA. So be kind to them, and wear your sunscreen. Our story looks at a very minute detail in this vast and messy picture, like zooming on a single arrowhead in the Trojan war. But even an arrowhead can be important, if it lands in the right spot...

4

Thymine dimers: when DNA gets sunburn

Contents

4.1 Introduction

UV light can cause different kinds of DNA damage, catalysing reactions that radically change the structures of the involved bases. Of these lesions, the most common is the

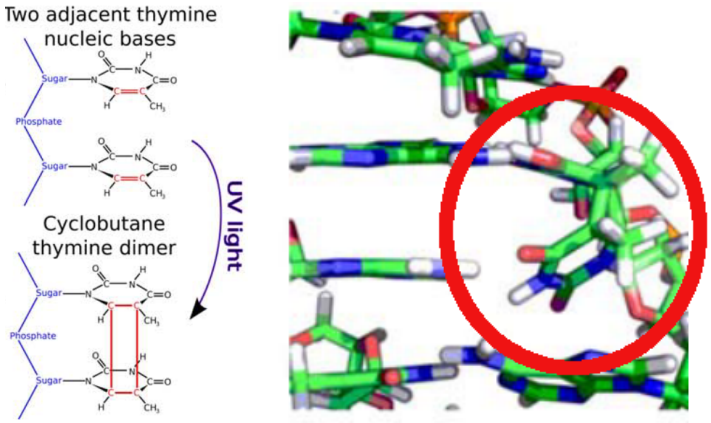

Figure 4.1: Left: schematic depiction of the formation of a *cys-sin* cyclobutane thymine dimer (TD). UV light interacts with two successive thymines along the same strand, inducing covalent bonds between the aromatic rings. Right: a TD (circled red) in a crystal structure (Park, Zhang, et al. 2002). Adapted from (Knips and Zacharias 2015).

cis-syn-cyclobutane pyrimidine dimer lesion between adjacent thymines, commonly referred to as thymine dimer (TD); the aromatic rings of the thymines are joined by covalent bonds (Morita et al. 2010) (as shown in fig. 4.1), with carcinogenic, cytotoxic and mutagenic effects if left unrepaired. These make UV light one of the major risk factors for skin cancer, including malignant melanoma (Mahroos et al. 2002).

The presence of the TD hinders gene expression by halting the RNA polymerase during the transcription phase. Moreover, the presence of the TD halts DNA polymerase: the block can be removed with the SOS response, at the price of introducing genetic mutations (Park, Lee, et al. 2007). These processes are also the source of sunburn in humans: prolonged exposure to direct sunlight causes an increase in the rate of production of DNA photoproducts such as TDs, leading to cellular death, which can be experienced as skin burn (reddening due to an inflammatory response, blistering, and peeling skin).

DNA can get damaged in many ways in the cell, and DNA repair mechanisms are crucial for cell health. The breakdown of repair mechanisms play a key role in many diseases, most notably cancer. For that reason, the study of DNA repair is a large and active field in biology. TDs are an important model system for the study of DNA repair. The importance of this system was signaled by the 2015 Chemistry Nobel prize being awarded to the Turkish-American biochemist Aziz Sancar for the the elucidation of the repair mechanisms for TD in bacteria (Sancar 2016).

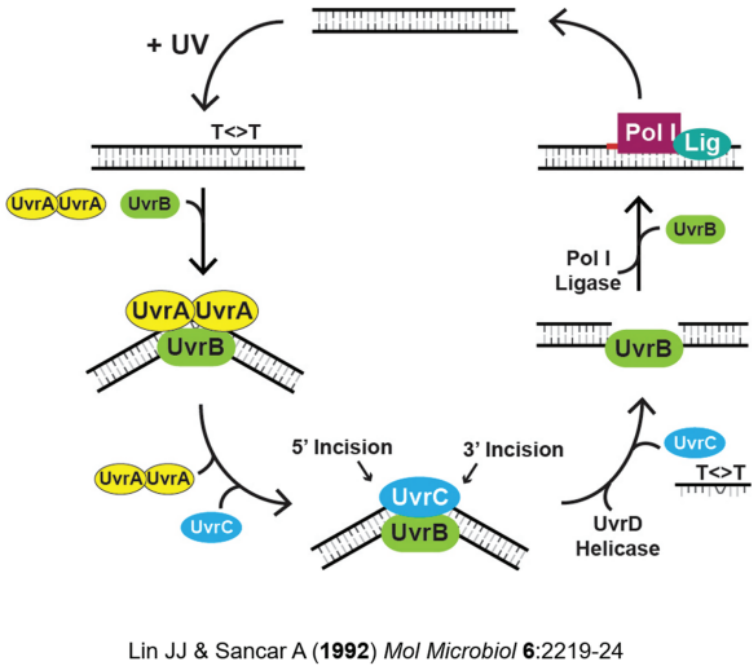

Figure 4.2: Nucleotide Excision Repair (NER) pathway in *E. Coli*. The damage is first bound by the $(UvrA)_2$ homodimer, which then recruits UvrB. UvrB forms a complex with DNA in an ATP hydrolysis dependent-reaction (not shown). The nuclease UvrC displaces $(UvrA)_2$, binds to the damaged strand, and nicks it. Finally, the helicase UvrD displaces the UvrC-TD complex, and the gap is filled by a newly synthesised strand, which is then ligated to the rest of the duplex. Reproduced from (Sancar 2016).

Bacteria can repair TDs with a photolyase enzyme, which binds to the TD-containing sequence and uses energy provided by an incoming blue light photon to break the covalent bonds connecting the two thymines in the dimer. The detailed chemistry of the repair of the TD in the fully-bound duplex in the photolyase has been the object of a recent review (Liu et al. 2015). Differently from bacteria, humans and other placental mammals can't use exposure to blue light to repair TDs. Interestingly, humans do have two homologues of the bacterial photolyase in their genome; these enzymes maintain a response to blue light, however they do not perform any DNA repair functions. Instead, they take part in the human circadian clock. The reasons leading to the loss of photolyase in placental mammals are unknown, but it has been proposed that weak natural selection might be one of them (Lucas-Lledó and Lynch 2009).

Alternatively, bacteria can use the nucleotide excision repair (NER) to repair the damage in the absence of blue light, when in presence of glucose: the proteins

UvrA, UvrB, and UvrC bind to the TD-containing sequence, and the dodecamer containing the TD is cut out of the genome by a helicase and replaced with an undamaged dodecamer of the same sequence (fig. 4.2). The NER pathway is also present in humans, where the genes UvrA-C are replaced by the RPA and XPA-G genes. Mutations in any of these genes block the TD-repair mechanism, leading to the disease *xeroderma pigmentosum* from which the genes take name: disruptions in the NER pathway causes extreme sensitivity to sunlight, including a much higher incidence of skin cancer.

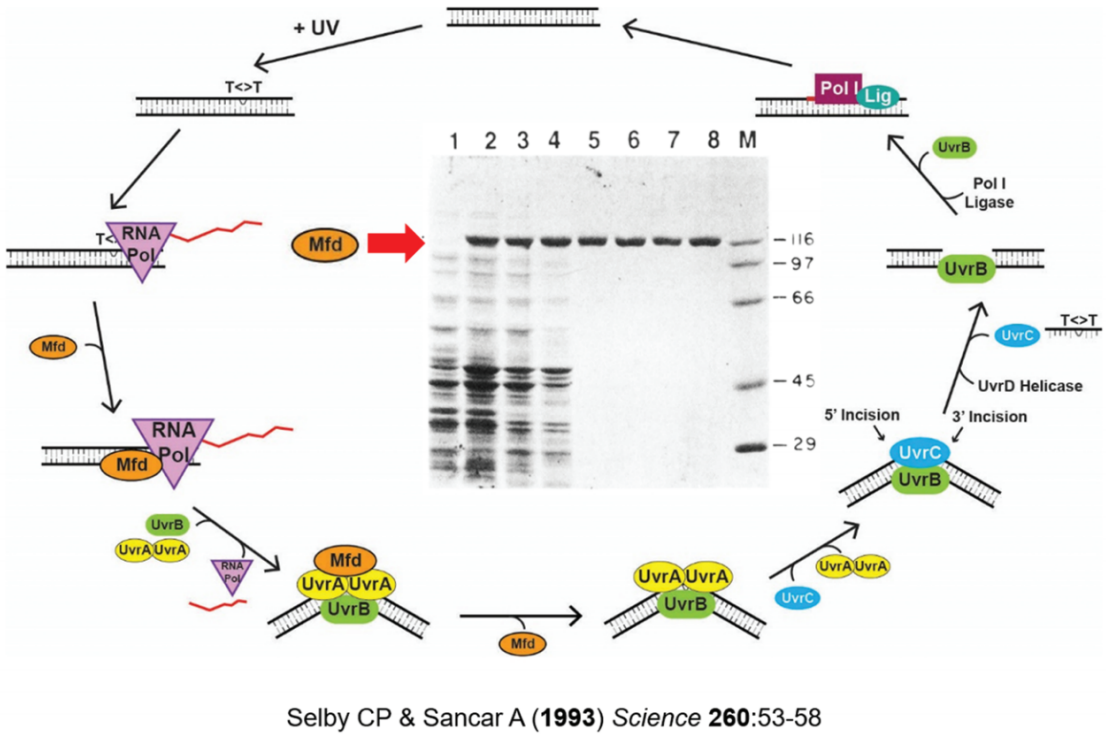

Figure 4.3: Transcription-coupled repair in *E. Coli*. A stalled RNA-polymerase (RNA-pol) forms a complex with DNA and RNA that is stable for hours. Because the TD is not exposed in the complex, repair is inhibited. The transcription-repair coupling factor (TRCF, also called mutation frequency decline, Mfd, as in figure) is a translocase that binds to the complex and displaces RNA pol and RNA. Mfd then recruits UvrA, which recruits UvrB, forming a transient Mfd-UvrA-UvrB complex. Finally, Mfd unbinds, and the damage is repaired as in the NER pathway (see fig. 4.2). From (Sancar 2016).

Finally, TDs in bacteria can also be recognised by transcription coupled repair (TCR) (fig. 4.3): a stalled RNA-polymerase recruits a transcription-repair coupling repair factor (TRCF), which in turns recruits the UvrA-C genes and initiates the NER pathway. The existence of this pathway strongly enhances TD-repairing

during transcription. In humans, the role of TRCF is taken by Transcription Factor II Human (TFIIH).

From a structural point of view, thymine dimers (TD) have been shown to locally bend the duplex and thermodynamically destabilise it, by weakening the hydrogen-bonding of the affected bases. As shown in chapter 3, both these effects should have an impact in plectoneme-localisation. The biological importance of the TD and its status as a key model system for DNA damage make it particularly interesting to assess whether it localises plectonemes. If the localisation is strong enough, this could affect how crucial DNA repair mechanisms work.

This chapter is divided as follows. Section 4.2 reviews the structural impact of thymine dimers on the structural and thermodynamic properties of DNA duplexes, as discussed in the literature. Section 4.3 describes how we model TDs in oxDNA. Section 4.4 presents simulated magnetic tweezers (MT) experiments on TD-containing DNA duplexes: the TD causes significant plectoneme pinning (as predicted by the theory presented in chapter 3), which in turn causes the TD to assume an extra-helical configurations. Some possible ways these effects may be biologically relevant are proposed in section 4.5 and section 4.6. Finally, section 4.7 reports the conclusions.

4.2 Features of thymine dimers

Over the past several decades, several experiments have studied the impact of TDs on the thermodynamics of DNA duplexes (Kemmink, Boelens, Koning, Kaptein, et al. 1987; Taylor et al. 1990; Barone et al. 1995; McAteer et al. 1998), using techniques such as UV absorbance, microcalorimetry, and the introduction of reactive chemical probes. We will review these results below.

4.2.1 A TD causes a drop in duplex melting temperature

Melting experiments on several different sequences (table 4.1) show that the presence of the TD causes a noticeable drop in the duplex melting temperature T_m. The thermodynamic data relative to these systems is summarised in table 4.2. This table provides a pretty comprehensive review of results from the literature. [1]

# bp	Sequence	Thermodinamic measures
8	5'-GCG**T-T**GCG-3'	(Kemmink, Boelens, Koning, Marel, et al. 1987)
10	5'-CGTA**T-T**ATGC-3'	(Taylor et al. 1990)
12	5'-GCACGAA**T-T**AAG-3'	(McAteer et al. 1998)
21	5'-GTG**T-T**AACGTGAGTATAGCGT-3'	(Barone et al. 1995)

Table 4.1: Sequences of duplexes containing TDs whose thermodynamic parameters have been compared with the ones of the native duplexes, and their relative references. The locations of the TDs are highlighted by the dash between the relevant thymines (all in boldface). The second strand of each duplex (not shown) completely matches the first.

# bp		ΔH (kcal)	ΔS (cal)	Experiment type	T_M(°C) at 1 mM
21	Native	186 ± 20	530 ± 60	UV melting (Barone et al. 1995)	68.1
	TD	167 ± 18	490 ± 50	5.4μM [duplex], 0.1M[NaCl],0.01M[Tris-HCl]	57.5
12	Native	-88 ± 9	-248 ± 25	UV melting (McAteer et al.	61
	TD	-87±9	-250 ± 25	1998), 0.25[NaCl], 0.1mM EDTA,0.01mM [sodium cacodylate]	53
10	Native	-64 ± 6	-178 ± 18	UV melting (Taylor et al. 1990),	56
	TD	-66 ± 7	-189 ± 19	1M [NaCl] 0.10mM [NaPh]	48
8	Native	-49 ± 3	-131 ± 7	NMR melting of T4's methyl	59
	TD	-51 ± 7	-143 ± 20	group (Kemmink, Boelens, Koning, Marel, et al. 1987), 3mM [duplex], 0.2M [NaCl], 0.05mM [KPh], 0.02%NaN_3	47

Table 4.2: Thermodynamic parameters ΔH and ΔS for several sequences, as reported in the literature for various DNA and salt concentrations, and resulting melting temperatures as computed from eqn. 4.1.

By measuring the duplex melting temperature T_M at different duplex concentrations c, the differences in enthalpy ΔH and entropy ΔS between the duplex and

[1]We note that two recent papers have been omitted from the table. Firstly, the duplex studied by (Taylor et al. 1990) has also been the object of an NMR investigation (Wenke et al. 2013). However, this work presents results which strongly disagree with (Taylor et al. 1990) and the model of (SantaLucia and Hicks 2004), which has been shown to be able to predict the melting temperature of most duplexes within a few degrees at most. This suggests that the results might be spurious, therefore we do not include it in our analysis. Similarly, another study (Rumora et al. 2008) presents a thermodynamic analysis of a 19-mer duplex, but the inconsistencies in the two estimates of the thermodynamic parameters obtained with microcalorimetry and UV melting suggest the presence of large systematic errors.

single-stranded states can be computed (Marky and Breslauer 1987) as

$$T_m = \frac{\Delta H}{R\ln(2c/x) + \Delta S} \quad \text{with} \quad x = \begin{cases} x = 4 \text{ for non-self-complementary strands} \\ x = 1 \text{ for self-complementary strands} \end{cases} . \tag{4.1}$$

In order to compare data obtained at different c, table 4.2 reports the T_M at $c =$ 1 mM DNA. The TD causes a significant drop of T_M, between 8°C and 12°C for the sequences studied. Notice that, because of the large errors in ΔH and ΔS, extrapolations of T_M at different c from that used in experiments shouldn't be relied on too much, but just be used to get a feeling for what the thermodynamic impact of the TD is.

4.2.2 A TD bends the duplex

The pioneering circularisation and electrophoretic mobility experiments of (Husain et al. 1988) and (Wang and Taylor 1991) showed that the TD causes the host duplex to bend, yet thirty years later there is still disagreement on the magnitude of the bending angle θ_0. Results from NMR studies (Kim, Patel, et al. 1995; McAteer et al. 1998; Wang and Taylor 1991) are compatible with $\theta_0 \approx 10°$ or smaller, while a circularisation assay (Husain et al. 1988) and a crystal structure (Park, Zhang, et al. 2002) show $\theta_0 \approx 30°$. Predictions from theoretical calculations have shown θ_0 ranging between 0 ° (Rao et al. 1984), 10° (Miaskiewicz et al. 1996), and 27° (Pearlman et al. 1985).It has been suggested that this degree of variation in θ_0 is due to a significant flexibility of the TD (Park, Zhang, et al. 2002; Knips and Zacharias 2015).

Another important question is how the TD fits within the duplex. The overwhelming consensus, mainly from crystal and NMR structures of short duplexes, is that the TD is not exposed to the solution in an extra-helical state, but sits in the double helix like two adjacent thymines would (Wenke et al. 2013; Knips and Zacharias 2017; Rumora et al. 2008), and that the local disruption of the duplex is limited to the bases neighbouring the TD. Several studies have also shown a local weakening of the hydrogen bonding (McAteer et al. 1998; Rumora et al. 2008; Wenke et al. 2013; Knips and Zacharias 2015).

The impact of the TD on the helical parameters of the affected base-pairs is even less clear than the overall duplex bending. A study of a crystal structure of an

unbound duplex (Park, Zhang, et al. 2002) reports that the base-pairs within the TD have a relative roll of 22°, and a tilt of 17.8°, though they unfortunately don't report any data regarding the same sequence without the TD. They also report a twist of 27° in the 5' neighbour of the TD. Recent atomistic simulations of the same sequence (Knips and Zacharias 2015) report the average helical parameters in the four base-pairs encompassing the TD; they confirm the minor-groove width and show compatible values for the increased roll and reduced twist, but do not show any tilt. By contrast, a NMR structure (McAteer et al. 1998) reports that the TD-containing base-pair has a tilt decreasing from ∼-1° in the native duplex to ∼-10° in the TD-containing duplex, a roll increasing from ≈ 0° to ≈ 26°, and a twist decreased[2] by 15°, from 36° to 21°. This change in the twist affects the global twist of the duplex as well. The backbone has also been observed to be pinched (McAteer et al. 1998).

In summary, the impact of TDs on the detailed chemical structure of DNA is still an active area of investigation. However, it is well established that the presence of the TD destabilises the duplex enough to significantly lower its melting temperature, but not enough to cause the TD to spend a significant fraction of time flipped outside of the duplex inside the solution. There is a consensus that the TD causes substantial bending in the duplex, though the exact value of this angle is unclear. The structural influence of the TD on its immediate vicinity is also unclear, but it is generally thought that it doesn't significantly disturb the chemical structure of base-pairs a few steps away. These details have all been considered when developing the model of the TD implemented in oxDNA, described in the next section.

4.3 Modelling of the TD in oxDNA

The previous section briefly reviewed what is known of the structural impact of TDs on the DNA duplex: its global structure is bent, its thermodynamic stability is lowered, and the helical parameters of the neighbouring base-pairs are changed, but those further away are unchanged. This section describes how we decided to

[2] Actually, the text most likely contains a typo, as it literally reads "the twist between the [dimer containing] T8·A17 and T9·A16 base-pairs decreases by ∼25° from ∼36° to ∼21°" (McAteer et al. 1998). However, the bottom right panel of fig. 6 in the same paper clearly shows a decrease in twist from 36° to 21°, so the decrease in twist of the TD-containing base-pair must be 15°.

incorporate this knowledge in oxDNA in order to construct a coarse-grained model of the TD. As shown in chapter 3, our theory predicts that base-pair stability (through its effect on the bending modulus B_1) and bending angle θ_0 are two factors that affect plectoneme pinning, and therefore we model TDs in a manner that can faithfully depict these features.

The covalent bonds within the TD cause a widening of the minor groove (Knips and Zacharias 2015; Park, Zhang, et al. 2002; Kim, Patel, et al. 1995), which has been associated with a global bending of the duplex towards the major groove (Park, Zhang, et al. 2002). Because the data available (Park, Zhang, et al. 2002; McAteer et al. 1998; Knips and Zacharias 2015) suggests that the base-pair-step containing TD has an increased roll, it seems natural to introduce the bending caused by the dimer by forcing a roll in oxDNA so that the minimum energy configuration for a TD-containing duplex reproduces this roll. Because θ_0 is not known precisely, we choose to develop distinct models of the TD, with either $\theta_0 = 30°$ as shown in the crystal structure of (Park, Zhang, et al. 2002), or $\theta_0 = 15°$, a value closer to the one suggested by NMR experiments.

One of the main advantages of oxDNA is a faithful depiction of the thermodynamic properties of DNA. Since the hydrogen bonding within the base-pairs containing the TD has been shown to be weakened (Park, Zhang, et al. 2002; Knips and Zacharias 2015; McAteer et al. 1998), especially the ones involving the 5' thymine (McAteer et al. 1998), we also weaken the hydrogen bonding as needed to reproduce the melting behaviour of TD-containing duplexes. We take the drop in melting temperature ΔT_m as a measure of the destabilisation induced by the TD, and aim for our model to reliably reproduce this value. In particular, we choose to parametrise our model to reproduce the ΔT_m of a DNA octamer first studied by (Kemmink, Boelens, Koning, Kaptein, et al. 1987). Its melting temperature is reported at a duplex concentration (3 mM) high enough that the melting temperature can easily be computed in oxDNA, so that they can be directly compared. By contrast, other studies either report T_M at much lower DNA concentration, which would require too expensive simulation for them to be computed in oxDNA, or they report only the

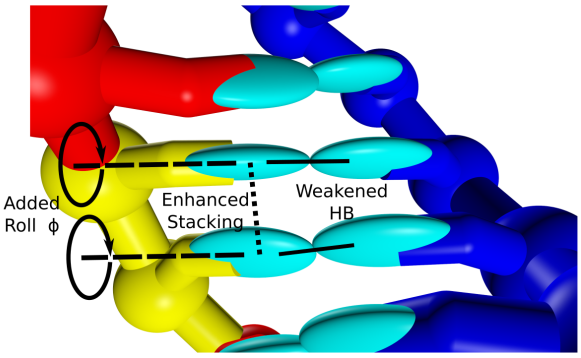

Figure 4.4: Close-up of the TD model (yellow) within the duplex. Notice the weakened hydrogen bonding interaction (continuous straight line), the enhanced stacking (dotted), and the added roll ϕ, generated by rotating the orientation matrix of the affected nucleotides around the base-pair axis (dashed) before computing the stacking interaction potential. Notice that, in TD3, only the roll in the TD-containing base-step was changed.

thermodynamic parameters ΔH and ΔS, but with errors larger than the measured variation between the native duplex and the TD-containing one.

Finally, we want to test the hypothesis that sequences with the same thermodynamic stability and bending angle have very similar plectoneme pinning properties, regardless of the details of their microscopic interaction potential. Therefore, two of the models have $\theta_0 = 15°$, but one has a stacking interaction enhanced by a factor of ten, and the bending angle is generated in a different manner, as shown in the next subsection. In total, we developed three different models of the thymine dimer, all reproducing the ΔT_m measured for the octamer of (Kemmink, Boelens, Koning, Kaptein, et al. 1987), and all with values of θ_0 compatible with the ones reported in the literature. The three models, dubbed TD1, TD2 and TD3 respectively, are discussed in more detail in the rest of the section.

4.3.1 Reproducing the duplex bending angle

Given the experimental reports of increased roll angle induced by the TD, we set a non-zero bending angle θ_0 by generating a non-zero roll angle between the base-pair steps within the thymine dimer and the one immediately adjacent to it in the 5' direction. This is done by slightly modifying the stacking term in the oxDNA interaction potential. In the oxDNA model (Ouldridge, Louis, and Doye 2011; Ouldridge, Louis, and Doye 2010), two successive nucleotides on the same strand interact depending on their separation vector and on their respective orientations with each other. To generate the roll, we change the interaction potential so that before computing the interaction energy, the orientation matrix of the nucleotide in

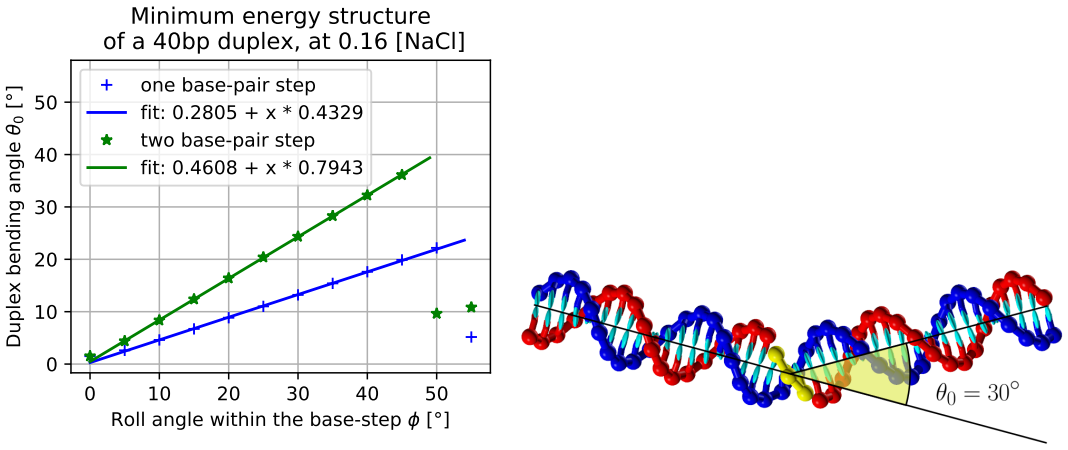

Figure 4.5: Left: duplex bending angle θ_0 as a function of the roll angle ϕ introduced either in one (blue) or two (green) base-pair steps, as measured from minimisation simulations (symbols) and as computed from a linear fit of the simulation data for $\phi > 0$ in the monotonically increasing regime. Notice that when $\phi \approx 50°$ the stacking breaks and the duplex bending angle decreases a lot. Right: snapshot from a minimisation simulation. The black lines show the direction of each half of the duplex, while θ_0 is shown in yellow. The two base-steps with a non-zero roll angle are also shown in yellow.

the 3' direction is rotated by an angle ϕ around the base vector of the nucleotide. This introduces the desired roll. A sketch of the model, showing both this modification and the ones discussed below, is shown in fig. 4.4. In turn, the roll can generate a noticeable bending of the duplex, as shown in 4.5.

To precisely measure the angle, we perform energy minimisation on a 40bp-long duplex, increasing the roll angle until θ_0 is close to 30° (for the TD1 model) and 15° for (for the TD2 and TD3 models). The angle is defined in the following way (Chatterjee et al. 2017):

- Each base-pair of the duplex is assigned to one of two groups, depending on which side of the TD it is located.
- The position of each base-pair is computed, by computing the average of the positions of the hydrogen-bond interaction sites of each base in a pair.
- We fit a line of direction $\hat{d}_{1,2}$ through the base-pairs of each group of nucleotides by a least-square distance minimisation.
- The bending angle is the angle between the lines that fit each half of the

duplex, as computed through the cosine rule:

$$\theta_0 = \arccos(\hat{d}_1 \cdot \hat{d}_2)\,.$$

A picture of the molecule exhibiting the $\theta_0 = 30°$ in the TD1 model is shown in the bottom part of fig. 4.5. The configuration of the duplex containing the TD was minimised with the steepest descent algorithm as implemented in the oxDNA program, for a total of 4×10^5 simulation steps. Notice that the bending angle didn't change appreciably after the first 2×10^5 simulation steps. Values of θ_0 measured after the minimisation, for several different values of ϕ, are shown in fig. 4.5. Setting $\phi = 37°$ and $\phi = 18.3°$ generates a bend of $\theta_0 = 30°$ and $\theta_0 = 15°$ respectively for TD1 and TD2. We choose to include this roll both in the TD-containing base-step and in its 5' neighbour. The plots were repeated after weakening the hydrogen bonding in both the base-pairs involving the TD, as discussed in the next section, without any appreciable variation. Attempts were made to bend the duplex by modifying the stacking interaction of one base-pair step instead of two, but such a modification was found to be insufficient to introduce the θ_0 required in the TD1 model.

In the TD3 model, the stacking interaction (highlighted in fig. 4.4) within the base-step containing the TD is enhanced by a factor of ten. This in turn strengthens the dependence of θ_0 with ϕ, such that setting $\phi = 27.1°$ only in the base-step containing the TD was sufficient to generate $\theta_0 = 15°$.

4.3.2 Reproducing the melting temperature drop

Next, we turn to the thermodynamic properties of the TD. The melting temperatures of the octamer studied by (Kemmink, Boelens, Koning, Marel, et al. 1987) were measured in oxDNA simulations at a duplex concentration $c = 3$ mM and at a salt concentration $s = 0.2$ [NaCl], both for native strands and with TD-containing strands. This was measured with the simulation protocol described in (Šulc et al. 2012): a single duplex was simulated in a cubic simulation box with periodic boundary conditions with the Virtual Move Monte Carlo (VMMC) algorithm in the oxDNA program, for 5 or 9 replicas of each system, until the estimated melting temperature

T_m of the octamer studied by (Kemmink, Boelens, Koning, Marel, et al. 1987)								
	Experiments	oxDNA	TD1-s	TD1	TD2-s	TD2	TD3-s	TD3
Native T_m	70 ± 2	66.4						
TD T_m	57 ± 2		56.8	53.7	66.2	52.0	67.5	53.2
ΔT_m	13±3		9.6	12.7	0.4	14.4	-1.1	13.2

Table 4.3: Melting temperatures T_m and melting temperature drops ΔT_m of the octamer studied by (Kemmink, Boelens, Koning, Marel, et al. 1987) in the native and TD-containing duplexes, as experimentally measured and as computed in oxDNA using any of the TD models. The columns marked with "-s" refer to simulations where the TD is modelled by merely changing the stacking interaction, i.e. without any weakening of the hydrogen bonding. All the oxDNA melting temperatures have an error of 0.2° or less. (Kemmink, Boelens, Koning, Marel, et al. 1987) do not report the error on ΔT_m, but it would be 3°C assuming uncorrelated errors on T_m.

Features of TD models			
	TD1	TD2	TD3
roll angle ϕ	37°	18.3°	27.1°
duplex angle θ_0	30°	15°	15°
stacking multiplier	1	1	10
# base-steps with modified stacking	2	2	1
HB weakening	10%	55%	60%
Melting temperature drop ΔT_m [°C]	12.7	14.4	13.2

Table 4.4: Main features of the TD models presented in this section. The melting temperature drop is relative to the octamer studied by (Kemmink, Boelens, Koning, Marel, et al. 1987), which has been measured to be 13±3°C(see table 4.3).

was known with a standard mean error of 0.3 or less. This took about 10^9 simulation steps per replica per particle, or about five days on a modern computer.

The TD was first modelled just by changing the stacking potential as described above. The melting temperatures T_m and melting temperature drops ΔT_M caused by the presence of the TD are listed in table 4.3, in the columns marked with "-stack". The data in table 4.3 shows that setting a non-zero ϕ destabilises the duplex, while increasing the stacking stabilises it. To reproduce the correct ΔT_m, the hydrogen bonding of both bases in the TD has to be weakened by 10%, 55% and 60% respectively for the TD1, TD2, and TD3 models. The features of the three models are summarised in table 4.4. A sketch of the modifications applied to the hydrogen bonding and stacking potentials in order to model the TD is shown in fig. 4.4.

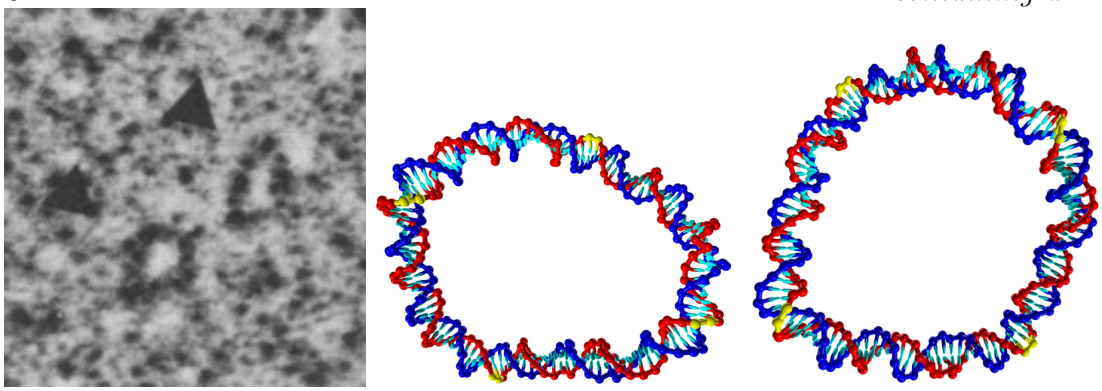

Figure 4.6: comparison of EM images of 128 bp TD-containing minicircles (left, adapted from (Husain et al. 1988)) with simulations (centre, right). Notice the preferential location of the sharply bent regions near the TDs (yellow).

4.3.3 TD-containing minicircles

As a final sanity check, we observe the impact of TDs on some of the minicircles first used in a famous study by (Husain et al. 1988) to quantify the bending angle induced by the TD. The minicircles are formed by ligation of identical 32 bp fragments, each fragment containing a TD. 128 bp minicircles imaged by electron microscopy (EM) are compared to minicircles of the same sequence simulated in oxDNA with molecular dynamics at a salt concentration of 100mM, with TDs modelled with the TD1 model. Both experiments and simulations show that the minicircles alternate sharply bent and mostly straight regions. However, simulations allow to observe explicitly what was hypothesised in the experiments, namely that these sharply bent regions are preferentially located near TDs, as shown in fig. 4.6. More generally, it would be very interesting to study how TDs affect the structure of minicircles, for example expanding on work by (Sutthibutpong et al. 2016), but we leave that question to future work.

4.4 Simulation of magnetic tweezers (MT) experiments on a 600 bp duplex containing a TD

The previous section introduces three different ways of modelling the TD in oxDNA, called respectively TD1, TD2, and TD3. All of them reproduce within error bars the melting temperature drop $\Delta T_m = 13°\text{C}$ observed in the octamer of (Kemmink, Boelens, Koning, Marel, et al. 1987), but they have different values of

the bending angle θ_0, and they have different nucleotide-level interaction potentials. The development of these models opens up the possibility of using simulations to investigate the impact of TDs on the properties of DNA systems with hundreds or thousands of base-pairs. In this section, we use these models to investigate the impact that the presence of the TD has on the properties of a 600 bp DNA duplex in a molecular tweezers experiment. For brevity, duplexes containing a TD1, a TD2 or a TD3 will be called TD1, TD3 and TD3 duplexes respectively.

4.4.1 Simulation techniques

Simulations were set up with as discussed in section 2.4: a straight duplex was slowly unwound at a constant rate of 10^7 simulation steps per turn, until reaching a superhelical density σ = -0.2, while pulled at a constant force F = 0.5 pN, 1 pN, 1.5 pN, and 2 pN respectively. Simulation snapshots from these runs were then used to start simulations at a given σ. For each F, 3 independent runs were simulated on GPU, for 5 or more days each, for a number of simulation steps between 5×10^8 and 10^9, until the measured observable (denaturation states, end-to-end distance and plectonemic state) were equilibrated (using the time-scales from section 2.2, 10^9 steps is about 140 μs). The reported values of observables extracted from simulations are the means of the time-averaged observables across all the independent runs, and the errors on each value are estimated with the sample standard deviation across the runs. When no errorbars are shown in a plot, the errors are smaller than the symbols used in the plot.

The position of the plectonemes was measured with the algorithm described in detail in the supplementary material of (Matek et al. 2015). In brief, plectonemes bring non-adjacent duplex segments into contact, either at the base of the plectoneme loop or all along the plectoneme tail, and detecting these contacts is a way to detect the relevant plectoneme. Among all the non-adjacent segments that are in contact with each other, those with the largest separation along the duplex are considered to be at the basis of the plectoneme tail, and their midpoint is the position of the plectoneme. This coincides with the position of the loop, up to a few base-pairs of difference. This algorithm cannot detect more than one plectoneme at a time, but

visual inspection showed that multiple plectonemes were never observed in simulation. For the algorithm to work, one must define a contact distance d^0_{lin} and a separation cut-off N_c to define what it means for two duplex segments to be non-adjacent. In this work, we take $d^0_{\text{lin}} = 10.65$ nm and $N_c = 40$bp, instead of $d^0_{\text{lin}} = 7.24$ nm as in (Matek et al. 2015), since this was observed to perform better at our salt concentration.

Sampling the equilibrium plectoneme position distributions required a more careful approach at $F = 2$pN (the same force investigated by (Matek et al. 2015), as shown in chapter 3): tip-bubbles (see section 1.7) were overwhelmingly present in all TD-containing simulations, which greatly slowed the diffusion of plectonemes. Two initial configurations, with plectonemes on the 200-th and 330-th base-pairs respectively, were extracted from simulations of an undamaged duplex at $F = 1$ pN and $\sigma = -0.071$, and used as starting configurations in simulations with $F = 2$ pN in 80 independent replicas each on two different duplexes: one without any TD, and one with a TD1 at 200bp from the first unconstrained base-step. The first $3{\times}10^7$ steps were discarded to allow the end-to-end distance to thermalise and the plectoneme to diffuse elsewhere, then the last 7×10^7 time-steps (about 10 μs using the time-scales of section 2.2) were used to measure the plectoneme position distributions. These distributions were used as a test set to further validate the distributions obtained from the simulations set up as follows. 94 plectoneme-containing simulation snapshots were extracted from the undamaged duplex simulations in the thermalised regime. The plectonemes in the snapshots were spaced by roughly 3bp, in the full range where plectonemes where observed (from the 150-th base-pair to the 425-th). Each snapshot was used as the initial configuration of a simulation of 10^8 steps, of which the first half was used for further thermalisation. Each of the 94 replicas was then simulated with any of the TD1, TD2, or TD3 model on 200bp from the first unconstrained base-step, or without any TD model. In this way, we significantly improved the sampling of the probability of finding a plectoneme at a particular site on the DNA strand. Each simulation step corresponds to 2×10^{-3} simulation time units. Since the simulation time unit corresponds to 70ps (as described in section 2.2), the total simulated time is 120μs.

4.4.2 Plectoneme position distributions

Fig. 4.7 shows the plectoneme position distributions at $F = 2\text{pN}$, $\sigma = -0.071$ for the undamaged duplex and for duplexes with each of the TD models on the 200th unconstrained base-step from the beginning of the duplex. A duplex with the hydrogen-bonding energy on base-pairs #200 and #201 reduced to 50% (called HB0.5), but without any bending, was also analysed to discriminate the impact of the bending from that of the lower bending stiffness. Plectonemes diffuse around the undamaged duplex with a roughly uniform probability, and show a preference for the damage sites in the damaged duplexes. The HB0.5 duplex shows a peak of the same intensity as the ones caused by statistical noise elsewhere in the duplex, but peaks in the TD duplexes are much more prominent. The fact that the peaks observed for TD2 and TD3 (two models with the same θ_0, but where the bending has different origins) are very similar suggests that the approximation of considering the loop energetics independent of the microscopic effects that generate the bending angle is justified.

As discussed in chapter 3, the plectoneme pinning coefficient can be measured by computing the ratio of the probability density associated to the plectoneme loop being located on the TD and elsewhere, i.e.

$$Q_{TDi} = \frac{p(x \in TD)}{p(x \notin TD)} \tag{4.2}$$

where x is the plectoneme position and $p(x)$ its associated probability density function.

The theory shown in chapter 3 relates Q_{TDi} to the loop energy difference ΔE due to the presence of the dimer via its associated Boltzmann factor

$$Q = \exp(-\beta \Delta E)\,, \tag{4.3}$$

with ΔE depending on B_1 and θ_0

$$\Delta E = -2F\tilde{z}L_1\left(\frac{B}{B_1} - 1\right) - \theta_0\sqrt{2BF\tilde{z}}\,, \tag{4.4}$$

where $\tilde{z}$ is the relative chain elongation, B is the bending stiffness of regular DNA, $B_1 < B$ is the stiffness of the tract containing the TD, and L_1 is its length. While $\theta_0 = 30^\circ$ for TD1 and $\theta_0 = 15^\circ$ for TD2 and TD3, the value of B_1 is unknown.

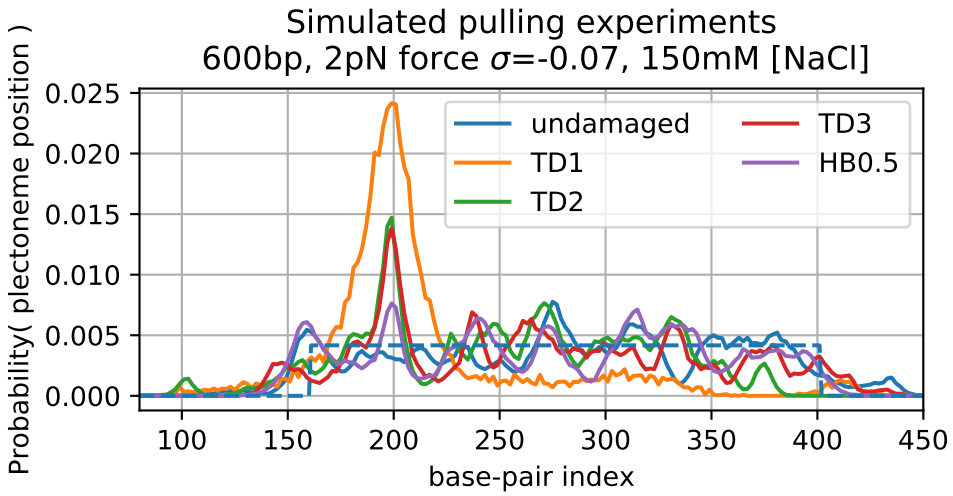

Figure 4.7: Plectoneme position distributions for the undamaged duplex, for duplexes containing a TD model starting on the 200-th unconstrained base-step, and for a duplex with hydrogen bonding within the 200-th and 201-st base-pairs weakened by 50% (labelled with HB0.5). The dashed line shows a uniform fit of the damage-free plectoneme position distribution. Notice that no plectoneme pinning is observed in the undamaged duplex, and the peak in the HB0.5 duplex is not high enough to be distinguished from the background.

The three models have roughly the same ΔT_m, so that if B_1 is determined by the thermodynamic destabilisation of the duplex then they should all have the same B_1. However, efforts to estimate the pinning induced by such a thermodynamic destabilisation alone proved fruitless, due to statistical fluctuations: the peak caused by the presence of the TD is not markedly different from a background fluctuation. One way to estimate B_1 is to assume that it is similar to the B_1 induced by a 1-bp mismatch in oxDNA, since the enthalpy reduction of the duplex state is quantitatively similar: the hydrogen bonding of a 1 bp mismatch is completely turned off, while the hydrogen bonding in both bases of the TD is attenuated by roughly 50%.

Notice that, while the two contributions in eqn. 4.4 should add up for a flexible polymer like the one modelled by TEP, it's not obvious that this should apply to oxDNA: possibly only one of the two contributions might be relevant at any given time, since the pinning induced by the floppiness of the TD might well be due mostly to an increase in prevalence of kinks on the TD, which would possibly neutralise the effects of the $\theta_0 \neq 0$. Nevertheless, keeping these caveats in mind, we consider both $B_1/B = 1$, as for regular DNA, and $B_1/B = 0.85$, as for a 1bp mismatch. We

use the length of $L_1 = 2.5$ nm, which was used in conjunction with B_1 in chapter 3 in order to reproduce the bending stiffness of mismatches.

Measuring Q in a way that can be compared with the theory is non-trivial: the theory assumes a circular loop, with ΔE independent of whether the TD is located at the top of the loop or somewhere else along its side. The peak caused by the presence of the TD in this highly oversimplified picture would be a hat function. However, fig. 4.7 shows that these probabilities are actually acutely peaked. There are several reasons for this. Firstly, of course the system has thermal fluctuations. Secondly, visual inspection shows that the presence of tip-bubbles strongly affects the geometry of the loop, so that only about 10bp are contained in the loop. To extract Q, we measure $p(x \in TD)$ as the probability density that the plectoneme loop is located exactly on the TD. The errors on this value can be conservatively estimated as the fluctuations observed in the undamaged duplex, i.e. 0.003 (see fig. 4.7).

The value of $p(x \notin TD)$ is expected to be constant, barring finite size effects. However, $p(x \notin TD)$ varies from 10^{-3} to 6×10^{-3} due to statistical fluctuations. For the undamaged duplex, $p(x \notin TD)$ can be obtained by performing a least-square fit $p(x)$ to a uniform distribution, which yields a uniform distribution with support on the bp interval $U = [153, 405]$ and a non-zero constant value of 4×10^{-3}. Assuming that the presence of the TD does not meaningfully change this support allows us to estimate $p(x \notin TD)$ by integrating $p(x)$ outside of the peak and dividing it by the number of base-pairs within the support but outside of the peak K.

$$p(x \notin TD) = \frac{\int_{x \notin K} p(x)dx}{|U| - |K|} \tag{4.5}$$

where $|U|$ and $|K|$ are the lengths in base-pair of U and K respectively, with $K = [164, 236]$ for TD1 and $K = [184, 214]$ for all the other models. While the choice of where to place the peak is somewhat arbitrary, it doesn't alter $p(x \notin TD)$ substantially, as long as all of the peak is cut out: extending K by 10bp in each direction leads to values of Q that differ by less than 0.2, well within the errors.

Values of Q and $\beta\Delta E = -\log Q$ computed from simulations are compared with theoretical predictions in table 4.5. Choosing $B_1 = 1$ consistently underestimates the measured values of Q. Instead, choosing $B_1 = 0.85B$, as one would for a

			Theory	
		Simulations	floppy + bent ($B_1/B = 0.85$)	only bent ($B_1/B = 1$)
Q	TD1	35 ± 5	37.9	26.1
	TD2	8 ± 2	7.4	5.1
	TD3	7 ± 2		
	HB0.5	4 ± 2	1.5	1
$\beta\Delta E$	TD1	-3.5 ± 0.2	-3.6	-3.3
	TD2	-2.1 ± 0.3	-2	-1.6
	TD3	-2.0 ± 0.3		
	HB0.5	-1.3 ± 0.4	-0.4	-0

Table 4.5: Plectoneme pinning coefficients Q and loop energy difference ΔE due to the TD as extracted from simulations (see fig. 4.7), and as predicted using eqn. 4.4, with $\theta_0 = 30°, 15°, 15°, 0°$ for TD1, TD2, TD3, HB0.5 respectively, as explained in the main text.

1bp-mismatch, yields a surprisingly good agreement, given the many approximations taken in the model and the large statistical errors. Interestingly, the models TD2 and TD3 show very similar plectoneme pinning despite their differences in the details of the stacking potential: this is expected by theory, since they have the same thermodynamic stability and bending angle, as described in section 4.3. The theory works less well for the HB0.5 model, but the large statistical errors (the peak due to the weakened hydrogen bonding is roughly of the same height of peaks in the background around the basepairs 240 and 310) limit the amount of information that they can yield.

4.4.3 Duplex end-to-end extension

The duplex extensions, measured as the average distance from the first and last nucleotide in the first strand of the duplex, are shown for different F and σ in fig. 4.8.

As expected, once σ goes beyond the buckling transition, the extension decreases with increasing σ as more of the strand enters into the plectoneme. For a given force, at a fixed extension, the overall extension decreases for increasing σ. This can be understood for the following two reasons

- The loops are smaller for larger force, and

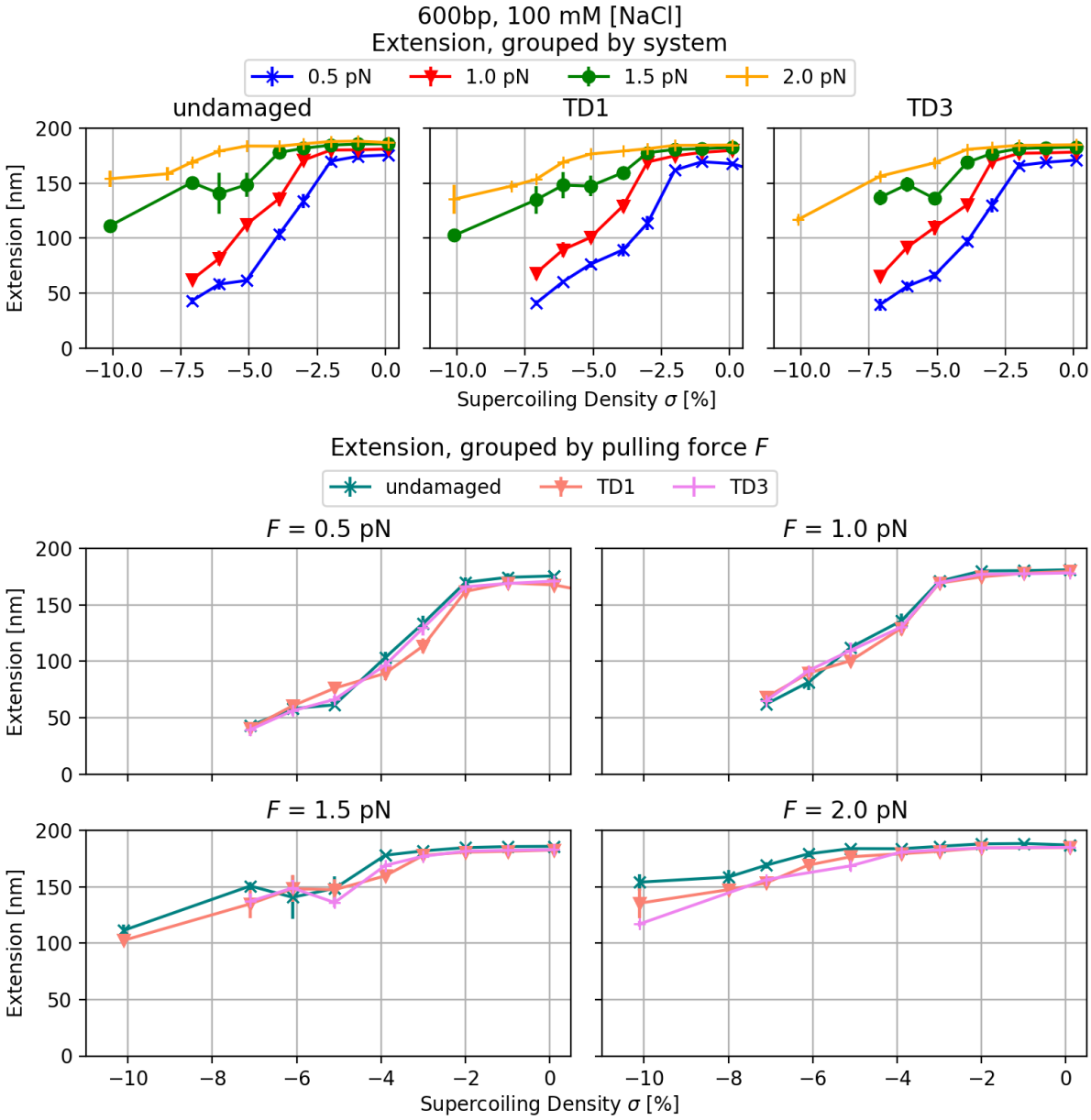

Figure 4.8: Plots of extension versus supercoiling density σ in duplexes containing (or not) a TD, grouped by TD model (top) and by value of the pulling force F (bottom).

- At larger σ especially, bubbles are also stabilised, reducing the amount of twist that gets converted into the plectonemes.

It is also interesting to compare some subtle differences between the strands with a TD in them, and those without a TD.

- Before the buckling transition, the extension of the undamaged duplex is slightly but systematically larger than that of any TD-containing duplex, possibly due to an enhancement of writhe from the TD. This is observed at all

F. It can be ascribed to the lower stability of the plectoneme-containing phases in the damage-free DNA, which prefers to relax twist by opening bubbles.

- After the buckling transition, for $F = 1.5$ pN, 2 pN, the undamaged duplex has a larger extension than the TD-containing duplexes, due to the enhanced stability of the plectoneme phase in the latter.

- After the buckling transition, for $F = 0.5$ pN, 1 pN, the trend is reversed: as predicted by the theory, the presence of the TD lowers the loop size and therefore increases the duplex extension.

The duplex extensions also shows a shoulder at $\sigma \approx -0.05$ for $F = 1.5$ pN. This is probably due to sampling errors.

4.4.4 Denaturation probability

The degree of stability of the TD-containing base-pairs may be important for the process that DNA repair enzymes use to selectively bind and repair the TD, since the TD has to be flipped outside of the duplex in order to fit in the binding pocket of the photolyase: a weak base-pairing could play a role in the binding specificity of TD-binding enzymes, as well as affect the free energy of the extra-helical state. Recent atomistic simulations (Knips and Zacharias 2015; Knips and Zacharias 2017) have investigated the impact that a TD has on the solution state of the duplex crystallised by (Park, Zhang, et al. 2002). In particular, they have shown that the TD preferentially assumes an intra-helical configuration when it's found in unconstrained duplexes, though the TD is up to 150 times[3] more likely to be extra-helical than two undamaged thymines in the same duplex.

In order to study the impact of F and σ on the denaturation probability (i.e. the fraction of time spent in an unbound state) of a TD in our models, we measured the binding state of each nucleotide in the duplex at an interval of 10^5 simulation steps, then computed the denaturation probabilities of the TD-containing base-pairs, shown in fig. 4.9. As in previous work (Šulc et al. 2012), a base-pair is considered to be formed if the binding energy is at least 10% of its maximum possible value.

[3]depending on how one draws the line between an extra-helical and intra-helical state

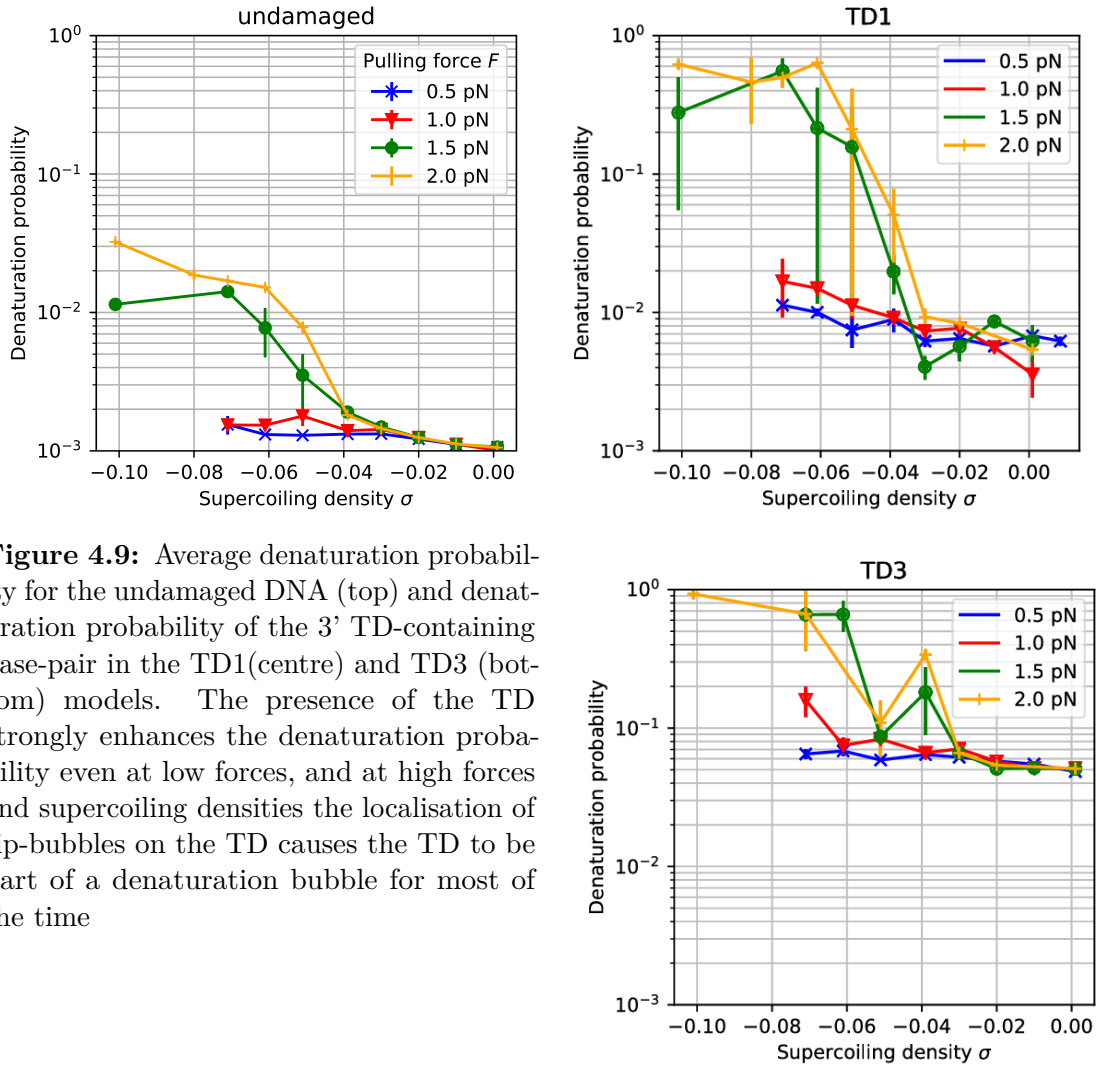

Figure 4.9: Average denaturation probability for the undamaged DNA (top) and denaturation probability of the 3' TD-containing base-pair in the TD1(centre) and TD3 (bottom) models. The presence of the TD strongly enhances the denaturation probability even at low forces, and at high forces and supercoiling densities the localisation of tip-bubbles on the TD causes the TD to be part of a denaturation bubble for most of the time

The presence of the TD strongly increases the denaturation probability at low supercoiling, from $\sim 10^{-3}$ for the damage-free DNA to $\sim 10^{-2}$ and $\sim 10^{-1}$ respectively for TD1 and TD3. For the TD1, the hydrogen bond is only 10% weaker than for the undamaged case, so the enhancement in the denaturation probability is mainly caused by the destabilisation that also causes the bend. For TD3 where the bonding is 60% weaker the probability is, as expected, higher, and the reduced strength of the hydrogen bonds plays a more important role.

Next we consider the effect of force and supercoiling on the denaturation probabilities. Firstly, for the undamaged duplex, we observe a marked increase in the average denaturation probability for $F > 1.5$ pN and $\sigma >$ -0.04. This enhancement is due to the well known effect of tip-bubbles, also observed by (Matek

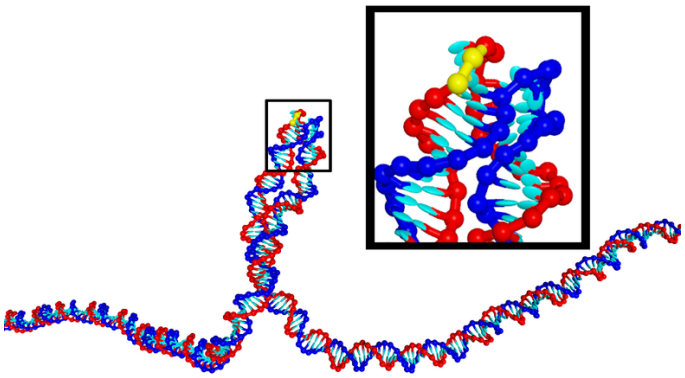

Figure 4.10: Example of a plectoneme with tip-bubble in a duplex containing a TD1 (yellow), for a pulling force $F = 2$ pN and a superhelical density $\sigma =$ -0.1. Note that at the plectoneme tip (shown in the detail) there is a bubble, not a loop.

et al. 2015). However, because the average denaturation probability remains lower, the probability that a particular base is open remains small.

A similar phenomenology can be observed for both damaged duplexes. For F = 1.5 pN, 2 pN, and for supercoiling $|\sigma| \gtrsim 0.4$, the denaturation probability rises markedly, reaching close to 100% denaturation. The reason for this is that the plectoneme is localised at the TD, and also develops a tip-bubble (see fig. 4.10).

Interestingly, even at $F = 0.5$ pN, 1 pN, the denaturation probability increases substantially with σ regardless of the presence of the TD, but the effect is much stronger for larger forces.

4.4.5 Phase behaviour

More insight can be gained by looking at different phases or states of DNA and how they vary with F and σ. These are:

B-DNA : no plectonemes nor bubbles are present.

Bubble : at least two consecutive base-pairs are denatured, according to the 10% binding energy criterion described above.

Plectoneme : A plectoneme is present anywhere along the chain, as detected with the algorithm described in section 4.4.1.

Tip-bubble : A plectoneme is present within $M_c = 20$ bp of the centre of a bubble.

Bubble+Plectoneme : both plectonemes and bubbles that are not tip-bubbles are present.

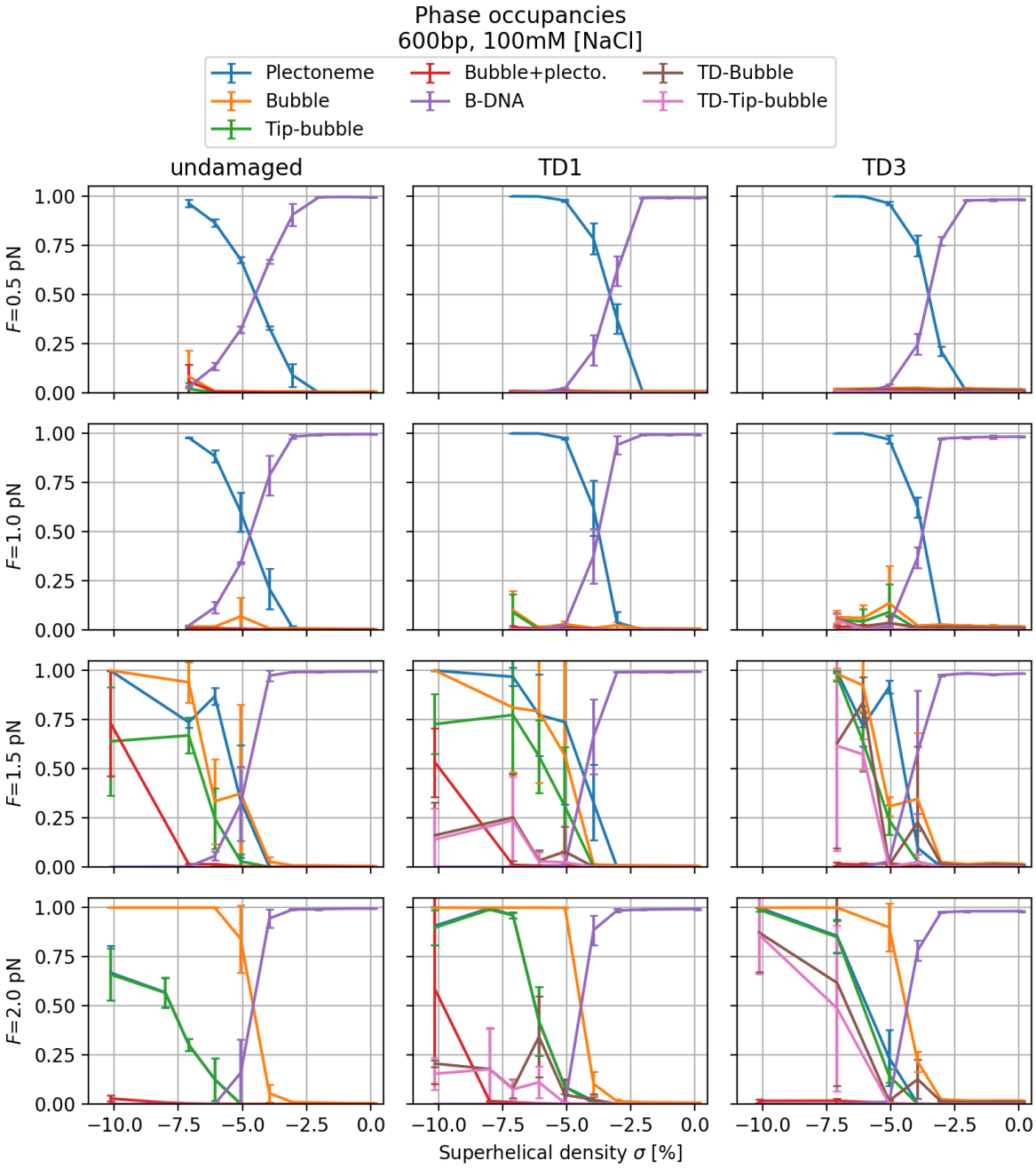

Figure 4.11: Phase occupancies in undamaged and TD-containing duplexes as a function of supercoiling density σ for different pulling forces F. At $F = 0.5$ pN, 1 pN, only the plectoneme and B-DNA phases are significantly observed. The presence of the TD slightly moves the buckling transition. At $F = 1.5$ pN and 2 pN, the tip-bubble and bubble phases are also observed. These are very likely to be located on the TD.

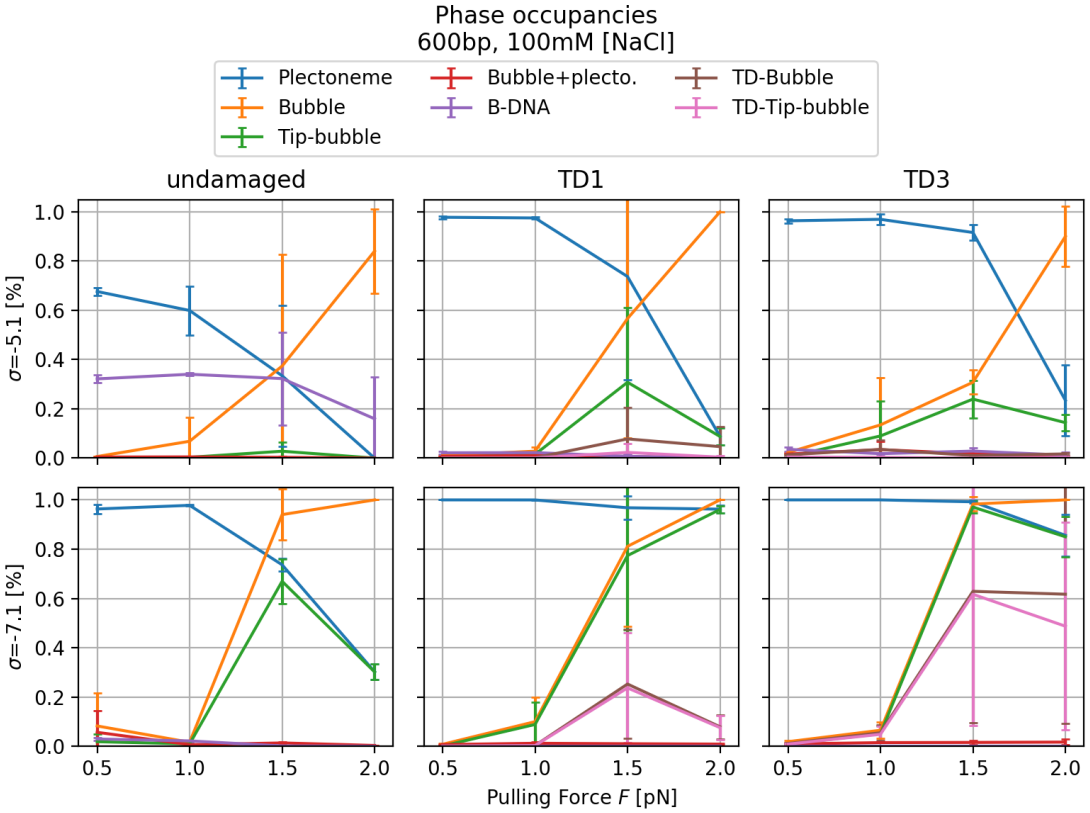

Figure 4.12: Phase occupancy of different phases in an undamaged duplex, a TD1-containing duplex, and a TD3-containing duplex, as a function of the pulling force F for different supercoiling densities σ.

TD-bubble : both the TD-containing base-pairs are denatured.

TD-tip-bubble : a tip-bubble plectoneme is detected within a distance M_c of a denatured TD.

Phase occupancies for each phase and for each duplex are shown in fig. 4.11 as a function of σ for different F. At $F = 0.5$ pN, 1.0 pN the B-DNA and the plectoneme phase dominate all the duplexes. At $F = 1.5$ pN and 2 pN, the bubble and tip-bubble phases are also present, and plectonemes without tip-bubbles are exceedingly rare. As $|\sigma|$ increases, tip-bubbles get more and more frequent. The conditional probability of having a plectoneme located on the TD given that a plectoneme is present is significantly increased, and described by the plectoneme pinning coefficient Q discussed in section 4.4.2. Similarly, in Fig 4.11, we show the phase occupancy as a function of force, for different supercoiling densities.

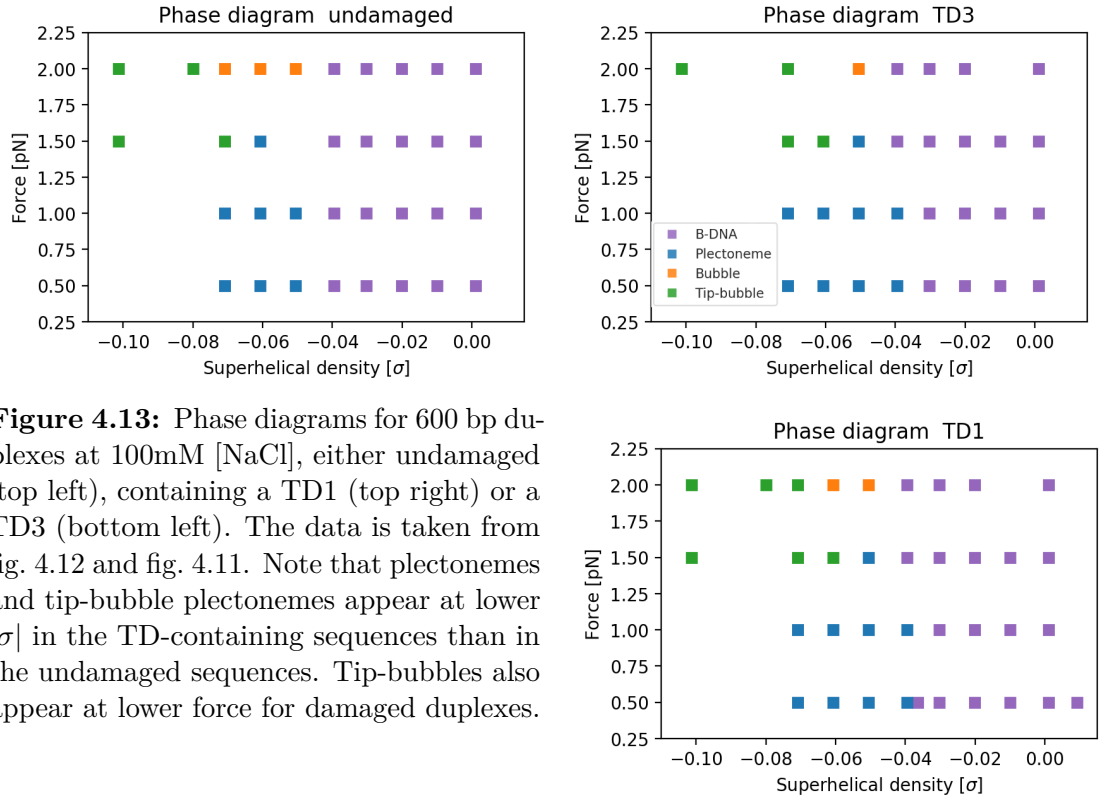

Figure 4.13: Phase diagrams for 600 bp duplexes at 100mM [NaCl], either undamaged (top left), containing a TD1 (top right) or a TD3 (bottom left). The data is taken from fig. 4.12 and fig. 4.11. Note that plectonemes and tip-bubble plectonemes appear at lower $|\sigma|$ in the TD-containing sequences than in the undamaged sequences. Tip-bubbles also appear at lower force for damaged duplexes.

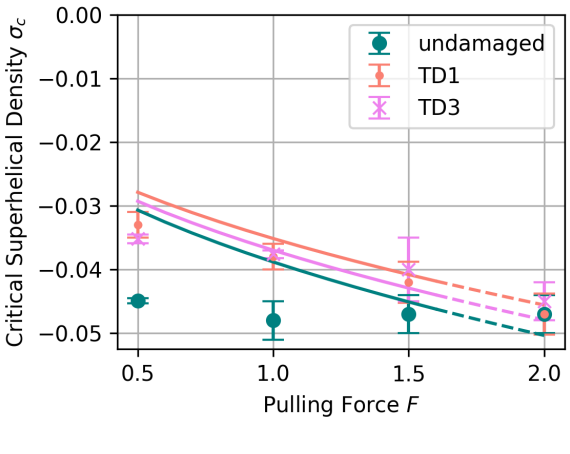

Figure 4.14: Critical superhelical density σ_c as for the TD-containing and undamaged 600 bp duplexes, at 150 mM [NaCl], measured from oxDNA simulations (symbols) and predicted by combining eqn. 3.53 and eqn. 4.4 (lines). The theory can't be compared to the simulated σ_c when denaturation bubbles appear, roughly at $F \geq 1.5$ pN (hence the dashed lines, see text), but is reasonably accurate for the TD-containing duplexes before that point.

Interestingly, there is also a range of biologically relevant F and σ where the plectonemes are only stabilised by the TD, and not present in the undamaged DNA, as seen in section 3.6. For example, no plectonemes are observed in undamaged DNA at $\sigma = -0.05$, while plectonemes are present respectively 9% and 60% of the time in the TD1 and TD3 duplexes at the same conditions. This trend can also be seen in fig. 4.12. The data in figs. 4.11 and 4.12 is used to construct the phase diagrams in fig. 4.13.

The data in fig. 4.11 can also be used to determine the critical supercoiling density σ_c, at which the occupancy of the B-DNA phase drops below 50%, shown in fig. 4.14. For F =0.5 pN, 1 pN, a well-defined buckling transition is observed, with the dominant phase shifting from B-DNA to plectonemes as $|\sigma|$ increases. The theory described in section 3.6, combined with the theory for the energy of a TD introduced in this chapter, using the values of torsional stiffness $C = 94$ k_BT nm, $B = 45$ k_BT nm, and $d = 10.55$ bp/turn reported for oxDNA2 (Snodin et al. 2015), provides a reasonable estimate of the buckling transition for both the TD containing duplexes and the undamaged DNA. However, this estimate drops faster with increasing force than observed in our simulations. The same theory, developed for positively supercoiled systems, cannot be applied to F =2 pN: negatively supercoiled systems at these forces relax the excess linking number by opening bubbles, which are not treated in the theory. Despite this, the theory works reasonably well, so we choose to show its predictions even for such high forces. In order to highlight this, theoretical prediction of the buckling transitions where there are bubbles are shown with a dashed line in fig. 4.14.

4.5 Biological significance of TD in supercoils

Previous sections described the biology of TD-repair (section 4.1), the modelling of TDs in oxDNA (section 4.3), and the impact of TD on duplexes in simulated MT experiments on a 600 bp duplex (section 4.4). Briefly, the presence of a TD in a duplex in a simulated MT experiment causes a significant degree of plectoneme pinning, summarised by the plectoneme pinning coefficient Q, causes a shift of the coexistence lines in the phase diagram (including a region where plectonemes are energetically stable only when located on the TD), and dramatically increases the opening probability of the TD when compared to a system with $\sigma = 0$. This section discusses how the simulation results presented in the latter section could apply to biology.

4.5.1 Active/passive behaviour of photolyase base-flipping

For the well known repair mechanism with photolyase, the TD-containing bases needs to flip out of the DNA duplex (Mees et al. 2004), as shown in fig. 4.15.

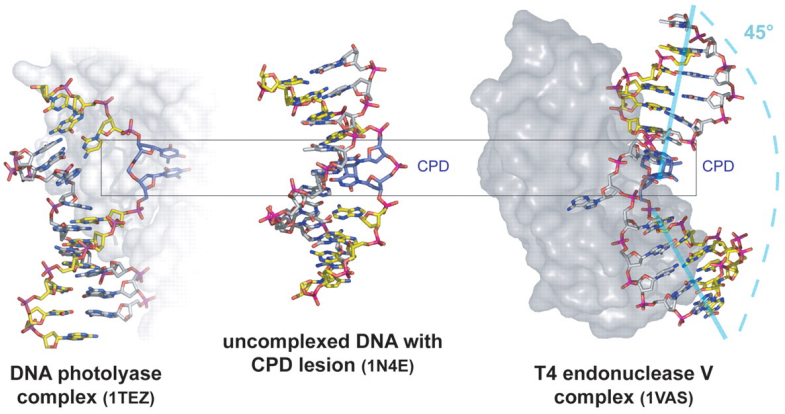

Figure 4.15: Crystal structures of a TD duplex in complex with a photolyase (left), a T4 complex (right), or unbound. The TD assumes an intra-helical conformation in the unbound duplex and an extra-helical conformation in both complexes. Adapted from (Mees et al. 2004).

The photolyase enzyme binds very specifically to TD dimers: the dissociation constant of 30 nM is 75000 lower than the dissociation constant for unspecific DNA binding (Husian and Sancar 1987). An open question is whether this is a passive phenomenon, e.g. the photolyase enzyme has to wait for a base flipping event to occur before binding, or whether it is active (induced fit), that is the protein first binds and then actively flips the base out of its usual intra-helical position. Reactive probes, NMR, and crystal structure experiments (Rumora 2008; McAteer et al. 1998; Park, Zhang, et al. 2002) and simulations (Knips and Zacharias 2015; Knips and Zacharias 2017) show an intra-helical state of the TD in short duplexes. Therefore, the general consensus in the literature is that the TD is preferably found in the intra-helical states, and that enzymes binds the TD-containing duplex and then exposes the TD, with the TD facilitating this process (Sancar 2003). Recent studies with isothermal titration calorimetry (Wilson et al. 2011; Schelvis et al. 2015) have however shown that the photolyase does not bind if the TD is not in the extra-helical state, signifying that the question is still far from answered.

However, none of these studies have been performed on a TD within a long (of the order of 1kbp or more) duplex under force and supercoiling, which is arguably a more biologically relevant context: DNA in both bacteria and eukaryotes is kept at a superhelical density of $\sigma \approx -0.05 \div -0.07$, while relatively strong forces are routinely applied by different molecules, such as RNA polymerases (20 pN), DNA

polymerases (12 pN) and microtubules (47 pN) (Milstein and Meiners 2011). The results of this chapter show that the denaturation probabilities can vary significantly with σ, increasing by up to about a factor of one hundred from $\sigma = 0$ to $\sigma = -0.06$ or above. This shows that both experiments and simulations on unsupercoiled DNA are missing an important feature that could have a profound effect on the TD flipping probability, its binding with photolyase in a biological context, and more.

4.5.2 The DNA-protein search problem

A well known open problem is the mechanism that allows DNA repair enzymes to find DNA damage. Locating such a damage via either 1D diffusion along the duplex or 3D diffusion in the cytosol would be too time consuming for the small number of DNA repair enzymes present at any given time inside the cell[4]: if either diffusion were the only mechanism involved, environmental and metabolic factors would damage the DNA faster than the cell could repair it, inevitably leading to cell death (Brackley, Cates, et al. 2013). One important part of this puzzle is the realisation that a mixture of 1D and 3D diffusion can speed up the diffusion significantly, as reviewed in (Mirny et al. 2009).

Much work has been done on the protein-DNA search problem (Mirny et al. 2009), but a great deal of it neglects the intricate structure of DNA in prokaryotes and eukaryotes. It would not be surprising if this structure had a profound effect on the search problem, even if the 1D + 3D mechanism continues to play a role. One exception to this picture is the work by (Brackley, Cates, et al. 2012; Brackley, Cates, et al. 2013), showing that the time τ_s to locate a specific sequence can be strongly affected by the supercoiled structure: in a string of rosettes, the location time of a sequence on a loop $\tau_{s,l}$ could be up to ten times shorter than the location time of a sequence in a rosette centre $\tau_{s,c}$, depending on the non-specific binding energy ϵ, as seen in fig. 4.16. This occurs at a comparatively low affinity $\epsilon \simeq 2$, due to the fact that the centre is harder to find for a protein, since the surrounding loops are in the way.

As shown in the past sections, TDs do preferentially locate on loops. Therefore, we propose that negative supercoiling aids TD repair not only by increasing the

[4]specifically, only about ten photolyases are present at any given time within *E. coli* (Sancar 2016).

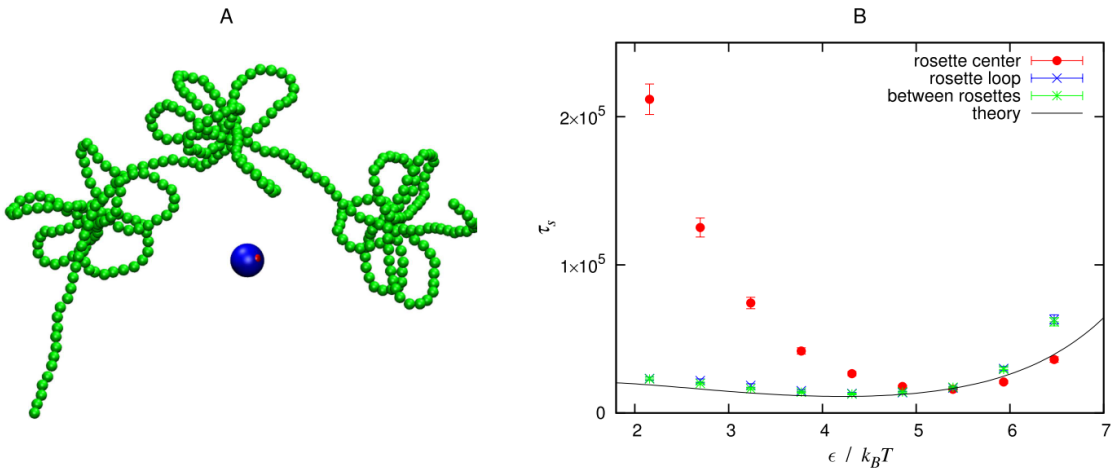

Figure 4.16: A: Simulation snapshot of the DNA rosettes (green) and protein (blue, with red active site) from (Brackley, Cates, et al. 2012). B: Plot of the mean search time τ_s as a function of protein-DNA binding affinity ϵ, for sequences on different positions of the rosettes, in units of 0.64μs. τ_s for sequences with low values of ϵ can vary by a factor of ten depending if the sequence is located in the centre of a rosette or elsewhere.

denaturation probability of the TD, thereby facilitating the flipping of the TD into the photolyase, but also by favouring the location of TDs in the relatively easy-to-find loops, away from the crowded and hard-to-sample environment at the centre of the chromosome, thereby lowering τ_s. This could explain how the few photolyases present at any given time in a bacterial cell can be sufficient to maintain a whole nucleoid. We must point out that, for the specific rosettes of (Brackley, Taylor, et al. 2013), $\tau_{s,l} < \tau_{s,c}$ only for $\epsilon < 5k_BT$: given that the non-specific binding energy of *E. Coli* photolyase is[5] 8 k_BT, in such a system moving, a TD to a loop would not lower τ_s. However, the model of (Brackley, Cates, et al. 2012) is quite crude, and so it is likely that a more sophisticated calculation could substantially change these boundaries. It's hard to tell how the details of the supercoiled structure influences the threshold at which becomes $\tau_{s,l} < \tau_{s,c}$, so that it wouldn't be surprising if a system more closely resemble a crowded biological supercoiled structure had a higher ϵ. More research is necessary to tell whether the proposed mechanism could actually apply to TDs in a biological setting, potentially including other protein-DNA search systems, such as mechanisms to repair other DNA damages,

[5]The non-specific binding energy in k_BT can be obtained by taking the natural logarithm of the binding constant 3 470 M^{-}1, reported in (Husian and Sancar 1987).

or the Lac repressor, which has been shown to find its target site on DNA much faster than what diffusion in solution would allow (Riggs et al. 1970).

4.6 The effect of TD on plasmid replication

It is known that plasmids are negatively supercoiled in most organisms. This facilitates strand separation, therefore easing replication and transcription. It is interesting to consider what happens when a TD is added to a supercoiled plasmid: will the presence of the TD strongly affect replication? (Yao et al. 2001) have indeed observed that introducing a (6-4)-photoproduct (and, to a lesser extent, a TD) can substantially hinder plasmid replication efficiency depending on the damage location within the plasmid. Biologically-relevant quantities such as the replication efficiency are affected by the denaturation profile of plasmids (Fye and Benham 1999), which might in turn be affected by the presence of the TD. This was indeed proposed almost forty years ago (Benham 1979).

To investigate this question, we extended the Fye-Benham model (described in section 2.5) to model TDs: the base-pairing enthalpy of two consecutive AT base-pairs can be changed to model the impact on the base-pair's thermodynamic stability, generalising the base-pairing free energies eqn. 2.17 to

$$b_j = \Delta H_j \left(\delta - \frac{T}{T_{m_j}} \right) . \tag{4.6}$$

For a regular base-pair we have $\delta = 1$, and the analysis presented in section 4.3 shows that setting $\delta = 0.4$ for a TD is consistent with experimental data. However, we also consider $\delta = 0.6$ and 0.7, in order to test the robustness of the results.

It would be interesting to use the extended Fye-Benham model to study the impact of the denaturation profile on plasmid pZ189 (Seidman et al. 1985), since this is the plasmid investigated by (Yao et al. 2001). However, the sequence of the plasmid is not available in the literature, and we couldn't obtain it from the authors of both studies above. Instead, we consider the well studied plasmid pSP189 (Parris and Seidman 1992), derived from pZ189.

We used the Fye-Benham model to analyse the denaturation profile of pSP189 (see fig. 4.17), identifying the two base-pairs #4011 and #3016 as the local maxima

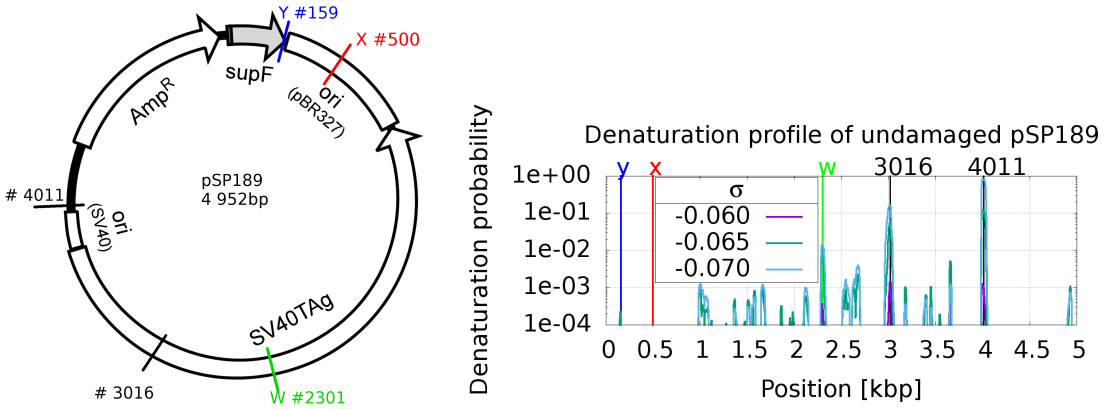

Figure 4.17: Left: sketch of plasmid pSP189, with highlighted the approximate positions of the most prominent denaturation sites in the undamaged plasmid (#3016 and #4011). TDs were added to one of the base-pairs #500 (X, in red), #159 (Y, in blue), or #2301 (W, in green). Adapted from (Jain et al. 2013). Right: denaturation profile of the undamaged plasmid, for biological solvent conditions at 300 K and supercoiling. Base-pairs #4011 and #3016 are the local maxima of denaturation probability at every shown σ.

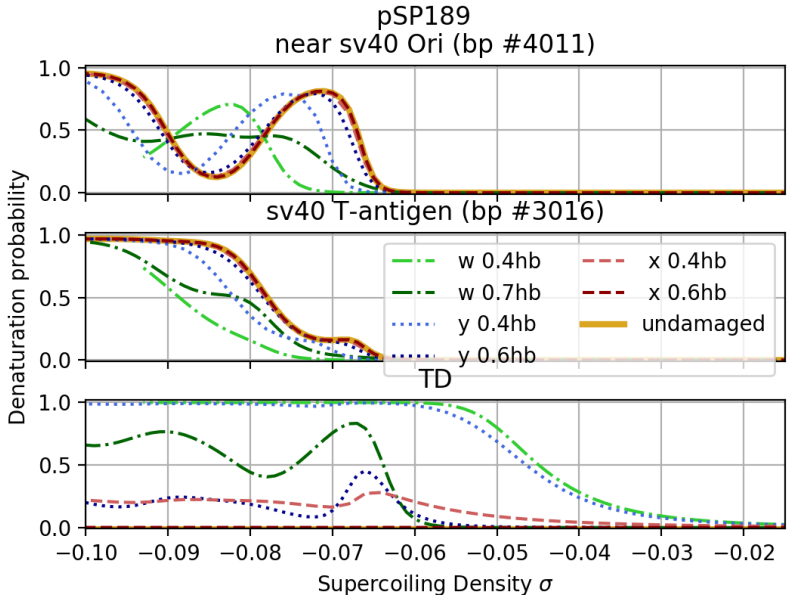

Figure 4.18: Denaturation probability of the most prominent denaturation sites of the plasmid pSP189 (top and centre) and of TD-containing base-pairs (bottom), in the modified Fye-Benham model. The top two plots compare the denaturation probability for an undamaged plasmid (gold, continuous) to what happens to the denaturation profile at this same location when a TD is introduced in the following base-pairs: #500 (x, red dashed), #159 (y, blue dotted), and #2301 (w, green dot-dashed). Models of TD where the hydrogen bonding enthalpy has been reduced to 40% (light) and 60% or 70% (dark) have been considered. The bottom plot shows the denaturation profile of the TD-containing base-pairs. Denaturation at a TD lowers the amount of twist available for denaturation at bp #4100 or #3016, and so makes these biologically important positions less likely to open.

of the two sites with the highest denaturation probabilities at biologically-relevant σ and solvent conditions ([NaCl] = 150mM, T = 310K). We then analysed the denaturation probability at these base-pairs as a function of σ, both in an undamaged plasmid and in a plasmid containing a TD at either base-pair #500 (referred to as site x), #159 (site y), or #2301 (site w). Sites x and y have been arbitrarily chosen among the ones sufficiently far from the outer two denaturation sites and with low denaturation probability even at biological supercoiling: their denaturation probabilities are at most 3×10^{-7} and 2×10^{-4} respectively. Site w is instead the third most-prominent denaturation site.

The denaturation probabilities of the base-pairs discussed are shown in fig. 4.18. Adding a TD on site x doesn't appreciably change the denaturation probabilities of the denaturation sites for any of the TD models considered. On the other hand, adding a TD on site y causes a significant shift of roughly $\Delta\sigma = 0.005$ to the denaturation probabilities of both bp #4011 and bp #3016, assuming that our estimate that a TD reduces the hydrogen bonding enthalpy by 40% is correct, and a smaller shift if the correct value were to be 60%. Finally, a TD on site w causes a strong disruption of the denaturation profile of the plasmid regardless even with a hydrogen-bonding reduced to 70%. The TD also massively increases the denaturation probability of the host base-pairs by up to six orders of magnitudes, although the exact amount of probability increase is dependent on the TD model considered. The difference between plasmids containing the TD at different sites is due to the different denaturation probability of the sequences in which the TD is embedded, which in turn is affected by the local AT and GC content.

The location and base-pairing energy of a TD-containing base-pair are crucial in determining whether the TD will barely affect the denaturation profile or radically change it. However, even if the base-pairing enthalpy of the TD-containing base-pair were to be 70% of a standard AT base-pair (a very conservative estimate, which would contradict the thermodynamic data presented in section 4.2), a TD on site w would still cause a major alteration of the denaturation profile.

(Yao et al. 2001) did observe that the location of the photoproduct (either a (6-4)-photoproduct or a TD) had a substantial impact on the plasmid replication

efficiency, though they ascribed it to whether the photoproduct was on the lagging strand or on the leading strand: our results suggest that the different location of the TD could also play a role. More experiments, as well as further theoretical analyses on the pZ189 sequence, are needed to elucidate the impact that TDs can have on biological processes in plasmids in particular and in other forms of DNA in general. However, the general principle that a TD can drastically alter the probability that other, possibly biologically important, sites can open is likely to be a general effect. One can imagine other more complex scenarios as well.

4.7 Conclusions

The impact of supercoiling density and pulling force on the supercoiled structure of long TD-containing DNA duplex has so far been overlooked in the literature. Here we show what is, to the best of our knowledge, the first study of such a system. Our simulations show that supercoiling density and pulling force can alter dramatically the conformation of the TD, making the extra-helical state as much as 10 000 times more likely than it would be in relaxed DNA. The exact value of the duplex bend θ_0 generated by the TD is not known, as different experiments show different value. However, our main results are qualitatively similar for two different values of θ_0, and the plectoneme pinning is unaffected by the detailed form of the microscopic potential that generates the bending (since TD1 and TD3 have approximately the same Q): this shows that our results are somewhat robust, and the system is worthy of more careful investigation by both theory and experiments. The first obvious step would be performing real MT experiments to validate the results of the simulated ones. The presence of tip-bubbles and their location could possibly be ascertained by using using selective ssDNA-binding enzymes, such as Replication Protein A (RPA). Useful insight might also be gained by investigating the structure of plasmids with or without TDs, both with experiments and simulation. Notice that this mechanism can work for any sequence that has a bending angle (stiffness) significantly greater (lower) than that of surrounding DNA: in this chapter we have applied it to the TD as a simple model of DNA damage, but other possible examples include (6-4)-photoproducts (for which the bending angle is $\theta_0 = 44°$), the

Dewar isomer ($\theta_0 = 21°$), naturally bent sequences such as the ones of the *Crithidia Fasciculata* (Shpigelman et al. 1993), and easily-denatured sequences such as the replication origins of plasmids (Fye and Benham 1999).

The protocol described in section 4.3 to generate θ_0 in an oxDNA duplex can be easily generalised to model other bent sequences: fig. 4.5 already shows how to generate any $\theta_0 \leq 40°$, but more base-pair-steps can be altered to further increase θ_0.

The simulation protocol used in section 4.4.2 can also be generalised to sample the plectoneme position distribution. Similarly, the enhancement of finding damage by an enzyme through the mechanism of the damage localising on supercoiling tips, presented in section 4.5.2, can also be generalised to other kinds of inhomogeneities. The mechanism is quite simple and neglects many important features of the DNA. Future work could include performing simulations along the lines of those of (Brackley, Cates, et al. 2012) on more realistic systems, as well as experiments, for example imaging fluorescent labelled photolysases as they bind to the TDs in a bacterical chromosome.

The results on the denaturation profile of plasmids described in section 4.6 are also valid for any DNA modification that significantly lowers the hydrogen-bonding enthalpy, not only thymine dimers.

The structural and thermodynamic impact of anomalous sequences is largely unknown, despite its numerous possible biological implications, and the results presented in this chapter merely scratch its surface. In a sense we've raised more questions that we have answered, but many of these questions may be very promising to try to answer.

Where it is shown what happens when DNA is pulled very hard and untwisted very strongly, which surprisingly doesn't involve DNA snapping "Hey you fella could you please really NOT!?". DNA is much more tolerant than I'd be under those conditions.

5 L-DNA

Contents

As discussed in the previous chapter, DNA absorbs excess linking number by forming supercoiled structures (writhe) and by overwinding or unwinding the double helix (twist). Depending on solvent conditions, DNA sequence, superhelical density σ and pulling force F (if present), runs of contiguous open base-pairs, commonly called denaturation bubbles, will form in the DNA duplex to store part of the twist. In brief, opening a base-pair to create a bubble requires energy, but the bubble locally reduces the energy cost required to twist the helix. The energy cost of denaturation is, to a good approximation, constant in F and σ, while the cost to extend a plectoneme or to twist the DNA increases with F and σ. Therefore, there is a range of forces and supercoiling densities where such bubbles are energetically stable. In magnetic tweezers (MT) experiments, this has been measured to happen at forces on the

order of F=1 pN, depending on salt conditions. Experiments report values ranging from $F = 0.8$ pN (Salerno et al. 2012; Tempestini et al. 2013) at a monovalent salt concentration of 150 mM, to $F > 1$ pN for a salt concentration of 100 mM (Kriegel, Ermann, Forbes, et al. 2017), to $F = 1.5$ pN (Matek et al. 2015; Matek 2014) for an effective monovalent salt concentration of 500 mM in oxDNA1 simulations.

As the duplex is unwound and σ is decreased, more and more of the duplex transitions to the bubble state, until the bubble is so extended that its structural properties (such as bending B and twisting C moduli) strongly affect duplex-wide properties such as torque and end-to-end distance. This highly denatured phase has been dubbed L-DNA (Bustamante et al. 2003), where the L stands for left-handed[1] (see also the phase diagram in 1.4).

The properties of DNA in MT experiments with $\sigma \sim -1$ have been the object of several experimental MT investigations (see (Kriegel, Ermann, and Lipfert 2017) for a recent review). Pioneering experiments by (Allemand et al. 1998) studied the behaviour of a 17 kbp duplex at $-5 < \sigma < +3$, in a 10 mM phosphate buffer, with $F < 50$ pN, providing the first experimental observation of the coexistence of B-DNA and denatured phases in MT experiments. For a sense of scale, note that at $\sigma = -1$ the two strands forming the duplex are so untwisted that they are in a conformation that is topologically equivalent to them being parallel to each other, and at $\sigma = -2$ they have a left-handed linking number equal in magnitude to the usual right-handed linking number.

(Sheinin, Forth, et al. 2011) have studied three duplexes of length 1.6kb, 1.8kb and 2.2 kbp respectively, at 150 mM [NaCl] and 23 $\pm$ 1°C, for F as high as 36 pN, and σ as high as -2.2. This study shows that L-DNA has a helical repeat of -13 bp/turn, a contour length of 0.48 nm/bp. The behaviour of the DNA at high force, where no significant supercoiled structures were expected to form, was also observed in order to establish the properties of L-DNA under high force both via a fit to the worm-like-chain model and with a phenomenological model of DNA in a mixed phase. This fitting approach yielded a bending persistence length of 3nm and

[1] This should not be confused with the left-handed enantiomer of DNA, commonly known as mirror-DNA, but occasionally also called L-DNA.

a twisting persistence length of 10 to 20nm, much lower for the respective values for B-DNA of the order of 50 nm and 100 nm respectively.

Further experiments, carried out by (Vlijm et al. 2015), have investigated the properties of an unwound 20.666 kbp duplex up to $F = 4.5$ pN and $\sigma = -2.5$, in a buffer of 30 mM NaCl, 60 mM Tris-HCl pH 7.5 and 0.1 mM Ethylenediaminetetraacetic acid (EDTA). The supercoiled structures formed at high values of σ were characterised with fluorescence spectroscopy, and the sharply decreasing duplex extension at $\sigma \sim -2$ was interpreted as the point where L-DNA plectonemes appear. Bubbles were shown to be uniformly distributed along the duplex by adding fluorescently-labelled replication protein A (RPA), a protein that selectively binds single-stranded DNA. The presence of L-DNA was also studied by adding the protein Zα to the solution. Zα binds to the left-handed sugar-phosphate backbone in Z-DNA, so that a change in the DNA extension at the time of adding Zα could elucidate the presence of a left-handed helix structure in the DNA. Z-DNA wasn't substantially detected, although it has been shown to form in sequences that have a high content of GC base-pairs (Zhabinskaya and Benham 2011).

None of the studies above could directly determine the supercoiled structure of DNA. One advantage of oxDNA simulations is that they are performed at nucleotide-level resolution, and allow direct observation of the supercoiled structure of the molecule. In this chapter, we use oxDNA to investigate this regime. In particular, we use the average-sequence parametrisation of oxDNA2 to simulate a 600 bp duplex at 23°C and 100 mM [NaCl], for σ between -2.8 and -0.4 and F between 0.5 pN and 32 pN. oxDNA was not parametrised to simulate DNA structures beyond the duplex phase and the usual single-stranded phase, therefore studying the highly unwound L-DNA state will be a challenge for oxDNA. Nevertheless, some of the main physics of bonding and flexibility of single and double stranded states, which oxDNA does well on, should also be the main determinant of the L-DNA phase, so it will be interesting to see how well oxDNA compares to experiments. Briefly, measurements of the duplex extension and torque show semi-quantitative and quantitative agreement with experiments, which is encouraging. Perhaps even more interesting, is our prediction that in the L-DNA phase, the form of the writhe is not the well-known

plectonemic state, but instead a solenoidal state. These oxDNA predictions are supported by theoretical calculations done by Romain Rollin, using a worm-like chain (WLC) model with parameters extracted from experiments, which suggest a fluctuation stabilised solenoidal phase for L-DNA.

The chapter is structured as follows: first, section 5.1 summarises the simulation protocol used in the oxDNA simulations. Then, measurements of duplex extension and torque are discussed in section 5.2 and section 5.3 respectively. The solenoidal phase is presented in section 5.4. Section 5.5 analyses the B/L-DNA transition and compiles a detailed collection of simulation snapshots. Finally, the conclusions are drawn in section 5.6.

5.1 Simulation protocol

The simulation protocol used in this chapter is almost identical to that described in section 2.4. A 624 bp duplex was generated in a straight configuration, where the 12 base-pairs at each end were used as handles. One of the handles was fixed both in position and in orientation by harmonic traps, simulating the fixed surface on which the duplex is tethered in MT experiments. The other handle was fixed in orientation, but its position was free to move along a straight line, simulating the magnetic bead. The optical trap had a stiffness of 10 simulation units, equivalent to 571 pN/nm. A pulling force of varying intensity was also applied to the freely moving handle. First, unwinding simulations were run with the molecular dynamics backend and the Langevin thermostat for every pulling force F by continuously unwinding the duplex by 1 turn per 10^6 simulation steps. The duplex was therefore turned ten times faster than the duplexes in other studies, because of the large linking number difference needed to study the L-DNA state. Using the time mapping from section 2.2 this is roughly equivalent to one turn per 15 ns, such that the duplex is unwound to e.g. $\sigma = -2$ over 1.7 μs. Then, simulations snapshots were extracted at the appropriate time to have the desired superhelical density σ. Finally, these final states were used as initial configurations for simulations at fixed σ. These were then run for a time of roughly 10^9 simulation steps, requiring up to five days each on a GPU machine, equivalent to 15 μs. Each system was simulated in three

independent replicas, and the quantities discussed in the following sections are estimated from the mean of the values measured in each replica. The error bars show the sample standard deviation among the three replicas. For F=0.5 pN, 2 pN and $\sigma > -0.4$, the same data of chapter 4 is used.

5.2 Duplex extension

The duplex extension as measured in oxDNA is compared with the results of (Vlijm et al. 2015) and (Sheinin, Forth, et al. 2011) in fig. 5.1. The extension curves measured in simulations have the following features:

- At $F = 0.5$ pN the extension rapidly decreases as the duplex forms a plectoneme, as in the standard scenario for low forces where no bubbles are expected to form.
- As F increases, the slope of the extension becomes less and less pronounced, as more and more bubbles form, and even initially changes direction for $F \geq 16$ pN.
- When σ becomes sufficiently negative, the slope of the extension changes. For F=32 pN, the slope becomes negative around σ=-1.6. For F=16 pN, the slope changes twice, at $\sigma = -2.0$ and $\sigma = -2.4$. The slope changes at $\sigma = -2.0$ at F=6.5 pN, 8 pN, and changes at both $\sigma = -2.2$ and $\sigma = -2.4$ for F=4 pN. The fact that the slope changes twice for F=4 pN might simply rise from sampling errors.
- The extension for F=2 pN is slightly concave, similarly to the data for F=1 pN and 2 pN in (Vlijm et al. 2015).

These features are similar to the experiments, except that the forces in the oxDNA are typically larger for the same extension behaviour. In particular, some discrepancies are:

- The force where the extension presents a plateau is respectively reported to be between 3.5 pN and 6 pN by (Sheinin, Forth, et al. 2011) and about 4.5

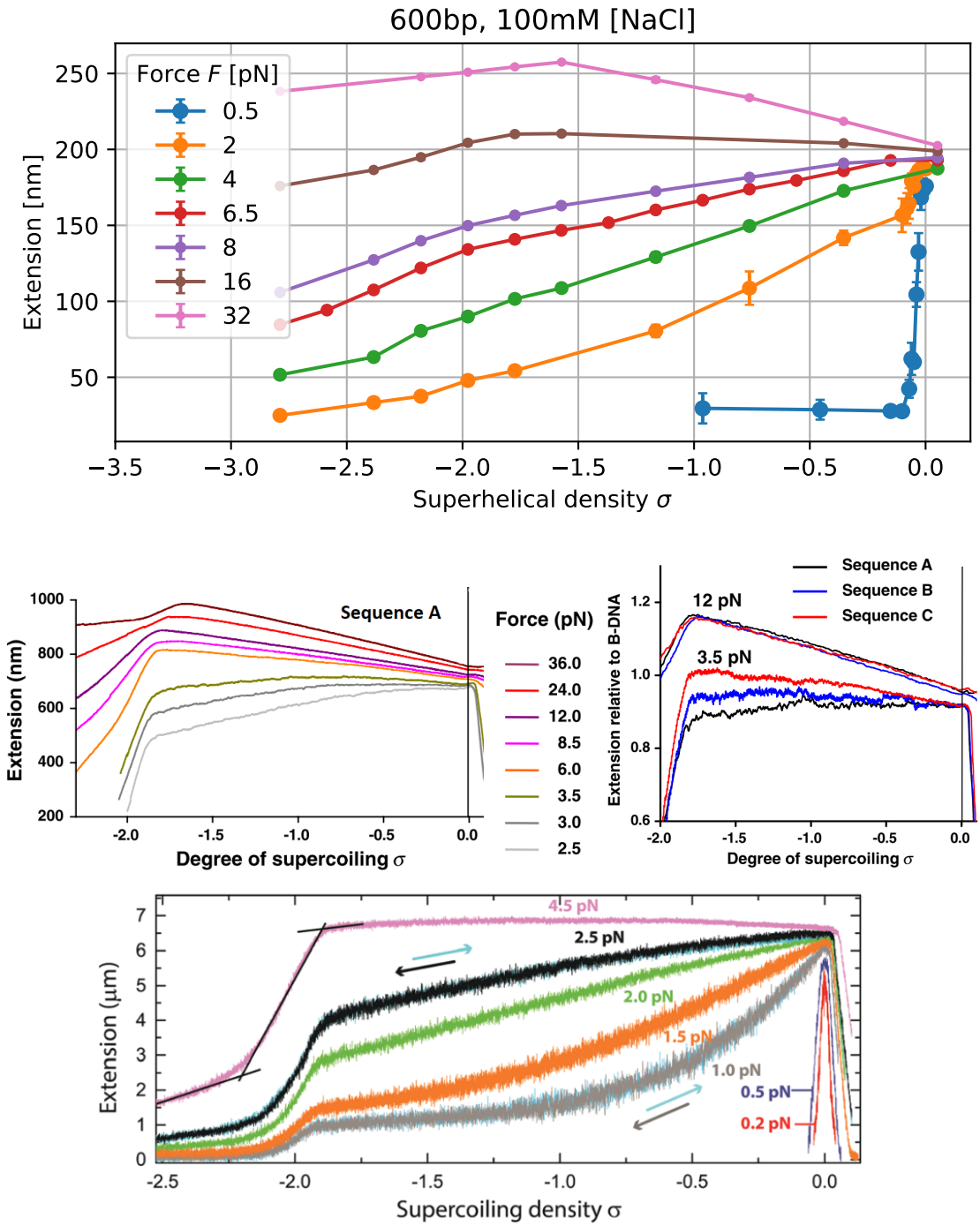

Figure 5.1: Duplex extension. Top: oxDNA simulation on a 600 bp duplex for an effective salt concentration of 100 mM. Centre: experiments at 150 mM NaCl, on sequences of length 2.2kb (A), 1.8kb (B), 1.6kb (C) respectively and varying GC content (adapted from (Sheinin, Forth, et al. 2011)). Bottom: experiments on a 20 666 bp duplex in a buffer of 30 mM NaCl, 60 mM Tris-HCl pH 7.5 and 0.1 mM Ethylenediamine- tetraacetic acid (EDTA) (reproduced from (Vlijm et al. 2015)).

pN by (Vlijm et al. 2015). Conversely, we observe a plateau for F between 8 pN and 16 pN, a value between a factor of 3 and a factor of 4 greater than the one observed in experiments.

- The change in slope observed at $\sigma \sim -2$ is much less pronounced in simulations than in experiments, both in relative and absolute terms.

The discrepancy could be partly explained by the presence of sequence effects, neglected in the average-base parametrisation of oxDNA used in these simulations. (Sheinin, Forth, et al. 2011) report significant variation in duplex extension depending on sequence at F=3.5 pN (see fig. 5.1); this shows that sequence can affect the plateau force, i.e. the force at which the extension becomes constant in σ for $\sigma \gtrsim -2$. Another origin of the discrepancy could be the difference contour length of the duplexes considered: while duplex treated in experiments are at least 1.6 kbp and 20.666 kbp respectively for (Sheinin, Forth, et al. 2011) and (Vlijm et al. 2015), our duplexes are 0.6 kbp long. This definitely has an effect when both F=2 pN and $\sigma < -2.2$, where the duplex collapses to a state where the extension can't decrease any further because of the entropic confinement of the duplex (as can be determined by visual inspection of simulation snapshots, see section 5.5), an effect that would probably appear at lower σ in a longer duplex. We are also working at a slightly lower salt concentration, which should destabilise the duplex.

However, even if the same sequences and experimental conditions were realised in simulations, there might well still be discrepancies due to limitations in the oxDNA model. OxDNA was not parametrised to study DNA phases different from the B-DNA state and the usual single-stranded phase, so using it to study heavily supercoiled, denatured structures means pushing its limits. For example, simulations of highly stretched configurations in oxDNA showed qualitative agreement with experiment, but significant quantitative differences (Romano et al. 2013).

5.3 Torque

Measurements of the torque acting on the moving end of the duplex, measured as described in section 2.4, are shown as a function of σ and compared to the

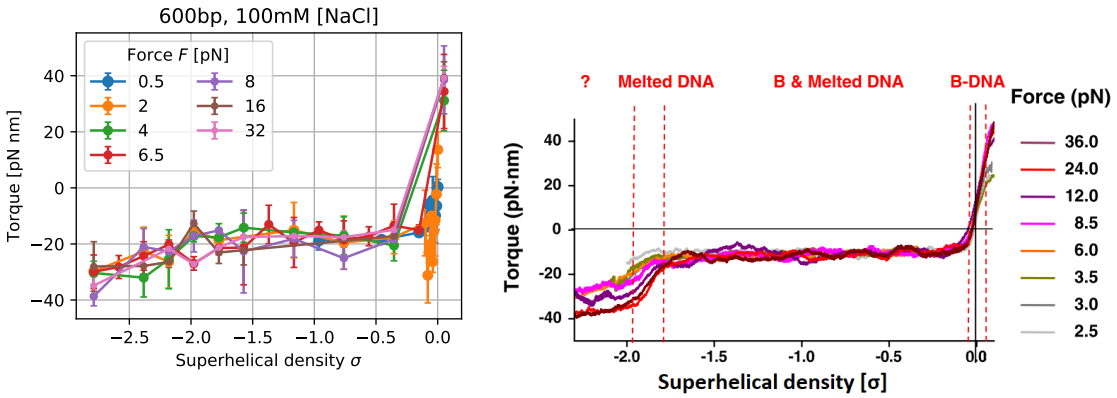

Figure 5.2: Torque on the moving end of the duplex, measured here with oxDNA (left) and in previous experiments (left, adapted from (Sheinin, Forth, et al. 2011)). Both plots show a plateau in the mixed B-DNA/L-DNA phase and a decreasing behaviour as the duplex is unwound for $\sigma \lesssim -1.8$.

experimental results of (Sheinin, Forth, et al. 2011) in fig. 5.2. Both the curves produced by oxDNA and MT experiments show a plateau for $\sigma \gtrsim -2$, and a decreasing trend as σ decreases beyond -2. The torque at the plateau can be estimated at 18.7 ± 0.5 pNnm by constant fitting all the data in the range, similar to the data by (Sheinin, Forth, et al. 2011) that report about 10 pNnm. The presence of the plateau can be linked to the linear slope in the plot of the denaturation fraction in fig. 5.6: as the torque required to open a base-pair is constant, the decrease in σ is proportional to the number of base-pairs in the L-DNA state.

5.4 Solenoids and plectonemes in L-DNA

The literature on DNA supercoiled structures has largely focused on plectonemes. However, in classical elastic theory, besides the plectoneme, it is also possible for an elastic rod to buckle to a solenoidal state, characterised by a helical shape of pitch increasing with the number of turns added to a system. The state has been considered early in the development of the theory of MT experiments, but it has since been largely disregarded, mostly because it was shown that plectonemes should prevail on solenoids for an elastic polymer with the bending and torsional moduli of B-DNA (Marko and Siggia 1994; Marko and Neukirch 2012), B=45nmk_BT and C=100nmk_BT respectively (with the Boltzmann constant k_B and temperature T).

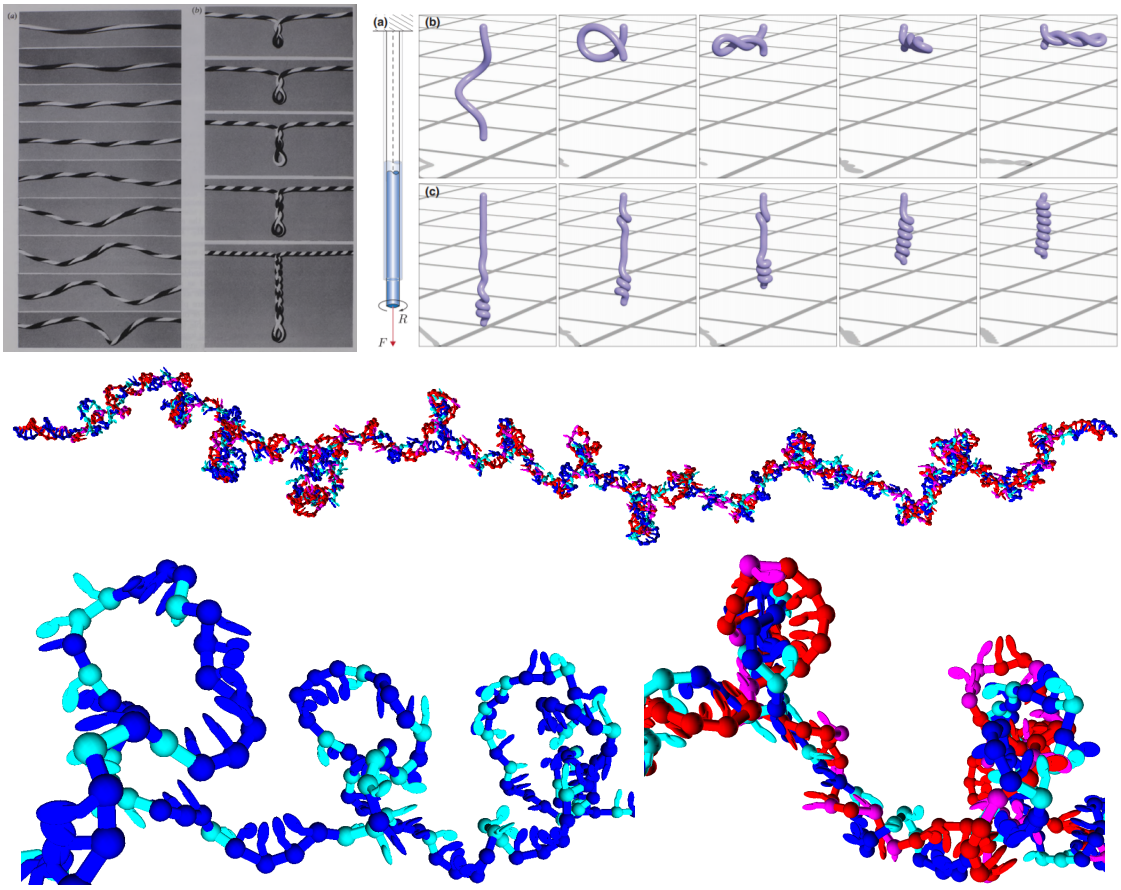

Figure 5.3: Competition between solenoids and plectonemes in macroscopic systems, from (Thompson and Champneys 1996) (top left) and (Gazzola et al. 2018) (top right). A typical simulation snapshot for F=8 pN and σ = -2.5 also shows a solenoidal shape (centre). The solenoid is highlighted in a detail, showing both strands (bottom left) or only one of them (bottom right). Light colours show unstacked bases, while stacked bases are darker. Stacked bases naturally form a curve that preferentially follows the curvature of the loops.

A recent exception is the work by (Dittmore and Neuman 2018), which proposes the presence of a possibly metastable solenoidal phase in B-DNA as well.

However, we do observe solenoids somewhat ubiquitously in the simulations described in this chapter (see fig. 5.3), as shown more in detail in the next section. For L-DNA, (Sheinin, Forth, et al. 2011) use a phenomenological model to estimate B=7 k_BT nm and C=19 k_BT nm, much lower than those of B-DNA, and it is therefore interesting to revisit the question of which phase is the most stable, the plectonemic one or the solenoidal one.

Romain Rollin, a master student in our group, used the worm-like chain (WLC) model in the mean-field limit to find that for this set of parameters, the solenoidal

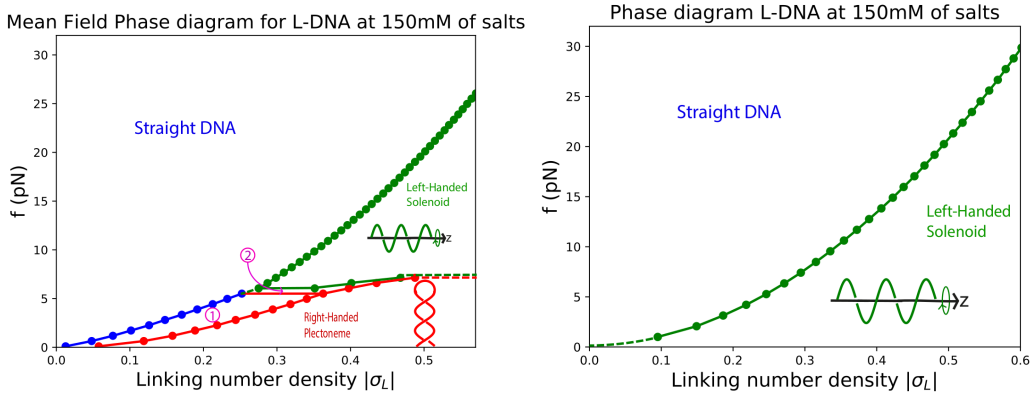

Figure 5.4: Phase diagram for L-DNA produced by Romain Rollin. When Gaussian fluctuations are accounted for (left figure), only solenoids are observed. At the lower force end, fluctuations become too large to treat, and the approximations used break down.

phase is favoured at high forces. Fluctuations can be treated within a Gaussian framework, in a similar way to what has been done for plectonemes in the much celebrated Moroz-Nelson theory (Moroz and Nelson 1997). Interestingly, it was found that these fluctuations destabilise the plectonemic phase more than the solenoidal phase, so that effectively the solenoids are stabilised by the presence of fluctuations. The phase diagrams he generated, both in the mean field limit and with Gaussian fluctuations, are shown in fig. 5.4. The large impact of the fluctuations is also observed in the simulation snapshots, which presents solenoids that are very dynamical and irregular. The approximations used do not hold in the limit of low F, or when both B-DNA and L-DNA are present, so they cannot be applied to the case where $F \leq 0.5$ pN or to $\sigma > -2$.

Such a picture, where L-DNA solenoids prevail over L-DNA plectonemes at intermediate to higher forces, could be compatible with the results of (Vlijm et al. 2015). Their images of fluorescently labelled, supercoiled DNA (see fig. 5.5) show the bright and highly dynamical spots identified with plectonemes for F=0.2 pN and, to a lesser extent, for F=1.5 pN. At F=2 pN, they report the structures observed to become rather stationary and condensed, suggesting that plectonemes have disappeared. On the other hand, solenoids like the ones discussed in this section would presumably look somewhat stationary in fluorescence microscopy, lacking the spike in longitudinal plectoneme density caused by a plectoneme. (Vlijm et al. 2015) observe that plectonemes are formed for $\sigma < -2$, but, as they don't

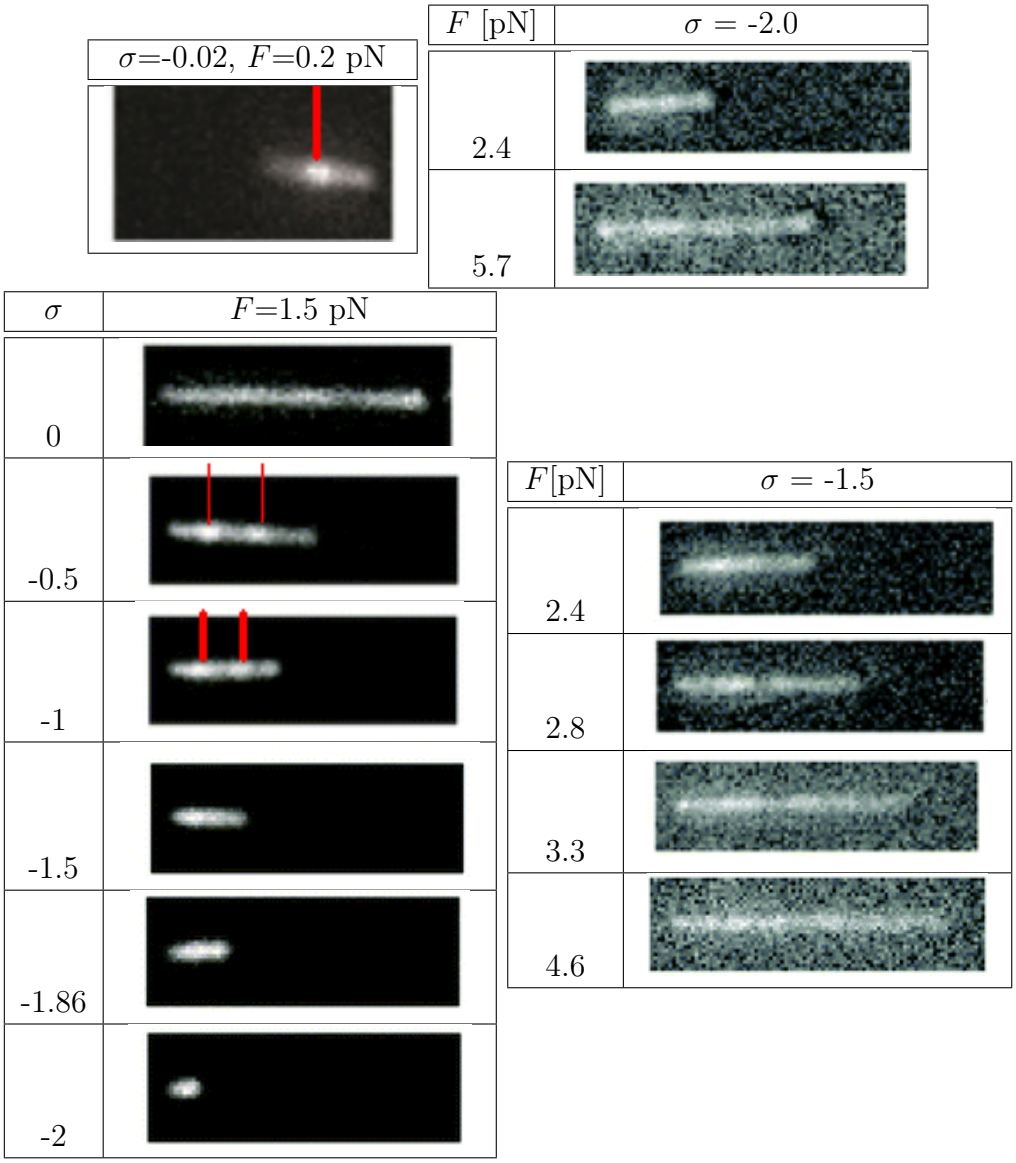

Figure 5.5: Background-corrected fluorescence images of unwound DNA, adapted from (Vlijm et al. 2015). For $F = 0.2$ pN, $\sigma = -0.02$, a plectoneme is observed as a bright blob (top left). At F=1.5 pN, similar bright blobs are observed for $\sigma = -0.5, -1$, which suggests the presence of plectonemes or possibly intermediate structures between plectonemes and solenoids. For $F > 2.4$ pN and $\sigma = -2$ (top right) and $\sigma = -1.5$ (bottom left), no bright spots are observed, which suggests the presence of supercoiled structures with a roughly constant linear density of nucleotides, such as solenoids.

report images, this statement seems to be based solely on the duplex extension. However both solenoids and plectonemes cause a drop in extension with increasing twist, and the difference between them is not that large, so that it is hard to distinguish them on this basis.

In the oxDNA simulations presented here, plectonemes were only observed in systems at either $F = 0.5$ pN or at $F = 2$ pN and $\sigma \leq -0.1$, where the B-DNA structure is intact, with solenoids dominating for $F \geq 2$ and $\sigma \leq -0.4$. However, our simulations cannot exclude the presence of plectonemes for $F \leq 4$ pN in longer duplex. If we assume the following statements:

- As a rule of thumb, one can expect finite-size effects to influence the probability of supercoiled structures with a length-scale comparable to or greater than the duplex contour length.

- The minimum free energy plectoneme state (that is, the plectoneme state with the lowest free energy of all the plectoneme states) and the minimum free energy solenoid (that is, the solenoid state with the lowest free energy of all the solenoid state states) have comparable characteristic length scales.

then we should admit the possibility that plectonemes are dampened because of finite-size effects when the solenoids have length scales comparable to the duplex contour length.

For $F \geq 8$ pN, solenoids seen in simulations (if any) have length-scales (i.e. solenoidal radius and pitch) much smaller than the duplex contour length, so we can safely expect plectonemes not to appear at these forces even in longer duplexes. However, at F=2 pN and F=4 pN, the solenoids have radius and pitch comparable to the duplex contour length. For such low F, plectonemes slightly bigger than the solenoids might not fit in our system, and therefore would not be observed due to finite size effects. According to this argument, we cannot rule out that longer oxDNA duplexes could potentially still show plectonemes even in the bubble phase for $F \leq 4$ pN. The question could be clarified by simulating longer duplexes. However, our attempts to simulate a 2kbp at these conditions have observed that the dynamic of supercoiled structure became so slow that simulations were impractical

given the scope of this study. Future simulations, perhaps performed on more advanced computer hardware, involving a simulation algorithm more elaborate than standard molecular dynamics, or performed with the current hardware and simulation algorithm but lasting several weeks or months per run, would be able to simulate such a long duplex and answer these questions.

To summarise, several hints point to the presence of a solenoidal phase in real DNA, as well as oxDNA: this is perhaps the most interesting outcome of this work. The simulations presented here reliably show that solenoids are prevalent for $F \geq 8$ pN in oxDNA. As oxDNA was not parametrised to faithfully reproduce the melting transition in supercoiled DNA under force, some care must be taken in applying insight to real-life DNA. The presence of solenoids is however compatible with the results of (Vlijm et al. 2015), for $F > 1.5$ pN. Furthermore, calculations by Romain Rollin for the WLC model predict that a polymer with the bending and torsional moduli measured by (Sheinin, Forth, et al. 2011) will prefer solenoids to plectonemes under such values of σ when not in the limit of low force; these rely on the WLC being a good model of L-DNA, which is still to be established, though it seems a reasonable assumption. The fact that oxDNA shows semi-quantitative agreement with previous measures of duplex extension and good agreement with previous measures of torque, as discussed in sections 5.2 and 5.3, further strengthens the argument. Future work, including detailed analysis and modelling of the experimental extension curves, is needed to establish whether L-DNA is actually present in MT experiments.

5.5 Characterisation of strongly negatively twisted DNA structures

One of the advantages of simulation methods such as oxDNA is that it provides access to detailed DNA structures. The supercoiled structure of oxDNA under extreme negative σ is strongly dependent on both σ and F. This section provides a thorough description of how this structure changes at different conditions and reports simulation snapshots, as well as showing the fraction of L-DNA as a function of σ. Solenoids are preferred to plectonemes at $F > 2$ pN and $\sigma \leq -0.4$, despite virtually all previous work in the literature assuming that the supercoiled structures

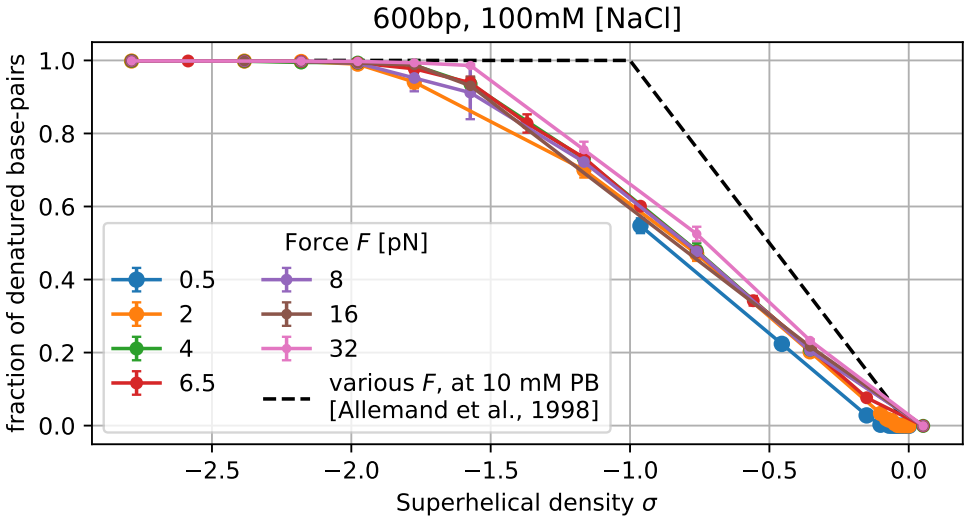

Figure 5.6: Fraction of denatured base-pairs as observed here (coloured, continuous lines) and in (Allemand et al. 1998) (black, dashed line) as a function of superhelical density σ. While in our simulations and in work by (Sheinin, Forth, et al. 2011) and (Vlijm et al. 2015) the torsional melting transition is complete at $\sigma \sim -1.8$, Allemand *et. al.* report $\sigma \sim -1$ (see main text).

present in L-DNA to be plectonemes, rather than solenoids. The solenoids are particularly evident for F=8 pN (where the force is high enough to significantly suppress the destabilising thermal fluctuations, but still low enough to generate solenoids with a significant radius of curvature) and $\sigma \leq -2.5$, where the large excess linking number causes the solenoids to be more pronounced.

5.5.1 Denaturation fraction

The fraction of denatured base-pairs, measured according to the criterion discussed in section 1.6 (Šulc et al. 2012), is plotted versus σ in fig. 5.6 and also compared to the experimental measurements of Allemand et al 1998 (none of the other experiments report this quantity). As σ becomes more negative, more base-pairs open to relax the excess turns, until all of the DNA is denatured. As expected, this doesn't appreciably depend on F: the torque is constant throughout this regime according to both experiments (discussed in the beginning of the chapter) and simulations (see section 5.3). A notable exception is the F=32 pN system, where the duplex fully transitions to L-DNA at $\sigma = -1.6$, while systems at other forces are in a B-DNA/L-DNA mixed phase until $\sigma = -1.8$. This can also be seen in the extension

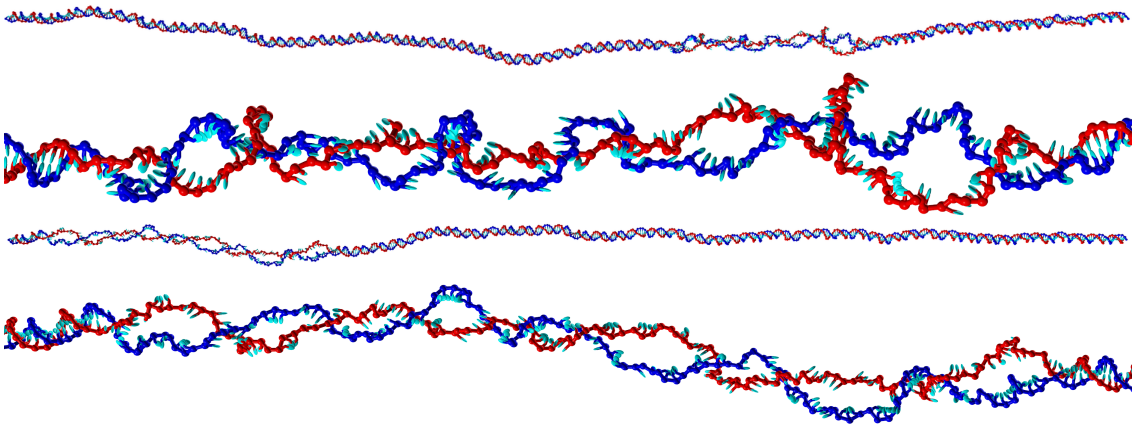

Figure 5.7: Simulation snapshots at $\sigma = -0.4$, F=16 pN (top two) and F=32 pN (bottom two), with close-ups of the denaturation bubbles. The bubble at F=16 pN is significantly less elongated than the one at F=32 pN, despite the bubbles being formed of roughly the same number of base-pairs at both forces (99 bp and 113 bp respectively).

plot (presented in section 5.2). These values of σ are compatible with what reported by (Sheinin, Forth, et al. 2011) and (Vlijm et al. 2015), but are in disagreement with those of (Allemand et al. 1998) (also shown in fig. 5.6). The discrepancy was previously (Sheinin, Forth, et al. 2011) ascribed to the unphysiologically low salt concentration of the buffer used by (Allemand et al. 1998), which would prevent the bubbles from twisting back in a left-handed helix.

5.5.2 Mixed phase

Despite the simple, substantially force-independent, piece-wise linear plot in fig. 5.6, the three-dimensional structure of DNA is very complex, so much that it's not obvious how to define an order parameter to describe it. Given the richness and unexpectedness of these structures, we feel that it is worth it to describe them in detail, and we proceed to do so in the remainder of the section.

At $F = 16$ pN, 32 pN, the bubbles in the mixed phase are twisted in a left-handed configuration and straightened by the high force. The denatured DNA at 32 pN does appear significantly elongated with respect to the 16 pN one (see fig. 5.7), despite the two systems having roughly the same number of denatured base-pairs.

At F=8 pN (see fig. 5.8), the bubbles assume a much larger variety of configurations. Elongated bubbles alternate with shorter bubbles, and sometimes one of the strands will assume a straight, mostly unstacked configuration, while the

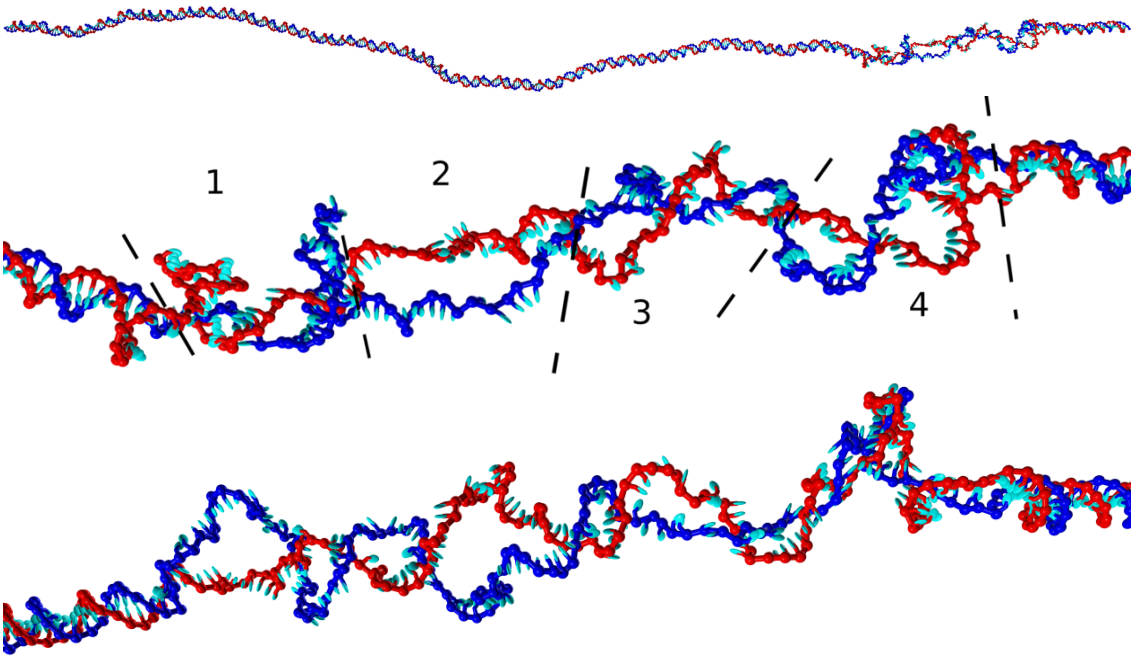

Figure 5.8: Simulation snapshot at $F = 8$ pN, $\sigma = -0.4$ (top), with close-up on the denaturation bubble (centre). Quasi-symmetric structures are separated by dashed lines. Notice in particular: 1) a section with alternated loops 2) an elongated bubble 3) a twisted three-turn bubble 4) another section with alternated loops. A bubble from another simulations snapshot of the same system is also shown (bottom) to highlight the highly dynamical nature of the bubbles.

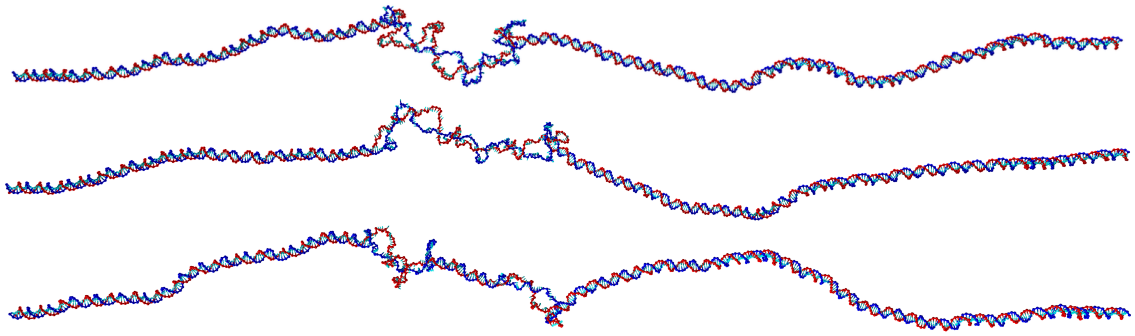

Figure 5.9: Simulations snapshots at $F = 4$ pN, $\sigma = -0.4$, showing one (top and centre) or two (bottom) denaturation bubbles.

other forms curved, partially stacked loops around it. This introduces significant slippage between the strands; because the strands are bound to return to the B-DNA phase at the same time, in any given snapshot they must take turns forming a loop and passing through it. This creates structures that, although chaotic and highly dynamic, can show some symmetry when the bubble is reasonably small (such as the ones observed at $\sigma = -0.4$, or even in higher σ when more than one bubble is present). Conversely, when the bubble is long enough this symmetry disappears, as the entropy

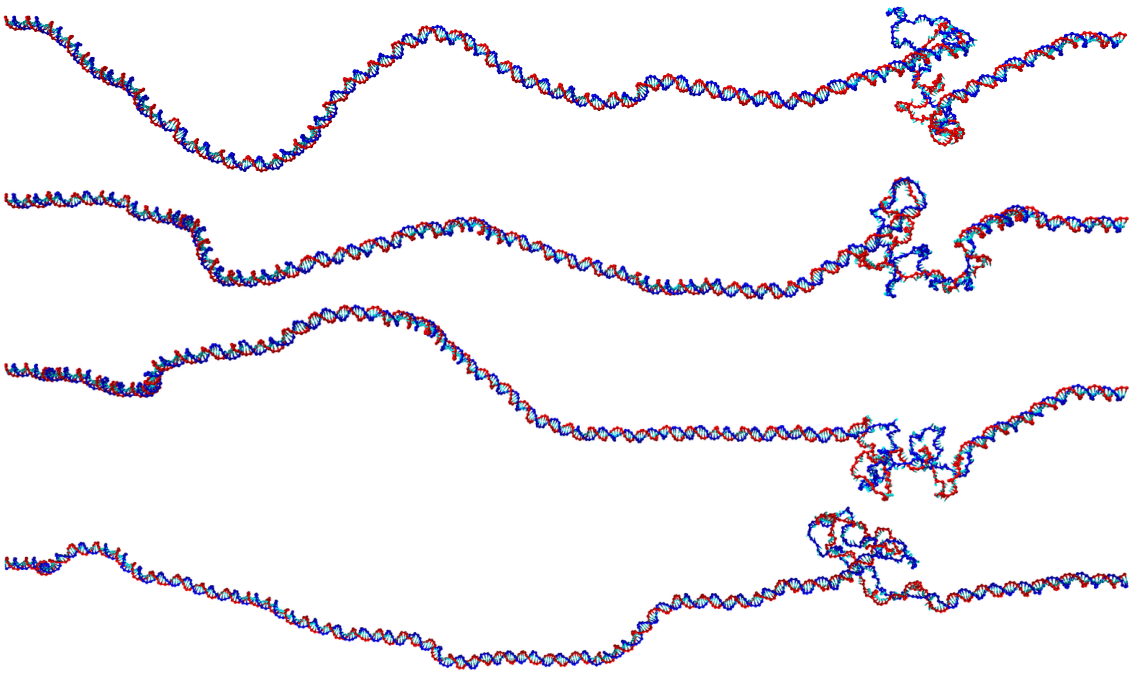

Figure 5.10: Simulation snapshot at $F = 2$ pN, $\sigma = -0.4$. Notice the highly dynamical, chaotically structured denaturation bubble.

of symmetric configurations is necessarily lower than that of asymmetrical ones.

At F=4 pN (see fig. 5.9), multiple bubbles are observed for a significant part of the simulations on two of the three simulations replicas. However, as the simulations equilibrate we see that bubbles sometimes disappear, but never appear spontaneously, which suggests that such a multi-bubble phase is, at least for duplexes of this length, a transient effect induced by the unwinding. It makes sense that the denatured DNA be contained by a single, large bubble, as opposed to several smaller bubbles: each separate bubble has to pay a large initiation cost (as discussed in section 2.5), and longer bubbles have more configurational entropy than smaller ones.

At F=2 pN the bubble twists on itself creating looped structures. These are sometimes similar to plectoneme loops, however no plectoneme tail is ever observed. Finally, at F=0.5 pN a hybrid collapsed plectoneme/bubble state is observed at $\sigma = -0.5, 1$ (see fig. 5.11). Both snapshots show multiple plectoneme tails, but this might well be a metastable effect caused by the highly stressed nature of these systems and the high speed at which the linking number was changed.

As σ is decreased further, the features of the denaturation bubbles for $F > 2$ pN remain roughly the same, except for the increased portion of duplex in the

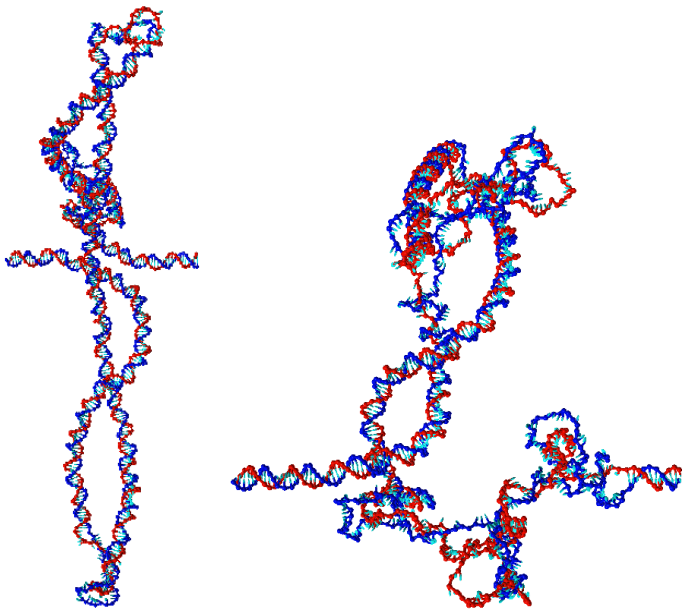

Figure 5.11: Simulation snapshots at F=0.5 pN, $\sigma = -0.5$ (left) and $\sigma = -1$ (right). Contrarily to all the other forces considered, a hybrid collapsed-plectoneme/bubble state is observed at these values of σ.

denatured state, as shown in fig. 5.12. The structures observed at F=2 pN instead appear somewhat solenoidal, similarly to what observed in the pure phase (as described in the next subsection).

5.5.3 Pure phase

At about $\sigma \sim -2$, the duplex is completely denatured. If σ decreases further, no excess twist can be stored by opening other base-pairs, so that significant writhing and overtwisting of the bubbles takes place.

At F=16 pN, 32 pN, thermal fluctuations have a reduced role and no significant supercoiled structures are observed even at $\sigma = -2.8$, the lowest σ considered in this study (see fig. 5.13). At F=16 pN, the strands alternate tightly twisted bubbles with a coaxial phase, where one of the strands is completely unstacked but straight, and another wraps around it. At F=32 pN, a simple tightly-wound bubble state seems to be preferred. There is similarity between the denatured state formed in this configuration and the P-DNA discussed in chapter 6. However, the two phases are different in at least two aspects:

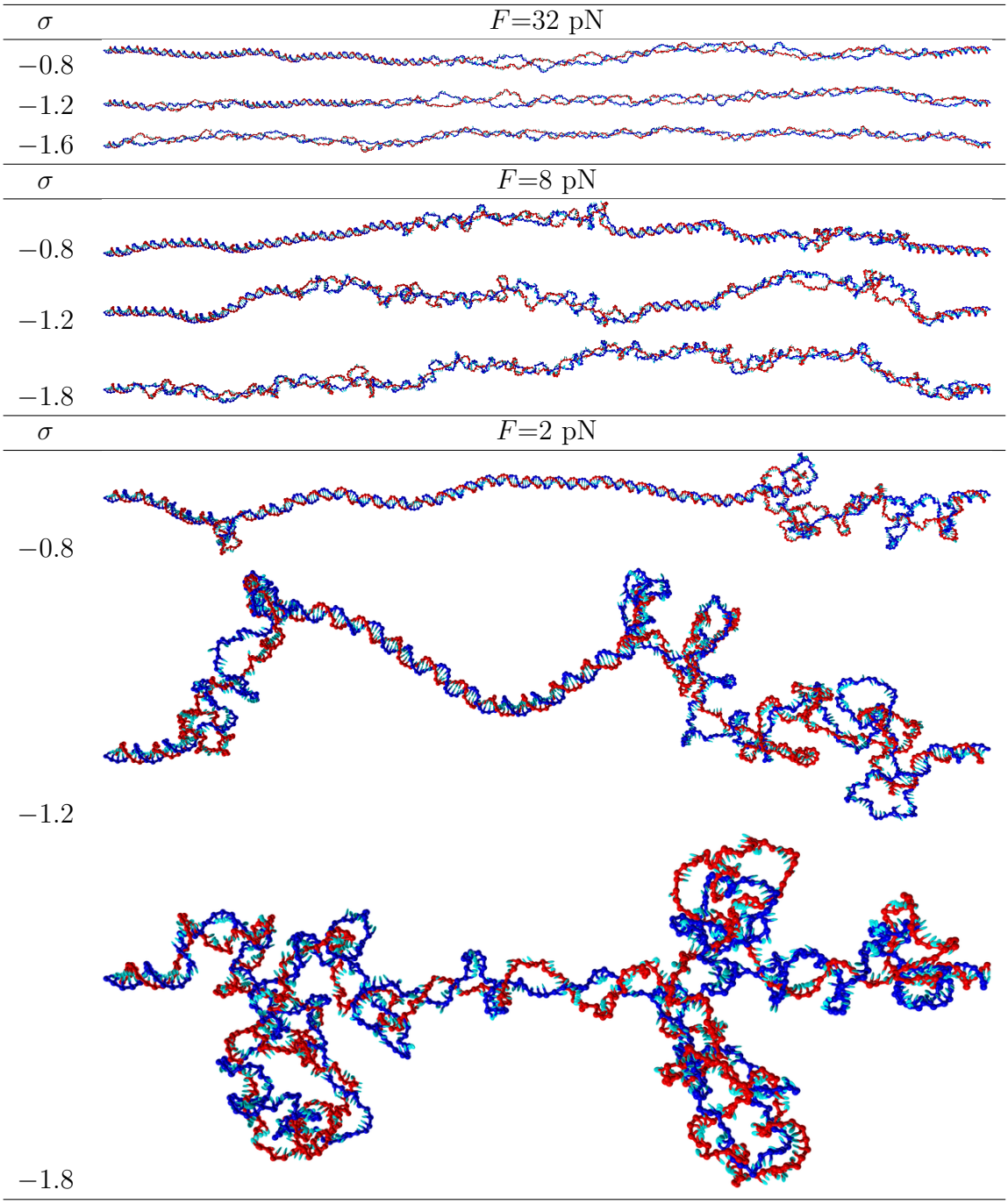

Figure 5.12: Simulation snapshots of the mixed phase and onset of the pure phase for several forces F and superhelical densities σ. As σ decreases, denaturation bubbles get longer and longer without fundamentally changing their structures. As F decreases, the supercoiled structures become more and more free to deviate from the line connecting the two handles, thereby changing the aspect ratio of the molecular structure.

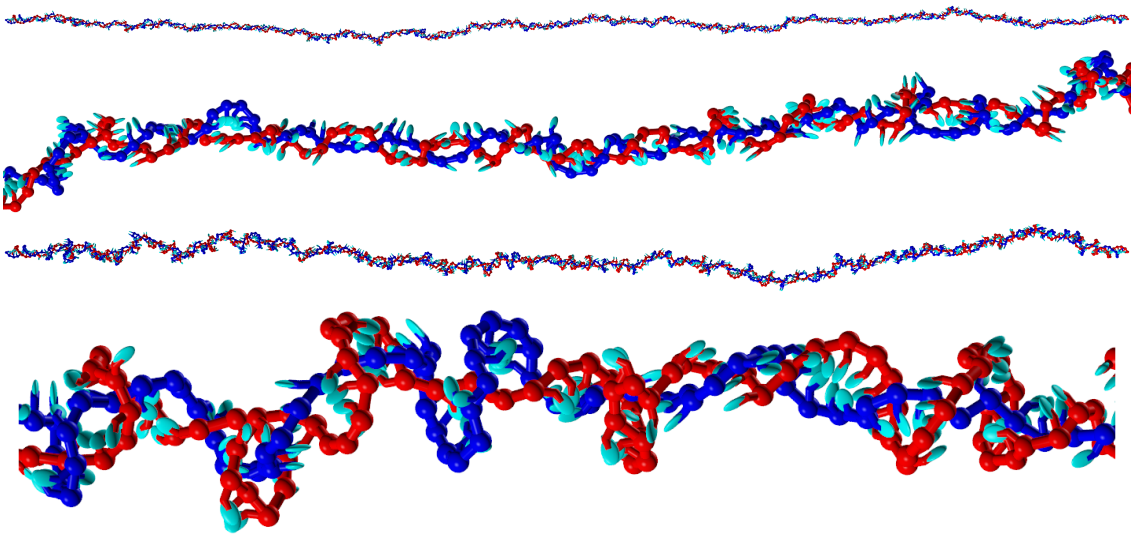

Figure 5.13: Simulation snapshots for $\sigma = -2.8$ at F=32 pN (top two) and F=16 pN (bottom two), for both the total system and detail of a denatured structure.

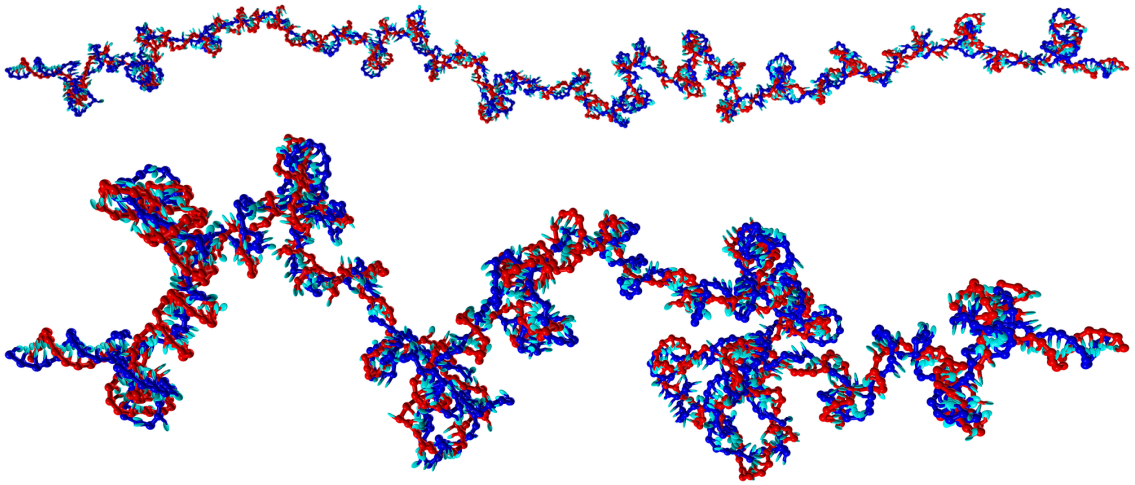

Figure 5.14: Simulation snapshots at $\sigma = -2.8$ for F=8 pN (top) and F=4 pN (bottom), showing typical solenoidal structures.

- They have different helicity: left handed for the L-DNA discussed here, and right-handed for the P-DNA of chapter 6.

- P-DNA is much more tightly wound than L-DNA.

The comparison is discused in more detail in Section 6.5.1.

At F=8 pN and lower (see fig. 5.14), the duplex forms solenoidal structures similar to the ones observed at F=2 pN.

5.6 Conclusions

This chapter investigates the response of an oxDNA duplex in simulated molecular tweezers (MT) experiments, for supercoiling density σ as negative as -2.8 and pulling force F between 0.5 pN and 32 pN. For sufficiently negative sigma, the duplex starts to relax some of the twist by denaturing base pairs. As σ becomes more negative, the fraction of denatured base-pairs increases, until the duplex is completely denatured at roughly $\sigma \sim -2$. The denaturation bubbles thus formed assume, for $F < 8$ pN, supercoiled structures in order to partially store the excess linking number as writhe. Surprisingly, these structures are not the almost ubiquitous, well studied plectonemes, but solenoids. These structures are shown and described in section 5.5.

OxDNA was not explicitly parametrised to correctly depict L-DNA. Given the current lack of experimental data on the nucleotide-level features of L-DNA, it's even unclear how one could extend oxDNA to make its predictions on L-DNA more reliable. Therefore, it's not straightforward to assume that just because solenoids are observed in oxDNA simulations of L-DNA then they must also be present in real L-DNA. However, a number of factors suggest that this is actually the case:

- The presence of supercoiled structures is qualitatively compatible with fluorescence spectroscopy by (Vlijm et al. 2015).

- Theoretical calculations in the framework of Gaussian fluctuations by R. Rollin (yet to be published) show that a polymer with the bending and torsional moduli measured for L-DNA by (Sheinin, Forth, et al. 2011) would, under these conditions, form solenoids, instead of plectonemes.

- The extension curves shows semi-quantitative agreement with experiment: the curves are roughly piecewise-linear, with two distinct slopes in the mixed B-DNA and L-DNA phase (i.e. for $\sigma \gtrsim -1.8$) and in the pure phase. In the mixed B-DNA and L-DNA phase, the slope is negative at low forces, but increases as F increases, changing from positive to negative. The slope in the mixed phase is always positive.

- The torque curves agree with the experiments of (Sheinin, Forth, et al. 2011): the torque is constant at about 20 pN in the mixed phase, regardless of F. It then decreases as σ decreased in the pure phase.

This work provides a qualitative to semi-quantitative agreement with experiment for properties such as the force-extension curves and the torque curves. In addition, it provides evidence for an unexpected solenoidal phase in L-DNA, instead of the more common plectoneme phase. This opens up several possible research lines, including:

- Establishing whether solenoids are actually present in L-DNA in real MT experiments.
- Extending theoretical models of DNA to consider the solenoidal phase, as well as the plectoneme phase.
- Systematically studying the impact of sequence and solvent effects in oxDNA.
- Exploring whether L-DNA would assume solenoidal configurations in a biological context, such as in the bacterial chromosome, and if so what could be the implications.

Images like the ones provided by (Vlijm et al. 2015) could possibly be analysed in more detail to ascertain the nature of the supercoiled structures in L-DNA. Another possibility could be the use of a DNA looping complex, such as the bidentate *lac* represser, to discriminate between plectonemes and solenoids fig. 5.15. If a *lac* repressor where to bind a plectoneme with a well defined tail, such as in the right panel of fig. 5.15, the interoperator distance would be of several hundred base-pairs. Conversely, if it were to bind two spires of a solenoid together (similarly to the left panel of fig. 5.15), the interoperator distance would be around 100 bp. Several technical challenges, such as using a looping complex that can bind denatured DNA and accounting for sequence specificity, might make this particular approach hard to execute. More in general, any approach able to detect the relative distances of a set of nucleotides could be useful in distinguishing solenoids from plectonemes.

Theoretical modelling of the solenoidal phase has partly being carried out by Romain Rollin; some of his findings include calculations of the free energy of the

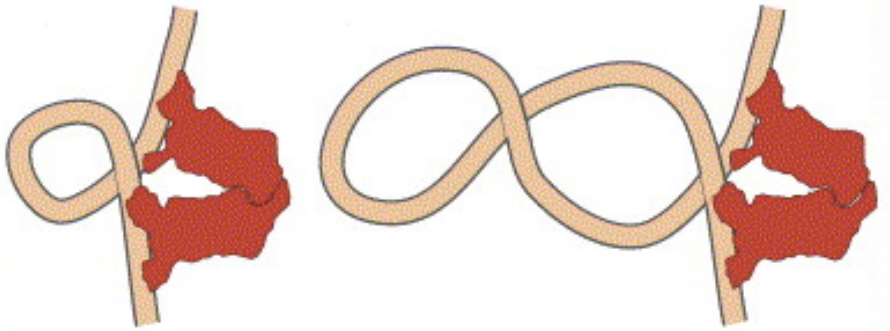

Figure 5.15: Sketches of a bidentate *lac* repressor (red) bound to a loop (left) and to a plectoneme with tail (right). Notice the different strand separation between the binding sites on DNA in the two supercoiled structures. Adapted from (Saiz and Vilar 2006).

solenoidal phase, both in the mean field limit and with Gaussian fluctuations. Work on the solenoidal phase in the absence of fluctuations has been modelled in the past mostly out of concerns for the buckling behaviour of cables under tension, such as in structural engineering of bridges (Thompson and Champneys 1996), but the presence of solenoids in DNA would provide a system to test theories about the stability of solenoids where strong thermal fluctuations are present.

These simulations are a first attempt at using oxDNA to study the behaviour of L-DNA in MT experiments, but they are by no means exhaustive. The full impact of duplex and solvent features has yet to be investigated. These include salt concentration, temperature, sequence, duplex length. Generating more precise measurements of torque and extension, as well as investigating other values of σ and F, might prove fruitful.

Finally, we don't find any obvious biological implications for the solenoidal behaviour of L-DNA, but these implications might indeed be present and their search deserves consideration, as both denaturation bubbles and picoNewton forces are routinely present in the cell. The solenoidal phase of L-DNA is unexpected and deserves further investigation; only future research will tell whether this work opens on an unexplored alley or on a dead end.

Underwound DNA has been the hero of our story... but like Elon has Kimbal, Cesare Borgia had Juan, and Corrado has Caterina, so too untwisted DNA has a less remembered, yet remarkable sibling. Overwound DNA is less prominent in biology, but when things get very hot it falls on its shoulders to make sure that life is not only possible, but triumphant.

6 Positive supercoiling

Contents

The properties of underwound DNA have been the focus of most of this thesis. Instead, this chapter looks at the properties of overwound DNA, also referred to as positively supercoiled DNA. From a biological perspective, negative supercoiling is more relevant than positive supercoiling, since most organisms (including mesophilic eukariotes and bacteria) have mostly negatively supercoiled DNA to facilitate strand separation during processes such as DNA replication and transcription. However, RNA polymerases (RNAp) introduce downstream positive supercoiling and upstream negative supercoiling during transcription (see fig. 6.1): the resulting supercoiling densities are so great that transcription can be hindered if topoisomerases don't remove them (Ma and Wang 2016). Recent studies have also suggested that positive supercoiling could also have a role in the regulation of chromosome

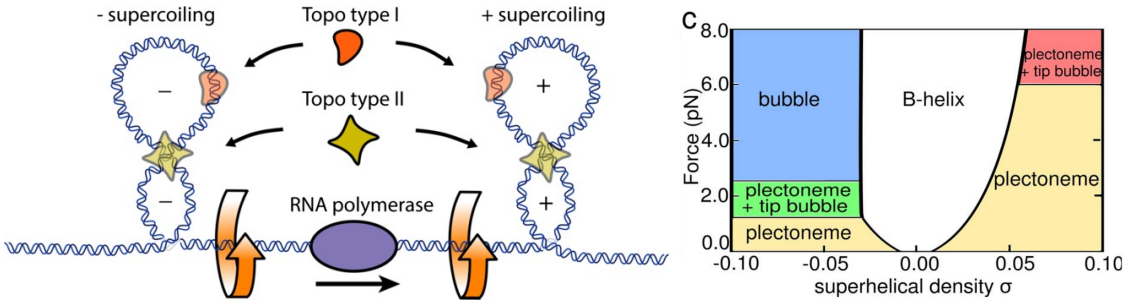

Figure 6.1: Left: sketch of the twin-supercoiled-domain model. RNA polymerase rotates around the DNA duplex during transcription. Because of the mass of the polymerase and associated machinery, the DNA is positively supercoiled upstream of the polymerase and negatively supercoiled downstream.((Ma and Wang 2016)). Right: phase diagram of supercoiled DNA from (Matek et al. 2015) for oxDNA1.

condensation, gene expression, mitosis, and more (Valenti et al. 2011). Finally, hyper-thermophiles have positive supercoiled DNA in order to cope with the high temperatures of their environment, which otherwise would make strand separation too probable (Galburt et al. 2014).

For low forces, the phenomenology of positively supercoiled DNA in molecular tweezers (MT) experiments is similar to that of negatively supercoiling DNA, as summarised in the phase diagram in fig. 6.1. Underwound DNA forms bubbles that compete with the plectonemes for forces greater than 1 pN or so at physiological salt. By contrast, overwound DNA continues to form plectonemes for significantly larger forces. However, for large enough $F > F_c$, positively supercoiled DNA also undergo a transition where bonds are broken in order to relieve twist. This denatured state has been dubbed P-DNA (Allemand et al. 1998), since it resembles the DNA structure initially proposed for untwisted DNA by (Pauling and Corey 1952): both structures have helices with the sugar-phosphate backbone inside and unpaired bases in solution, albeit P-DNA is a double helix and the Pauling-Corey structure is a triple helix.

In this chapter, we apply oxDNA2 to study positive supercoiling in MT experiments in a similar way to what has been done in chapter 5 and section 4.4.3 for negative supercoiling. As for the case of L-DNA, when we analyse P-DNA we are pushing oxDNA2 beyond the limits for which it was parametrised, since the specific structural and mechanical properties of P-DNA were not taken into account when parametrising the model. Traditionally, the torsionally melted phase, with roughly constant duplex extension and torque, is seen as contiguous to the

post-buckling phase, with extension decreasing in σ due to formation of plectonemes, and both extension and torque increasing with F. The transition between the two phases is reported at the critical force F_c. Instead, similarly to previous work by (Matek et al. 2015) for both positive and negative supercoiling and to what is discussed for negative supercoiling in chapter 4, we observe a richer picture at the torsional melting transition: a plectoneme tip-bubble phase with co-localised P-DNA bubble and plectoneme loop separates the phases with only plectonemes or bubble respectively. Instead of reporting a single critical force F_c, we find the two critical forces denoting the boundaries of the plectoneme tip-bubble phase with the post-buckling and torsionally melted phases, respectively $F_{c-} = 8$ pN and $F_{c+} = 9$ pN. These are much higher than $F_{c-} = 1.5$ pN, observed in 4.4.5 for tip-bubbles at negative supercoiling (at the same salt concentration). We measure the superhelical density at the buckling transition σ_b^p, as well as the post-buckling torque and slope, and find a good agreement with the model of (Brutzer et al. 2010), introduced in section 3.2. Finally, we investigate the P-DNA phase for up to $F = 50$ pN and $\sigma = 3.1$ in a duplex of length $L_0 = 600$ bp at an effective NaCl concentration of 100 mM. The P-DNA phase in oxDNA presents similar characteristics to the L-DNA phase discussed in chapter 5: the extension and torque-response curves are qualitatively similar, and both bubbles show strands tightly wound around each other (though P-DNA is much more tightly wound than L-DNA and of opposing helicity). Also, intriguingly, P-DNA shows solenoidal like configurations, much as L-DNA did.

The chapter is structured as follows: in section 6.1 we review the current literature on positively-supercoiled DNA, with particular regard to the behaviour at the buckling transition and P-DNA. Then, section 6.2 presents the simulation protocol. In section 6.3 we investigate the location of the buckling transition, discuss the measurements of duplex extension, and compare the results with theoretical modelling. Section 6.4 reports measurements of torque around the melting transition, while section 6.5 describes in more detail the P-DNA, torsionally melted phase. Finally, section 6.6 presents the conclusions and discusses future work.

6.1 Previous work

Experimental work by (Brutzer et al. 2010) has systematically investigated the superhelical density at the buckling transition σ_b in positively supercoiled DNA duplexes of length 1.9 kbp and 10.9 kbp, varying the NaCl concentration of the aqueous solvent between 30 mM and 320 mM. They also develop the theoretical model presented in section 3.2, which shows good agreement with their experiments. The simplicity of this modelling approach makes it a convenient tool to predict the location of the buckling transition when finite-size effects are taken into account, so we will be referring to this model in the following sections.

The first observation of torsionally melted DNA in a magnetic tweezers experiment was made by (Allemand et al. 1998) twenty years ago. They also modelled P-DNA in the rigid-base energy minimisation program JUMNA (Lavery, Zakrzewska, and Sklenar 1995), and measured the extension of a 17 kbp duplex in a 10 mM solution of phosphate buffer well into the P-DNA phase.

Important further work was carried out by (Bryant, Stone, et al. 2003), who used an optical trap to measure the torque plateau in the B-DNA/P-DNA transition at 34 pN nm (at F=45 pN) in a 14.8kb molecule in a solution with 100 mM NaCl and 40 mM Tris-HCl (pH 8.2).

The boundaries between pre-buckling B-DNA, post-buckling B-DNA, and the mixed B/P-DNA phases have also been investigated with optical tweezers in a phosphate-buffered saline (157 mM Na^+, 4 mM K^+, 12 mM PO_4^{3-}, 140 mM Cl^-, pH = 7.4) in a 4.3 kbp duplex (Sheinin and Wang 2009), finding that denaturation bubbles appear in positive supercoiling at about 6-8 pN and $\sigma = 0.06$, with a torque plateau of about 45 pN nm. (Marko and Neukirch 2013) report measurements of the twisting persistence length of P-DNA being 20 nm, from unpublished data by Z. Bryant.

Recent work by (Kriegel, Ermann, Forbes, et al. 2017) investigated the salt dependence of the torsional properties of a 7.9 kbp duplex with an NaCl concentration between 20 mM and 500 mM, reporting torsional melting transition for F >6 pN and $\sigma > 0.05$, for [NaCl] = 100 mM. The relative torque plateau has been measured to be about 35 pNnm.

6.2 Simulation protocol

Two slightly different simulation protocols were used to generate the results discussed in this chapter. The protocol used to study the case of $\sigma \leq 0.15$ is the one already discussed in section 2.4. A 624 bp duplex was generated in a straight configuration, where the 12 base-pairs at each end were used as handles. Harmonic traps were used on both handles to control their motion, thereby controlling σ, and a constant force was added to one of the ends to simulate the force F. First, overwinding simulations were run with the molecular dynamics backend and the Langevin thermostat of the oxDNA code for $F \in [6, 10]$ pN by continuously overwinding the duplex by 1 turn per 10^7 simulation steps. With the simulation step of 0.005 simulation time units, equivalent to 15 fs (using the mapping from section 2.2), this is roughly equivalent to one turn per 150 ns, such that the duplex is overwound to e.g. $\sigma = -0.1$ over approximately 850 ns. Then, simulations snapshots were extracted at the appropriate time to have the desired superhelical density σ. Finally, these were used as initial configurations for simulations at fixed σ. These were run for up to 10^9 simulation steps, requiring up to five days each on a GPU machine, equivalent to 15 μs. Other values of F were simulated by taking the last configuration obtained from systems at the same σ and at the closest F available, then setting the appropriate F. Each system was simulated in three independent replicas, and the quantities discussed in the following sections are estimated from the mean of the values measured in each replica. The error bars show the sample standard deviation among the three replicas.

The data for $\sigma \geq 0.4$ was instead generated with a slightly different protocol, with a lower period of rotation of 10^6 simulation steps (equivalent to 15 ns). This was necessary to overwind the system in a shorter simulation time - bringing σ up to 3.1 means that about 120 turns need to be added to the system, so that overwinding simulations were run for 1.2×10^8 simulation steps each. A similar protocol, already discussed in section 5.1 was used for L-DNA. Overwinding simulations were run for F=10 pN only. Simulations at different constant F were all started from the latest configuration at the closest force available, i.e. simulations at 10 pN were used to start those at F=6, 16 pN, simulations at 16pN were used to generate those at 30 pN, and so on. These simulations proved much faster to equilibrate

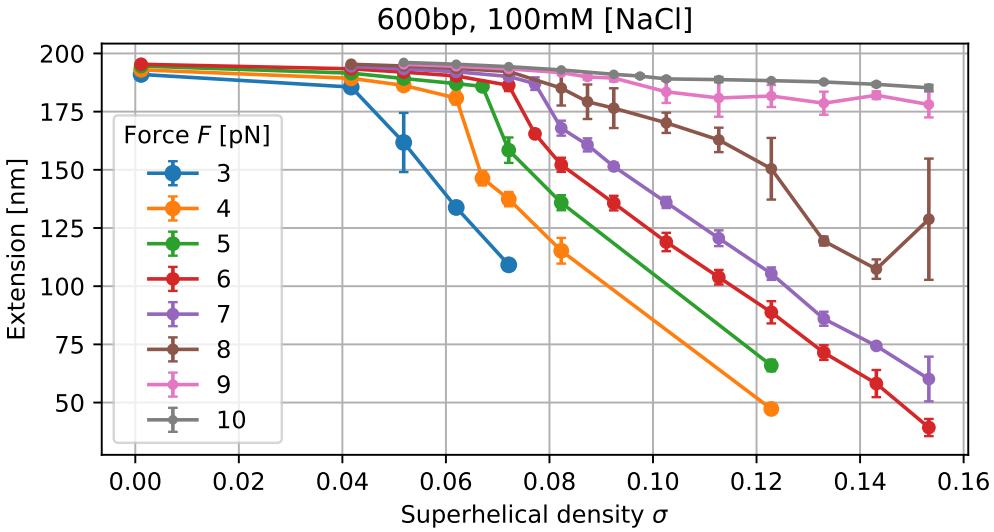

Figure 6.2: Duplex extension in positively supercoiled DNA for different pulling force F as a function of superhelical density σ.

than the ones at lower σ or F, such that simulating them for up to 2×10^8 steps was sufficient to produce the results in this chapter. Plots of the duplex extension, torque, and fraction of denatured base-pairs as a function of simulation time (often called running plots) were assessed when determining wether removing the first 10^7 steps was sufficient to account for equilibration.

6.3 Buckling transition and duplex extension at low superhelical density

This section describes the location of the buckling transition and the duplex extension for $\sigma \leq 0.16$. The extension curves generated during the simulations are shown in fig. 6.2. Up to $\sigma \approx 0.04$, the duplex extension is approximately constant for all F between 3 pN and 10 pN, and the duplex doesn't form any particular supercoiled structure beyond writhing fluctuations. At higher σ, the behaviour of the duplex is strongly dependent on F.

- For $F \leq$7 pN, the duplex extension has a sudden drop at the buckling transition supercoiling density σ_b^p, due to the formation of a plectoneme. Then, the duplex extension decreases linearly in the post-buckling regime.

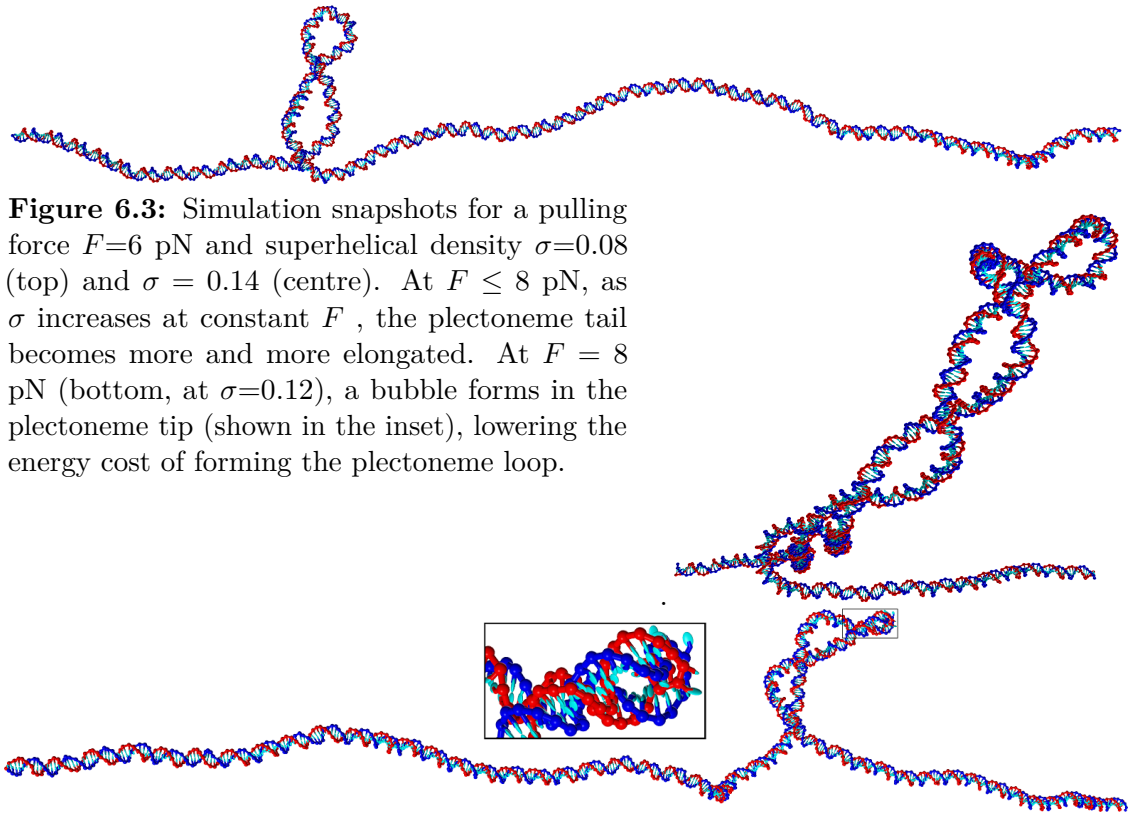

Figure 6.3: Simulation snapshots for a pulling force F=6 pN and superhelical density σ=0.08 (top) and $\sigma = 0.14$ (centre). At $F \leq 8$ pN, as σ increases at constant F , the plectoneme tail becomes more and more elongated. At $F = 8$ pN (bottom, at σ=0.12), a bubble forms in the plectoneme tip (shown in the inset), lowering the energy cost of forming the plectoneme loop.

- When F reaches $F_{c-} = 8$pN, the post-buckling slope is significantly lower, and no clear σ_b^p can be identified. This is due the presence of a tip-bubble phase, similar to the one observed by (Matek et al. 2015) and described in section 2.4, also observed in chapter 4: the denaturation bubble has a lower bending modulus than DNA, lowering the energy cost of loop formation in a similar manner to what a mismatch would do according to the theory in chapter 3. This then allows for the formation of a plectoneme tail. Images of plectonemes with and without tip-bubbles are presented in fig. 6.3.

- Finally, for F above $F_{c+} = 9$ pN, the duplex extension is approximately constant throughout the $\sigma < 0.16$ regime. Instead of forming plectonemes, the excess turns are stored by torsional melting, as discussed in the following section.

For $F \leq F_{c-}$, the location of the drop is a measurement of the buckling transition σ_b^p, and is plotted in fig. 6.4 together with the estimate provided by the model

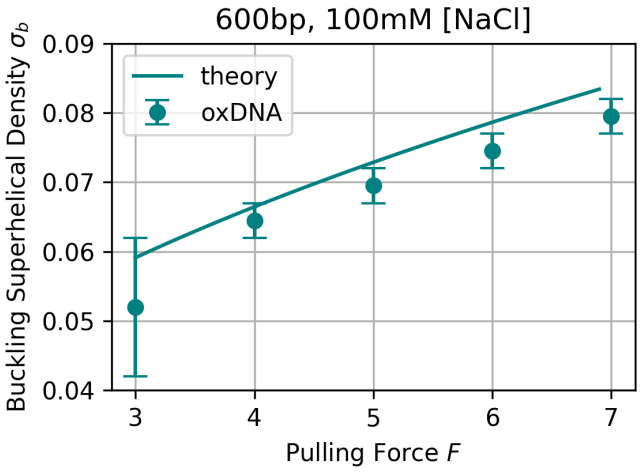

Figure 6.4: Buckling superhelical density σ_b^p as a function of F as obtained in oxDNA (points) and with the model by (Brutzer et al. 2010) (line).

by (Brutzer et al. 2010), see section 3.2

$$\sigma_b^p = \frac{E_{\text{tail}}}{\frac{C_s}{d}(2\pi)^2} + \sqrt{\frac{2(E_{\text{loop}} - E_{\text{tail}})}{\frac{C_s}{d^2}(2\pi)^2 L_0}}\,. \tag{6.1}$$

where $d = 10.55\text{bp/turn} = 3.4815$ nm/turn is the helical repeat of the oxDNA2 duplex (Snodin et al. 2015). The effective torsional stiffness (Moroz and Nelson 1997) is

$$C_s = C\left[1 - \frac{C}{4pk_BT}\sqrt{\frac{k_BT}{pF}}\right]\,, \tag{6.2}$$

with the temperature T, the Boltzmann constant k_B, the bending persistence length $p = 45$ nm, the torsional stiffness $C = 395$ fJfm measured for oxDNA2 at this salt concentration (Snodin et al. 2015). The loop energy is computed in the circular approximation

$$E_{\text{loop}} = 2\pi\sqrt{2pk_BTFz(F)}\,, \tag{6.3}$$

with the zero-torque relative extension (Marko 2007)

$$z = 1 - \frac{1}{2}\sqrt{\frac{pF}{k_BT}} - \frac{1}{32}\,. \tag{6.4}$$

Finally, the energy cost of storing one excess turn in the plectoneme tail is proportional to the post-buckling torque Γ

$$E_{\text{tail}} = 2\pi\Gamma = 2\pi\sqrt{\frac{2P(F - \sqrt{\frac{k_BTF}{p}})}{1 - \frac{P}{C_s}}}\,. \tag{6.5}$$

where we set $P = 26 k_B T$ nm as discussed in section 3.2. The model shows good agreement with the measurements, overestimating the location of the buckling transition by less than 0.005σ. The curves do not significantly change when P changes in the aforementioned interval. Qualitatively, the increase of σ_b^p as F increases is due to the fact that the energy cost of forming the loop also increases with F (see eqn. 6.3), so that more turns can be stored in the pre-buckling state before the formation of the loop becomes energetically convenient.

The post-buckling slope of the extension curves (Marko 2007), given in the form computed by (Brutzer et al. 2010), is

$$\frac{dL}{dN} = 2\pi \frac{1 - \frac{1}{2}\sqrt{\frac{k_B T}{pF}} - \frac{C^2 g \left(\frac{k_B T}{pF}\right)^{3/2}}{8(C_s k_B T)^2 \left(\frac{1}{P} - \frac{1}{C_s}\right)}}{\sqrt{2Pg\left(\frac{1}{P} - \frac{1}{C_s}\right)}}, \tag{6.6}$$

where g is the negative free energy of a stretched, nicked (torsionally unconstrained) duplex

$$g = F - \sqrt{\frac{k_B T F}{p}}. \tag{6.7}$$

Using $P = 26\ k_B T$ nm gives a slope varying from -27 nm/turn for $F = 4$ pN to -21 nm/turn for $F = 7$ pN, in reasonable agreement with the slope of our measured oxDNA curves extracted from fig. 6.2, varying from -31 nm/turn for $F = 4$ pN to -27 nm/turn for $F = 7$ pN.

6.4 The onset of torsional melting and torque at low σ

Measurements of torque and fraction of denatured base-pairs, defined as discussed in section 2.4, are shown in fig. 6.5. For $\sigma \leq \sigma_b^p$, the torque grows linearly for all forces: in this linear torque-response regime, DNA stores all the excess linking number as pure twist, so that DNA behaves effectively as a torsional spring (see section 3.2).

Next, the torque curves present an overshoot, plateauing after an initial drop. The plateau is expected from eqn. 6.5, which doesn't contain any σ dependence. It does, however, depend on F, increasing from about 25 pNnm at $F = 3$ pN

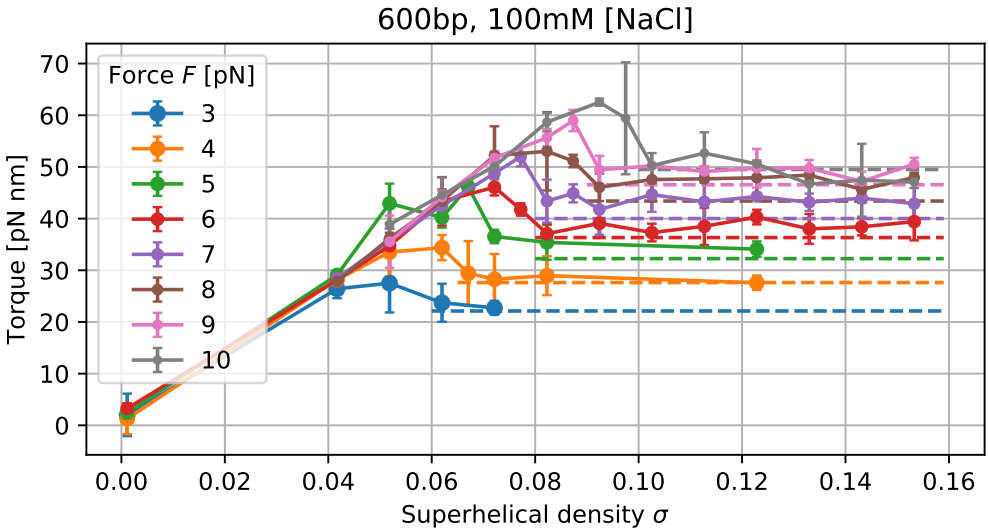

Figure 6.5: Torque as a function of superhelical density σ for several different pulling forces F. The dashed lines show theoretical estimates for the torque plateau Γ from eqn. 6.5.

Source	Solvent	Γ_m [pNnm]	σ_m	F_{c}- [pN]
this work	100 mM NaCl	50	0.09	8
(Kriegel, Ermann, Forbes, et al. 2017)	100 mM NaCl	35	0.06	6.5
(Sheinin and Wang 2009)	157 mM Na^+, 4 mM K^+,12mM PO_4^{3-}, 140 mM Cl^-, pH = 7.4	45	0.06	7
(Bryant, Stone, et al. 2003)	100 mM NaCl, 40 mM Trish-HCl (ph 8.2)	34	0.03	N/A
(Allemand et al. 1998)	150 mM NaCl, 10mM PB	N/A	N/A	6

Table 6.1: Torque at the torsional melting transition Γ_m, critical force F_c, and supercoiling density at the melting transition as reported in this work and in previous experiments.

to about 50 pNnm $F = 8$ pN. In this regime, simulations show good agreement with the value of the post-buckling torque Γ predicted by eqn. 6.5, when using $P = 26\ k_BT$ nm. The torque at the plateau does not increase any further for $F > 8$ pN, even at much higher force (as will be shown in section 6.5.3). The fact that the torque plateau (within error bars) at Γ_m for higher F can be ascribed to the onset of the torsional melting transition: forming a plectoneme becomes more energetically expensive than storing the excess linking number as denaturation bubbles, at a cost that doesn't depend on F nor σ.

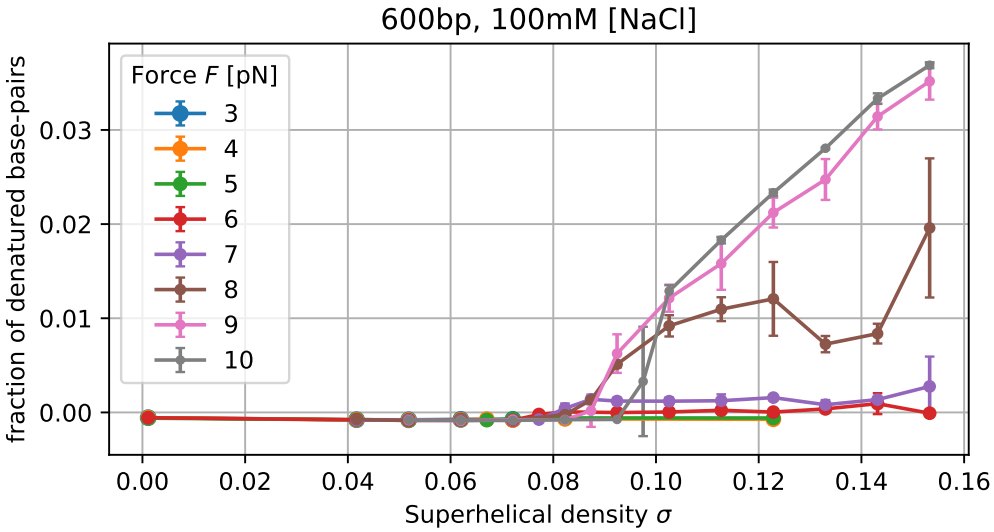

Figure 6.6: Fraction of denatured base-pairs as a function of superhelical density σ for several different pulling forces F.

We can observe directly the presence of the melted phase by measuring the fraction of denatured base-pairs: as shown in fig. 6.6, no significant fraction of open base-pairs is observed for $F \leq 7$ pN. At $F = 9$ pN, 10 pN, the fraction of denatured base-pairs increases roughly linearly as a function of σ, with a weak dependence of F. Values of the melting torque Γ_m, of the superhelical density σ_m at the torsional melting transition, and of F_{c-} (measured as the force when the torque plateaus at Γ_m for increasing F) are reported in table 6.1, showing semi-quantitative agreement with our simulations. At F=8 pN, the fraction of denatured base-pairs is affected by the presence of tip-bubble plectonemes, showing a non-linear trend in σ.

The torsionally melted phase has an interesting phenomenology, which is explored more in detail in the following section.

6.5 P-DNA at larger supercoiling density and force

At high enough F and σ denaturation bubbles are preferred to plectonemes, even to those containing tip-bubbles, as a way to store the excess linking number. This happens for about $F \geq 9$ pN and $\sigma > 0.09$ in oxDNA2, as shown in section 6.4. In order to study this torsionally melted P-DNA phase, simulations were ran for

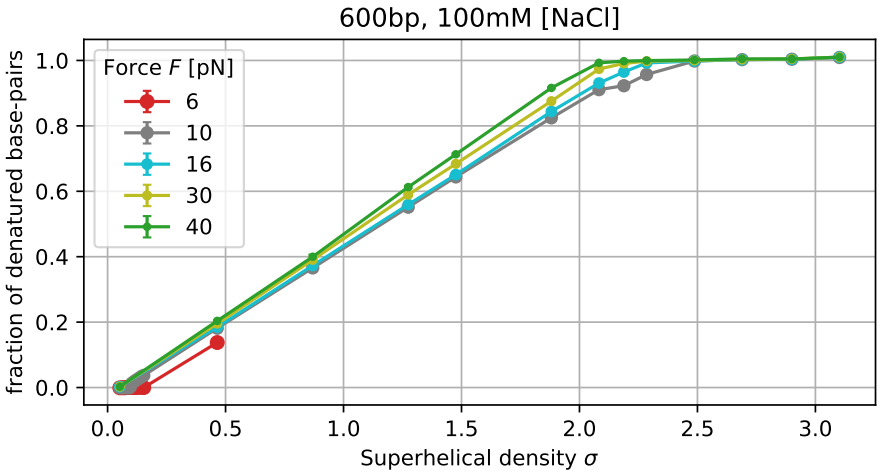

Figure 6.7: Fraction of denatured base-pairs in P-DNA. Notice how the superhelical density σ_P at which the duplex denaturation is complete moves from 2.5 for F=10 pN to 2.1 for F=40 pN.

σ up to 3.1 and F=6, 10, 16, 30, 40 pN in oxDNA. This section presents the results of these simulations.

6.5.1 Phase characterisation and supercoiled structures

Just like L-DNA, the fraction of denatured base-pairs (shown in fig. 6.7), defined consistently with the rest of the thesis according to the criterion of (Šulc et al. 2012) (see section 2.4) can be used as a proxy for the fraction of base-pairs in the P-DNA state.

After the onset of the torsional melting transition (discussed in section 6.4), the fraction of P-DNA rises linearly, until the whole duplex has transitioned to the P-DNA phase. The slope is weakly dependent on force, such that the supercoiling density at the pure P-DNA phase σ_p varies from approximately 2.5 for F= 10 pN to 2.1 for F = 40 pN. For F = 6 pN, the denaturation fraction is significantly lower than that of higher forces, because of the writhe stored in the collapsed plectoneme phase, as discussed below.

When in a straight configuration, P-DNA observed in oxDNA is similar to previous modelling by (Allemand et al. 1998) and (Wereszczynski and Andricioaei 2006), and to the L-DNA phase seen in chapter 5, shown side by side in fig. 6.8. All three present denatured base-pairs arranged in a double-helical structure, with

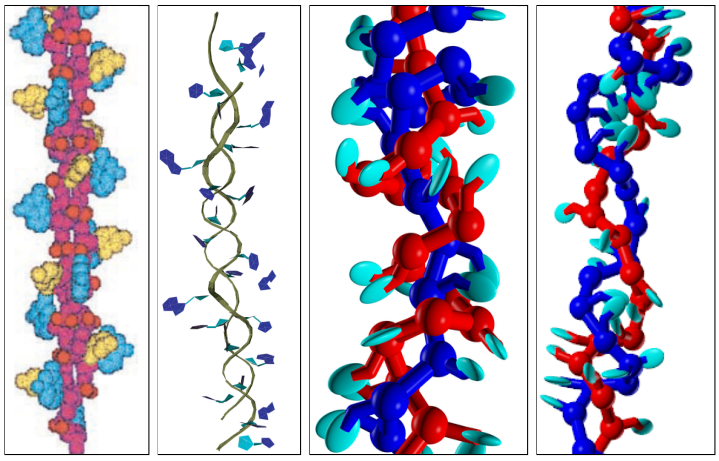

Figure 6.8: Left to right: comparison between P-DNA modelled in the coarse-grained model JUMNA (Lavery, Zakrzewska, and Sklenar 1995) (adapted from (Allemand et al. 1998)) (first), in CHARMM (Brooks et al. 2009) (adapted from (Wereszczynski and Andricioaei 2006)) (second), as observed in oxDNA (third), and L-DNA in oxDNA (right).

P-DNA being right handed and L-DNA being left handed. However, P-DNA has a much lower helical repeat than L-DNA. This is expected from the fact that, albeit with a similar density of *excess* turns σ, the total density of turns

$$\rho = \frac{\mathrm{Lk}}{\mathrm{Lk}_0} = 1 + \sigma, \tag{6.8}$$

is much higher, with Lk the linking number of a duplex in a given configuration and Lk_0 the linking number in absence of torque. In absence of writhe, ρ can be thought of as a measure of duplex winding relative to the equilibrium state, such that parallel strands have $\rho = 0$ and a left-handed duplex with the same helical repeat of usual B-DNA has $\rho = -1$. Right-handed turns added with positive supercoiling stack with the turns present in a duplex at equilibrium, while the left-handed turns added with negative supercoiling partly cancel them out. For example, P-DNA at $\sigma = 2$ has $\rho = 3$, while L-DNA at $\sigma = -2$ has $\rho = -1$.

Like for L-DNA, the P-DNA phase often shows significant writhe. To the best of our knowledge, these supercoiled structures have not been characterised before, so it's worth documenting in detail their wide phenomenology. For $\sigma < \sigma_p$, the duplexes are in a mixed B-DNA/P-DNA phase. For $F \geq 10$ pN (see fig. 6.9), the P-DNA phase writhes in solenoidal structures similar to the ones observed in L-DNA, albeit somewhat stiffer and less dynamical. As F increases from 10 pN to 16pN, the radius of the solenoids decreases, and the solenoids become barely visible at F=30, 40 pN. P-DNA's estimated torsional modulus $C = 20 \pm 10\ k_BT$ nm and bending modulus B=19 k_BT nm are significantly lower than those for B-DNA

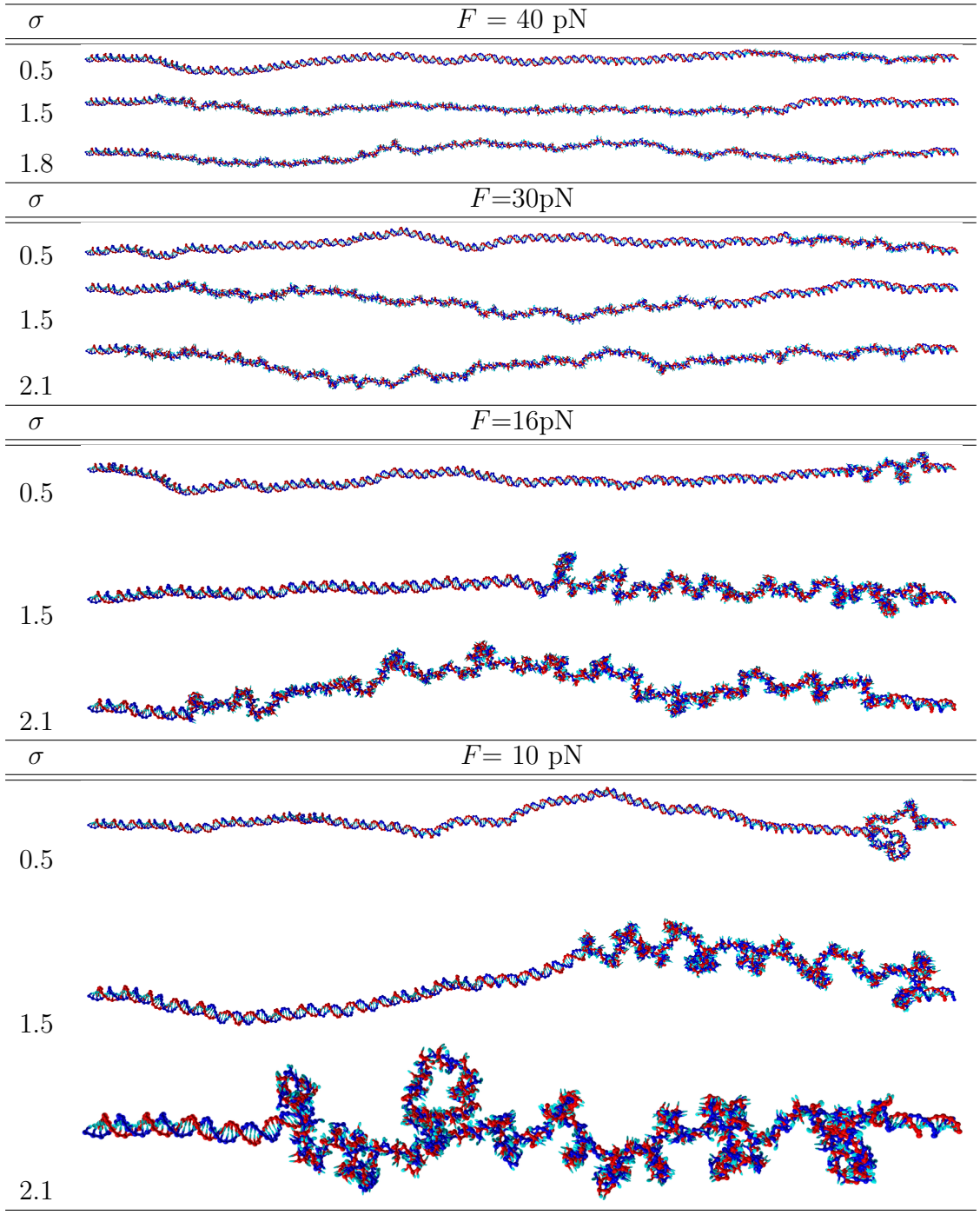

Figure 6.9: Simulations snapshots of duplexes in the mixed B-DNA/P-DNA phase for F between 10 pN and 40 pN. As the supercoiling density σ is increased, more and more of the duplex assumes the P-DNA phase to relax linking number. Note the solenoidal configurations of the P-DNA.

(Marko and Neukirch 2013), and preliminary results by the theory developed by Romain Rollin (introduced in section 5.4) predicts that for these values relaxing writhe through solenoids is energetically favoured to doing so by forming plectonemes. Still, more work is needed to assess whether these solenoids also form in P-DNA in experiments or only in oxDNA.

At $\sigma = \sigma_p$, the P-DNA transition is complete. Any further linking number is stored by increasing the writhe, i.e. increasing the size of the solenoids, as can be seen in fig. 6.10.

Finally, at F=6 pN the plectoneme phase is still preferred over base-pair opening, as seen in section 6.3. At $\sigma = 0.5$, we observe hybrid configurations, that look somewhat like a collapsed plectoneme with denaturation bubbles arranged along its length (see fig. 6.11).

6.5.2 Duplex extension

The duplex extension curves for $\sigma \leq 3$ for and different F are shown in fig. 6.12. At F= 6 pN, the duplex stores σ first by forming a plectoneme loop, then elongating the plectoneme tail, and finally denatures base-pairs only once that the plectonemic structure extends to the whole chain. On the other hand, at F= 10 pN the extension of the duplex decreases linearly until $\sigma = 2.9$, flexing slightly at $\sigma = 3.1$. The curves for $F \geq$16 pN have a similar behaviour, but change slope at σ_p. In particular, the curves for F=30, 40 pN are approximately constant (respectively weakly decreasing and increasing) for $\sigma < \sigma_p$.

The data shown is qualitatively similar to that of (Allemand et al. 1998). Both plots show a change of slope at $\sigma = 2 \div 2.5$ for $F = 40$ pN, an approximately constant extension for F=30 pN, and comparable trends for F=10, 16 pN. However, their extension curves for 6 pN do not show any relatively steep initial slope, signifying the absence of B-DNA plectonemes. This is probably due to the different buffer (10 mM phosphate buffer), which increases the energy cost of the plectoneme tail and lowers that of base-pair opening.

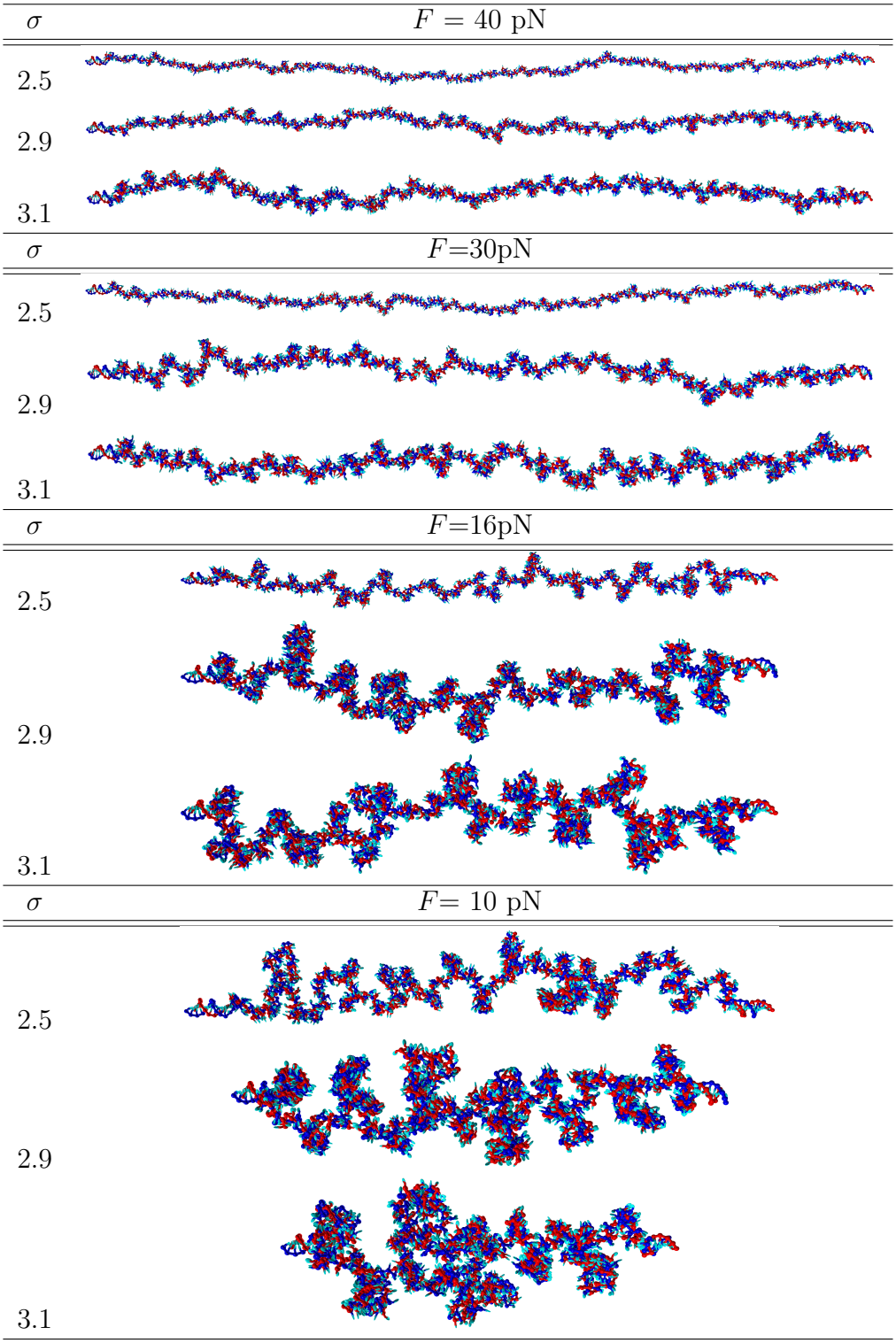

Figure 6.10: Simulation snapshots of duplexes in the pure P-DNA state, for different values of pulling force F and supercoiling density σ (not in scale). The solenoidal structures have a radius that decreases with increasing force, and increases with increasing σ. Larger σ also shows more complex writhed structures, especially for $F = 10$ pN.

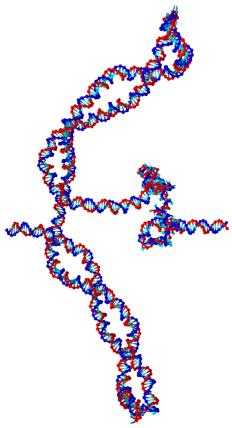

Figure 6.11: Supercoiled structure observed at a pulling force $F = 6$ pN and a superhelical density $\sigma = 0.4$. Notice the coexistence of collapsed plectonemes and P-DNA.

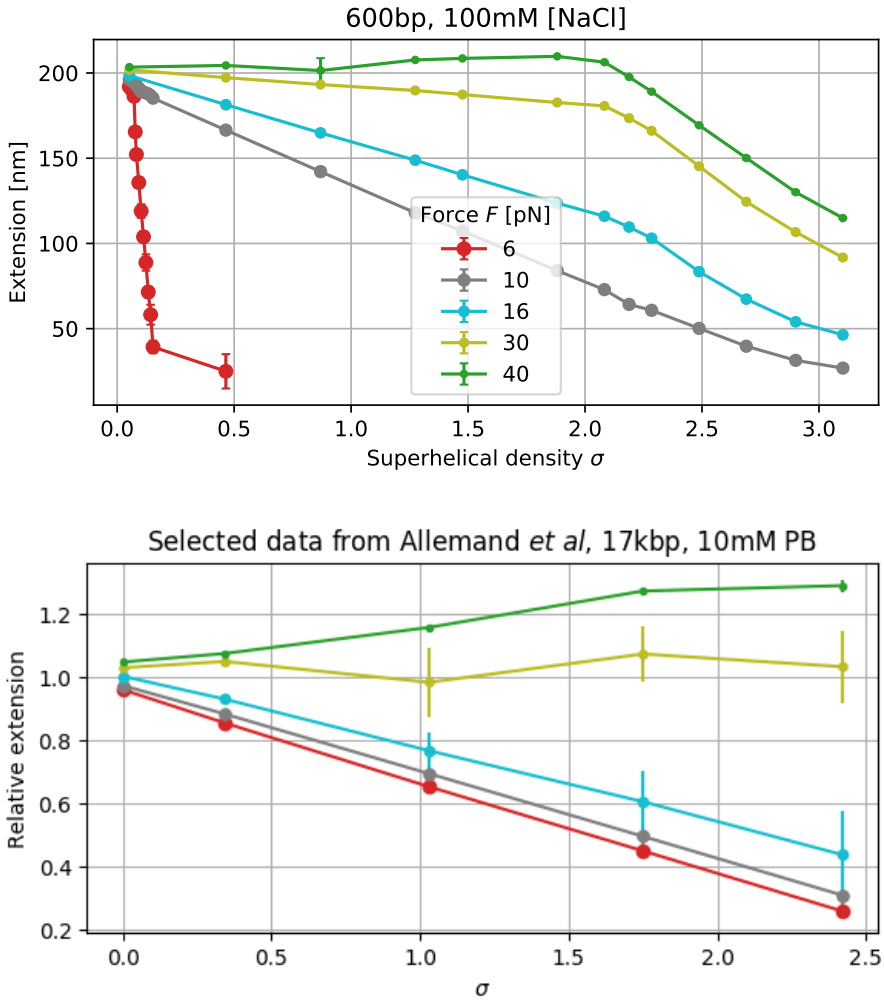

Figure 6.12: Duplex extension for P-DNA in oxDNA (top) and as reported by (Allemand et al. 1998) (bottom). The error bars show the hysteresis between winding and unwinding experiments.

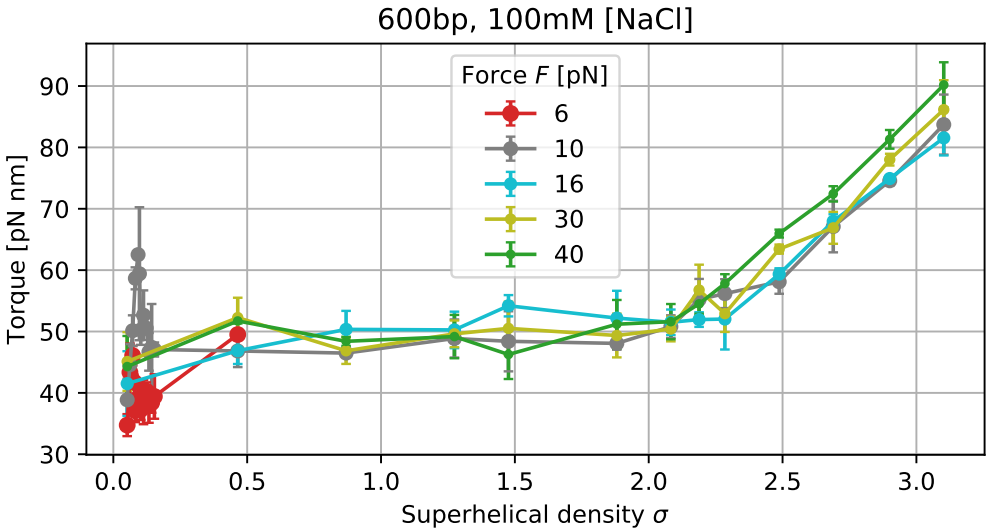

Figure 6.13: Torque curves for P-DNA.

6.5.3 Torque

The torque acting on the moving handle has been measured according to the protocol detailed in section 2.4. In the mixed B/P-DNA phase, the torque has a constant value of about 50 pNnm, regardless of F and of whether the torsional melting transition happens in the collapsed plectoneme state (such as for F= 6 pN) or in the B-DNA phase (for $F \geq 10$ pN), consistently with the observations of section 6.4.

Then, for $\sigma > \sigma_p$, the torque again rises linearly, up to 90 pNnm for $\sigma = 3.1$ and $F = 40$ pN. In the pure phase, the difference $\Delta\Gamma$ between the torque at $\sigma > \sigma_p$ and the torque at the onset of the pure phase Γ_p is expected to be proportional to $\sigma - \sigma_p$, and to the effective torsional modulus C_{eff} (Kriegel, Ermann, Forbes, et al. 2017)

$$\Delta\Gamma = \frac{C_{\text{eff}}}{d} 2\pi(\sigma - \sigma_p)\,. \tag{6.9}$$

Linear fits of the torque curves in the $\sigma > \sigma_p$ regime yield the values for C_{eff} in table 6.2. Notice that C_{eff} is not the true torsional modulus of P-DNA C: instead, it is a lower bound that takes into account the formation of supercoiled structures, and is expected to depend on both the bending torsional modulus B and F, in a similar way to the Moroz-Nelson model (Moroz and Nelson 1997). However, given the substantial absence of such structures for $F = 40$ pN all the way until $\sigma = 3.1$, it is reasonable to use the C_{eff} at this force as a measure of C, taking

F [pN]	C_{eff} [$k_BT$nm]	C_{eff} [pNnm^2]=[fJfm]
10	5.3(1)	22
16	4.9(1)	20
30	4.6(3)	19
40	5.0(1)	21

Table 6.2: Estimates of the torsional modulus C_{eff} of P-DNA obtained from a linear fit of the torque in the P-DNA pure phase $\sigma > \sigma_p$.

$C = 5\ k_BT$ nm for P-DNA, in reasonable agreement with the value of $C = 20 \pm 10\ k_BT$ nm reported by (Marko and Neukirch 2013).

6.6 Conclusion

This chapter presents the results of investigations of the effects of positive supercoiling in a 600bp duplex at an effective salt concentration of 100 mM. In particular, the chapter explores the boundaries between the pre-buckling, post-buckling, and torsionally melted phases, and characterises the torsionally melted P-DNA phase.

For $F \leq 7$ pN, the superhelical density at the buckling transition σ_b^p is well described by the model by (Brutzer et al. 2010). Theoretical estimates also provide semi-quantitative agreement with the simulation results for the post-buckling slope $dL/d\sigma$.

At a pulling force of F=8 pN, we observe a significant deviation from the classic buckling picture described above; a P-DNA bubble co-localises with the plectoneme loop. This reduces the loop size, and therefore its energy cost, allowing for an extension curve with a slope significantly different from that observed at lower forces. The tip-bubble phase is similar to that previously observed by (Matek et al. 2015), and in chapter 4 for negative supercoiling.

In agreement with previous studies (Janssen et al. 2012; Marko 2007), the torque rises linearly for low σ, then it suddenly decreases and plateaus. The value of the torque plateau increases with F at first, but it later stays constant for $F \geq 8$: these forces are so high that plectonemes are energetically unfavourable and excess turns are relaxed by denaturing the base-pairs. The torque plateau in this phase is observed at 50 pNnm, similarly to what has been reported in previous studies

at the same (Kriegel, Ermann, Forbes, et al. 2017) or similar (Bryant, Stone, et al. 2003; Sheinin and Wang 2009) salt concentration.

The torsionally-melted P-DNA phase is discussed in section 6.5. The fraction of bases in the P-DNA phase rises linearly as σ increases, until the pure phase is reached at a superhelical density σ_{P} that can vary from 2.5 for F= 10 pN to 2.1 for F = 40 pN. The P-DNA phase in oxDNA has a similar appearance to a previous model (Allemand et al. 1998), and to the L-DNA discussed in chapter 5, albeit it is more tightly wound and right-handed as opposed to left-handed. Like L-DNA, P-DNA forms solenoidal structures instead of plectonemes, with the radius and pitch of the solenoids increasing as F decreases and σ increases. The torsional modulus C of P-DNA is measured to be about 5 nmk_BT, not too far from a previous estimate of 20±10 nmk_BT (Marko and Neukirch 2013). For either of these values of C and for the bending modulus B=19 k_BT nm as reported by (Marko and Neukirch 2013), the theoretical model of Romain Rollin (see section 5.4) predicts that solenoids are favoured to plectonemes as a means to relax excess writhe, which corroborates the hypothesis that solenoids are indeed observed in DNA in experiments as well.

Despite the first measurements on P-DNA going back to at least 20 years ago (Allemand et al. 1998), the study of P-DNA is still in its infancy. Extension and torque curves for several F well within the pure phase, analogous to the ones for L-DNA produced by other experimental groups (Sheinin, Forth, et al. 2011; Vlijm et al. 2015), would be an obvious improvement and a way to further validate or refute the applicability of this modelling approach to real DNA. The importance of positive supercoiling for hyper-thermophilic organisms would suggest the importance of studying P-DNA at higher temperatures, similarly to what similar work did for the negative supercoiling melting transition (Galburt et al. 2014). Likewise, the influence of salt concentration on P-DNA is a topic of interest of basic biophysics: oxDNA simulations are currently running to repeat the measurements of section 6.5 in an effective NaCl concentration of 500 mM, which could be compared to experimental work at a similar salt concentration (Léger et al. 1999).

Where all the hanging threads are finally resolved, showing a bird's eye view on the grand scheme of things relating to this piece of science. Curtains fall.
Applauses ensue.
Or a volley of rotten cabbages, that's also cool.

7 Conclusions

Contents

7.1 Summary

In this thesis, we have applied different coarse-grained models to supercoiled DNA duplexes in several different conditions. In chapter 3, we extend the simple modelling approach of (Brutzer et al. 2010) to investigate how the presence of anomalous DNA sequences (i.e. naturally bent or significantly floppier than a regular duplex) influences plectoneme loop energetics. The equations stemming from the theory are surprisingly simple and elegant in the circular loop approximation, and can be obtained by either applying general dimensional arguments to the Hamiltonian of an elastic rod or by direct energy minimisation. Anomalous loops, i.e. loops containing an anomalous sequence, are energetically cheaper than regular loops; this affects the energetics of the whole supercoiled system, most notably favouring plectoneme localisation on anomalous sequences. This effect gets stronger as the pulling force F increases, and it can be summarised by the plectoneme pinning coefficient Q, which can be computed in the theory from the parameters of the anomalous sequence,

and in principle measured in molecular tweezer (MT) experiments. Theoretical prediction for Q show semi-quantitative agreement with simulations in an extension of the twistable elastic polymer (TEP) model of (Brackley, Morozov, et al. 2014). The model has been tested for duplexes containing segments with the same elastic parameters as mismatches of 1, 2, 4, and 10 base-pairs, as well as for bent sequences with different bending angles. Because of their lower energy cost, anomalous loops also nucleate at lower superhelical density σ than regular loops: this leads to a thin region in the $F - \sigma$ phase diagram where plectonemes are stable only when located on the anomalous sequence.

The theory developed in chapter 3 is applied to the thymine dimer (TD), the most commonly formed DNA photoproduct, in chapter 4. First, the TD is modelled in oxDNA, a somewhat more versatile and realistic DNA model than TEP. We modify the oxDNA interaction potential in order to introduce a non-zero bending angle θ_0 in the TD-containing duplex, and weakening the hydrogen bonding interaction to reproduce the melting temperature of a TD-containing oligonucleotide as reported in the literature. Because of the relative variance in the properties of the TD reported in the literature, three different models of the TD are considered. Then, the model is used to simulate a MT experiment on a 600 bp duplex containing a TD. The findings further corroborate the theory, showing semi-quantitative agreement with theoretical prediction of Q and of the superhelical density at the buckling transition σ_b^p.

Furthermore, the denaturation probability of the TD-containing base-pairs is shown to be strongly dependent on σ, increasing by up to two orders of magnitude between unsupercoiled and physiologically supercoiled DNA. It is thought that having denatured TD base-pairs would lower the barrier for enzymes to repair the damage. If this result were confirmed through further studies, this could limit the applicability of the many experimental and theoretical studies that investigate the efficiency of the DNA repair mechanism on unsupercoiled DNA. We also propose that the favoured plectoneme localisation on a TD could lower the time needed for a DNA repair enzyme to find the damage, possibly provide a piece of the solution to the famous DNA-protein search problem. Finally, we extend the Benham-Fye model of DNA denaturation to treat TD, and investigate the impact that a TD

can have on the denaturation profile of a plasmid. We find that the location of the TD can have significant impact on the opening of an origin of replication bubble, which could in turn affect the replication of plasmids.

In chapter 5, we use oxDNA to investigate the torsionally melted DNA phase, commonly known as L-DNA. Simulation results show semi-quantitative agreement with experiments in both duplex extension and torque. Interestingly, the pure L-DNA phase does not store writhe in the form of plectoneme, like B-DNA does. Instead, simulations show highly fluctuating solenoids, with a radius decreasing with F and increasing with σ. To the best of our knowledge, this is the first observation of solenoidal behaviour in denatured DNA. Characterising these solenoids is not straightforward and is not obvious what order parameter to use to classify them, so we show several simulation snapshots to make sure that we do not filter out potentially important information. An argument is made that fluorescence images of L-DNA from previous studies could be showing solenoids instead of plectonemes. The position that this behaviour is not unique to the oxDNA model, but is shared with real DNA, is corroborated by theoretical results that show how a polymer with the experimentally measured bending and twisting stiffnesses would preferentially form solenoids than plectoneme. Finally, a sketch of possible experimental tests of the solenoidal nature of L-DNA is proposed.

The focus of chapter 6 is studying the behaviour of positively supercoiled DNA in MT with the oxDNA model. In particular, the chapter begins by investigating the phase boundaries between the pre-buckling, post-buckling, and torsionally melted phase. Duplex extension and torque have semi-quantitative agreement with previous experiments, and the location of the phase boundaries is found to have good agreement with experiments at the same salt concentration. In particular, the model by Brutzer et al. 2010 shows a good agreement with the location of the buckling transition for all the forces considered. We also investigate the behaviour of the positively supercoiled, torsionally melted DNA phase, called P-DNA. The simulations show semi-quantitative agreement with previous experiments. Like L-DNA, P-DNA prefers to form solenoids to plectonemes in oxDNA, and theoretical

calculations show that solenoids are preferred to plectonemes in a polymer with the bending and twisting stiffnesses reported in the literature for P-DNA.

7.2 Future work

The theory in chapter 3 is worth testing more thoroughly, both with experiments and simulation, to determine its range of applicability. In particular, the fact that the loop energy difference due to an anomalous sequence has a different trend on F depending on whether the sequence is bent or floppy is yet to be verified. Longer, more thorough simulations like the ones performed in chapter 3 would be able to ascertain this. Combining fluorescence imaging and MT experiments would also be a way to observe plectoneme pinning by these sequences. Our theory of loop energetics is reasonably crude, and can definitely be improved upon, for example by abandoning the circular loop approximation

In order to make the results of chapter 4 more robust, we have considered three different models of the TD, finding qualitatively similar results. It would now be interesting to compare the results of these simulations with real MT experiments, where the position of the plectoneme can be detected with fluorescence microscope. The modelling approach used to treat the TD in oxDNA can be generalised for other kinds of photoproducts, such as the Dewar isomer and the (6-4) photoproduct, as well as other DNA modifications that induce a bend and change the floppiness of the duplex, and even for sequence-dependent base-pair crookedness. We have merely started to scratch the surface of possible biological implications of induced bend and floppiness by DNA modifications, so that the proposed position-dependent impact on the plasmid denaturation profile and the mechanism for the protein-DNA search problem are just early-stage ideas. Further research will be needed to assess the applicability of these and other effects for biology. OxDNA simulated MT experiments with the MD backend are probably much slower than they need to be. More efficient simulations approach, possibly involving umbrella sampling or global Monte Carlo (such as virtual move Monte Carlo), already implemented in the oxDNA code, should be attempted in order to treat longer systems.

In chapter 5, we use oxDNA to investigate torsionally melted underwound DNA (L-DNA). The main finding of the chapter is that L-DNA in oxDNA simulations relaxes writhe with solenoids rather than plectonemes. While limited experimental results are compatible with this picture, more research is needed to confirm this. In that case, the theory of L-DNA would have to be rewritten to take solenoids into account. Then, more studies can be carried out to investigate the effects that sequence, temperature, salt condition and contour length would have on the behaviour of these systems.

The remarks discussed above for chapter 6 also apply to torsionally melted overwound DNA (P-DNA), investigated in chapter 6, with the difference that there are no experiments to confirm or deny the presence of solenoids. Experimental studies are therefore particularly important to understand this system. The effect that salt, temperature, sequence and contour length have on the phase boundaries should also be investigated. More oxDNA simulations to investigate P-DNA in 500 mM effective salt concentration have already been run, but we did not include them in the thesis because of timing and space constraints. Investigating the behaviour of positively supercoiled at higher temperatures could be valuable to understanding the biology of hyper-thermophiles.

References

Allemand, J F et al. (1998). "Stretched and overwound DNA forms a Pauling-like structure with exposed bases". In: *Proceedings of the National Academy of Sciences* 95.24, pp. 14152–14157. URL: https://doi.org/10.1073/pnas.95.24.14152.

Allen, Michael P and Dominic J Tildesley (1991). *Computer simulation of liquids*. Oxford University Press.

Ambjörnsson, Tobias et al. (2007). "Master equation approach to DNA breathing in heteropolymer DNA". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 75.2, pp. 1–13. arXiv: 0610547v1 [cond-mat].

Argudo, David and Prashant K. Purohit (2012). "The dependence of DNA supercoiling on solution electrostatics". In: *Acta Biomaterialia* 8.6, pp. 2133–2143. URL: http://doi.org/10.1016/j.actbio.2012.01.030.

Avery, O T, C M Macleod, and M McCarty (1944). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III". In: *The Journal of experimental medicine* 79.2, pp. 137–58. URL: https://doi.org/10.1084/jem.79.2.137.

Ball, M P (2018). *No Title*. URL: https://en.wikipedia.org/wiki/File:DNA_chemical_structure.svg.

Bar, A., Y. Kafri, and D. Mukamel (2007). "Loop dynamics in DNA denaturation". In: *Physical Review Letters* 98.3, pp. 1–4.

Barone, Flavia et al. (1995). "Effect Of Thymine Dimer Introduction In A 21 Base Pair Oligonucleotide". In: *Photochemistry and Photobiology* 61.1, pp. 61–67. URL: http://doi.wiley.com/10.1111/j.1751-1097.1995.tb09243.x.

Bates, A D and A Maxwell (2005). *DNA topology*. Oxford University Press.

Bauer, William R. and Craig J. Benham (1993). "The free energy, enthalpy and entropy of native and of partially denatured closed circular DNA". In: *Journal of Molecular Biology* 234.4, pp. 1184–1196. URL: https://doi.org10.1006/jmbi.1993.1669.

Benham, Craig J (1979). "Torsional stress and local denaturation in supercoiled DNA". In: *Proceedings of the National Academy of Sciences* 76.8, pp. 3870–3874. URL: https://doi.org/10.1073/pnas.76.8.3870.

Benham, Craig J. (1992). "Energetics of the strand separation transition in superhelical DNA". In: *Journal of Molecular Biology* 225.3, pp. 835–847. URL: https://www.sciencedirect.com/science/article/pii/0022283692904048.

Benham, Craig J. and Chengpeng Bi (2004). "The Analysis of Stress-Induced Duplex Destabilization in Long Genomic DNA Sequences". In: *Journal of Computational Biology* 11.4, pp. 519–543. URL: http://doi.org/10.1089/cmb.2004.11.519.

Berger, Mitchell A and Chris Prior (2006). "The writhe of open and closed curves". In: *Journal of Physics A: Mathematical and General* 39.26, pp. 8321–8348. URL: https://doi.org/10.1088/0305-4470/39/26/005.

Bi, Chengpeng and Craig J Benham (2004). "WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA." In:

Bioinformatics (Oxford, England) 20.9, pp. 1477–9. URL: https://doi.org/10.1093/bioinformatics/bth304.

Biebricher, Andreas S. et al. (2015). "The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics". In: *Nature Communications* 6, p. 7304. URL: http://www.nature.com/doifinder/10.1038/ncomms8304.

Bouchiat, C. and M. Mézard (1998). "Elasticity Model of a Supercoiled DNA Molecule". In: *Physical Review Letters* 80.7, pp. 1556–1559. URL: http://link.aps.org/doi/10.1103/PhysRevLett.80.1556.

Brackley, C. A., J. Allan, et al. (2014). "Topological constraints strongly affect chromatin reconstitution in silico". In: *Nucleic Acids Research*, gku1085–. URL: http://doi.org/10.1093/nar/gku1085.

Brackley, C. A., M. E. Cates, and D. Marenduzzo (2012). "Facilitated diffusion on mobile DNA: Configurational traps and sequence heterogeneity". In: *Physical Review Letters* 109.16, p. 168103. URL: https://doi.org/10.1103/PhysRevLett.109.168103.

Brackley, C A, A N Morozov, and D Marenduzzo (2014). "Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS." In: *The Journal of chemical physics* 140.13, p. 135103. URL: https://doi.org/10.1063/1.4870088.

Brackley, Chris A., Mike E. Cates, and Davide Marenduzzo (2013). "Effect of DNA conformation on facilitated diffusion". In: *Biochemical Society Transactions* 41.2, pp. 582–588. URL: http://doi.org/10.1042/BST20120234.

Brackley, Chris a, Stephen Taylor, et al. (2013). "Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization". In: *Proceedings of the National Academy of Sciences* 110.38, E3605–E3611. URL: https://doi.org/10.1073/pnas.1302950110.

Brooks, B. R. et al. (2009). "CHARMM: The biomolecular simulation program". In: *Journal of Computational Chemistry* 30.10, pp. 1545–1614. URL: http://doi.wiley.com/10.1002/jcc.21287.

Brutzer, Hergen et al. (2010). "Energetics at the DNA supercoiling transition." In: *Biophysical journal* 98.7, pp. 1267–76. URL: http://doi.org/10.1016/j.bpj.2009.12.4292.

Bryant, Zev, Florian C. Oberstrass, and Aakash Basu (2012). *Recent developments in single-molecule DNA mechanics*. URL: https://www.sciencedirect.com/science/article/pii/S0959440X12000735?via%3Dihub.

Bryant, Zev, Michael D. Stone, et al. (2003). "Structural transitions and elasticity from torque measurements on DNA". In: *Nature* 424.6946, pp. 338–341. URL: https://doi.org/10.1038/nature01810.

Bustamante, Carlos, Zev Bryant, and Steven B. Smith (2003). "Ten years of tension: Single-molecule DNA mechanics". In: *50 Years of DNA*. Vol. 421. Nature Publishing Group, pp. 423–427. URL: https://doi.org/10.1038/nature01405.

Carlon, Enrico, Enzo Orlandini, and Attilio L. Stella (2002). "Roles of Stiffness and Excluded Volume in DNA Denaturation". In: *Physical Review Letters* 88.19, p. 198101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.88.198101.

Chakraborty, Debayan, Naoto Hori, and D. Thirumalai (2018). "Sequence-Dependent Three Interaction Site Model for Single- and Double-Stranded DNA". In: *Journal of Chemical Theory and Computation* 14.7, pp. 3763–3779. URL: http://doi.org/10.1021/acs.jctc.8b00091.

Chargaff, Erwin, Rakoma Lipshitz, and Charlotte Green (1952). "Composition of the desoxypentose nucleic acids of four genera of sea-urchin". In: *Journal of Biological*

Chemistry 1, pp. 155–160. URL: http://www.jbc.org/content/195/1/155.citation.

Chatterjee, Gourab et al. (2017). "A spatially localized architecture for fast and modular DNA computing". In: *Nature Nanotechnology* 12.9, pp. 920–927. URL: http://www.nature.com/doifinder/10.1038/nnano.2017.127.

Cornell, Wendy D. et al. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *Journal of the American Chemical Society* 117.19, pp. 5179–5197. URL: http://pubs.acs.org/doi/abs/10.1021/ja00124a002.

Coronel, Lucia, Antonio Suma, and Cristian Micheletti (2018). "Dynamics of supercoiled DNA with complex knots: large-scale rearrangements and persistent multi-strand interlocking". In: *bioRxiv*, p. 331314. URL: https://www.biorxiv.org/content/early/2018/05/25/331314.

Creeth, J. M., J. Masson Gulland, and D. O. Jordan (1947). "214. Deoxypentose nucleic acids. Part III. Viscosity and streaming birefringence of solutions of the sodium salt of the deoxypentose nucleic acid of calf thymus". In: *Journal of the Chemical Society (Resumed)* 0.0, p. 1141. URL: http://xlink.rsc.org/?DOI=jr9470001141.

Daniels, Bryan C. and James P. Sethna (2011). "Nucleation at the DNA supercoiling transition". In: *Physical Review E* 83.4, p. 041924. URL: http://link.aps.org/doi/10.1103/PhysRevE.83.041924.

Dauxois, Thierry (1991). "Dynamics of breather modes in a nonlinear "helicoidal" model of DNA". In: *Physics Letters A* 159.8-9, pp. 390–395. URL: https://www.sciencedirect.com/science/article/pii/037596019190367H.

Davis, Natalie A., Sangita S. Majee, and Jason D. Kahn (1999). "TATA box DNA deformation with and without the TATA box-binding protein". In: *Journal of Molecular Biology* 291.2, pp. 249–265. URL: https://doi.org/10.1006/jmbi.1999.2947.

De Vlaminck, Iwijn and Cees Dekker (2012). "Recent Advances in Magnetic Tweezers". In: *Annual Review of Biophysics* 41.1, pp. 453–472. URL: http://doi.org/10.1146/annurev-biophys-122311-100544.

Deufel, Christopher et al. (2007). "Nanofabricated quartz cylinders for angular trapping: DNA supercoiling torque detection". In: *Nature Methods* 4.3, pp. 223–225. URL: http://www.nature.com/articles/nmeth1013.

Ding, Yue et al. (2014). "DNA supercoiling: A regulatory signal for the λ repressor". In: *Proceedings of the National Academy of Sciences* 111.43, pp. 15402–15407. URL: www.pnas.org/cgi/doi/10.1073/pnas.1320644111.

Dittmore, Andrew, Sumitabha Brahmachari, et al. (2017). "Supercoiling DNA locates mismatches". In: URL: https://arxiv.org/pdf/1704.07815.pdf.

Dittmore, Andrew and Keir C. Neuman (2018). "Evidence for a Solenoid Phase of Supercoiled DNA". In: *bioRxiv*, p. 302661. URL: https://www.biorxiv.org/content/early/2018/04/16/302661.

Doye, Jonathan P K et al. (2013). "Coarse-graining DNA for simulations of DNA nanotechnology." In: *Physical chemistry chemical physics : PCCP* 15.47, pp. 20395–414. URL: http://www.ncbi.nlm.nih.gov/pubmed/24121860.

Dršata, T et al. (2014). "Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning". In: *Nucleic Acids Research* 42.11, pp. 7383–7394. URL: http://doi.org/10.1093/nar/gku338.

Emanuel, Marc, Giovanni Lanzani, and Helmut Schiessel (2013). "Multiplectoneme phase of double-stranded DNA under torsion". In: *Physical Review E* 88.2, p. 022706. URL: http://link.aps.org/doi/10.1103/PhysRevE.88.022706.

Engel, Megan C., Dustin B. Ritchie, et al. (2014). "Reconstructing folding energy landscape profiles from nonequilibrium pulling curves with an inverse Weierstrass integral transform". In: *Physical Review Letters* 113.23, p. 238104. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.238104.

Engel, Megan C., David M. Smith, et al. (2018). "Force-Induced Unravelling of DNA Origami". In: *ACS Nano* 12.7, pp. 6734–6747. URL: http://pubs.acs.org/doi/10.1021/acsnano.8b01844.

Forget, Anthony L. and Stephen C. Kowalczykowski (2012). "Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search". In: *Nature* 482.7385, pp. 423–427. URL: http://www.nature.com/articles/nature10782.

Forties, Robert A and Michelle D Wang (2014). "Discovering the power of single molecules." In: *Cell* 157.1, pp. 4–7. URL: https://doi.org/10.1016/j.cell.2014.02.011.

Franklin, Rosalind E. and R. G. Gosling (1953). "Molecular Configuration in Sodium Thymonucleate". In: *Nature* 171.4356, pp. 740–741. URL: http://www.nature.com/doifinder/10.1038/171740a0.

Fye, Richard M. and Craig J. Benham (1999). "Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA". In: *Physical Review E* 59.3, pp. 3408–3426. URL: http://link.aps.org/doi/10.1103/PhysRevE.59.3408.

Galburt, E. A. et al. (2014). "Force-dependent melting of supercoiled DNA at thermophilic temperatures". In: *Biophysical Chemistry* 187-188, pp. 23–28. URL: http://doi.org/10.1016/j.bpc.2014.01.001.

Ganji, Mahipal et al. (2016). "Intercalation-based single-molecule fluorescence assay to study DNA supercoil dynamics". In: *Nano Letters* 16.7, pp. 4699–4707. URL: http://pubs.acs.org/doi/abs/10.1021/acs.nanolett.6b02213.

Gazzola, M. et al. (2018). "Forward and inverse problems in the mechanics of soft filaments". In: *Royal Society Open Science* 5.6, p. 171628. URL: http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.171628.

Geller, Ben et al. (2011). *No Title*. URL: http://umdberg.pbworks.com/w/page-revisions/47555271/DNA%20spring%201%3A%20Spring%20constant%20of%20DNA.

Ghosh, Anirban and Manju Bansal (2003). *A glossary of DNA structures from A to Z*. URL: http://scripts.iucr.org/cgi-bin/paper?S0907444903003251.

Ghoshdastidar, Debostuti and Manju Bansal (2018). "Dynamics of physiologically relevant noncanonical DNA structures: an overview from experimental and theoretical studies". In: *Briefings in Functional Genomics*. URL: http://doi.org/10.1093/bfgp/ely026.

Goyal, Sachin and N. C. Perkins (2008). "Looping mechanics of rods and DNA with non-homogeneous and discontinuous stiffness". In: *International Journal of Non-Linear Mechanics* 43.10, pp. 1121–1129. URL: https://www.sciencedirect.com/science/article/pii/S0020746208001340.

Gross, Peter et al. (2011). "Quantifying how DNA stretches, melts and changes twist under tension". In: *Nature Physics* 7.9, pp. 731–736. URL: http://www.nature.com/articles/nphys2002.

Ha, Sung Chul et al. (2005). "Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases". In: *Nature* 437.7062, pp. 1183–1186. URL: http://www.nature.com/articles/nature04088.

Hacker, William C., Shuxiang Li, and Adrian H. Elcock (2017). "NAR Breakthrough Article Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome". In: *Nucleic Acids Research* 45.13, pp. 7541–7554. URL: http://academic.oup.com/nar/article/45/13/7541/3875533.

Hamdan, Samir M. et al. (2009). "Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis". In: *Nature* 457.7227, pp. 336–339. URL: http://www.nature.com/doifinder/10.1038/nature07512.

Harris, Sarah A., Zara A. Sands, and Charles A. Laughton (2005). "Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA". In: *Biophysical Journal* 88.3, pp. 1684–1691. URL: https://www.sciencedirect.com/science/article/pii/S0006349505732355.

Henrich, Oliver et al. (2018). "Coarse-grained simulation of DNA using LAMMPS: An implementation of the oxDNA model and its applications". In: *European Physical Journal E* 41.5, p. 57. URL: http://doi.org/10.1140/epje/i2018-11669-8.

Holmes, Frederic (1998). "The DNA replication problem, 1953–1958". In: *Trends in Biochemical Sciences* 23.3, pp. 117–120. URL: https://doi.org/10.1016/S0968-0004(98)01190-6.

Hoogsteen, K. (1963). "The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine". In: *Acta Crystallographica* 16.9, pp. 907–916. URL: http://doi.org/10.1107/S0365110X63002437.

Howan, Kévin et al. (2012). "Initiation of transcription-coupled repair characterized at single-molecule resolution". In: *Nature* 490.7420, pp. 431–434. URL: http://www.nature.com/articles/nature11430.

Husain, I, J Griffith, and A Sancar (1988). "Thymine dimers bend DNA." In: *Proceedings of the National Academy of Sciences of the United States of America* 85.8, pp. 2558–62. URL: http://www.pnas.org/content/85/8/2558.

Husian, Intisar and Aziz Sancar (1987). "Binding of E. coli DNA photolyase to a defined substrate containing a single T< >T dimer". In: *Nucleic Acids Research* 15.3, pp. 1109–1120. URL: https://doi.org/10.1093/nar/15.3.1109.

Jain, Aklank et al. (2013). "DHX9 helicase is involved in preventing genomic instability induced by alternatively structured DNA in human cells". In: *Nucleic Acids Research* 41.22, pp. 10345–10357. URL: http://doi.org/10.1093/nar/gkt804.

Janissen, Richard et al. (2014). "Invincible DNA tethers: covalent DNA anchoring for enhanced temporal and force stability in magnetic tweezers experiments". In: *Nucleic Acids Research* 42.18, e137–e137. URL: http://academic.oup.com/nar/article/42/18/e137/2434509/Invincible-DNA-tethers-covalent-DNA-anchoring-for.

Janssen, Xander J. A. et al. (2012). "Electromagnetic torque tweezers: A versatile approach for measurement of single-molecule twist and torque". In: *Nano Letters* 12.7, pp. 3634–3639. URL: http://pubs.acs.org/doi/10.1021/nl301330h.

Jeltsch, Albert (2002). "Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases". In: *ChemBioChem* 3.4, pp. 274–293. URL: http://doi.wiley.com/10.1002/1439-7633%2820020402%293%3A4%3C274%3A%3AAID-CBIC274%3E3.0.CO%3B2-S.

Kabakçioğlu, A., E. Orlandini, and D. Mukamel (2009). "Supercoil formation in DNA denaturation". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 80.1, p. 010903. URL: https://link.aps.org/doi/10.1103/PhysRevE.80.010903.

Kafri, Yariv, David Mukamel, and Luca Peliti (2000). "Why is the DNA denaturation transition first order?" In: *Physical Review Letters* 85.23, pp. 4988–4991.

Kahn, Jason D. (2014). "DNA, Flexibly Flexible". In: *Biophysical Journal* 107.2, pp. 282–284. URL: https://doi.org/10.1016/j.bpj.2014.06.007.

Kamien, Randall D. (2002). "The geometry of soft materials: a primer". In: *Reviews of Modern Physics* 74.4, pp. 953–971. URL: http://doi.org/10.1103/RevModPhys.74.953.

Kanhere, Aditi and Manju Bansal (2005). "Structural properties of promoters: Similarities and differences between prokaryotes and eukaryotes". In: *Nucleic Acids Research* 33.10, pp. 3165–3175. URL: https://doi.org/10.1093/nar/gki627.

Kannan, Srinivasaraghavan and Martin Zacharias (2009). "Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations". In: *Physical Chemistry Chemical Physics* 11.45, p. 10589. URL: http://xlink.rsc.org/?DOI=b910792b.

Kapanidis, Achillefs N. and Terence Strick (2009). *Biology, one molecule at a time*. URL: https://doi.org/10.1016/j.tibs.2009.01.008.

Kavenoff, Ruth (1972). *Supercoiled Chromosome of E. Coli*. URL: https://www.nature.com/scitable/content/supercoiled-chromosome-of-e-coli-44517.

Kemmink, J., R. Boelens, T. Koning, G. A. van der Marel, et al. (1987). "H nmr study of the exchangeable protons of the duplex d(gcgttgcg).d(cgcaacgc)containing a thymine photodimer containing a thymine photodimer". In: *Nucleic Acids Research* 15.11, pp. 4645–4653. URL: http://doi.org/10.1093/nar/15.11.4645.

Kemmink, Johan, Rolf Boelens, Thea M.G. Koning, Robert Kaptein, et al. (1987). "Conformational changes in the oligonucleotide duples d(GCGTTGCG). d (CGCAACGC) induced by formation of a cis-syn thymine dimer: A two-dimensional NMR study". In: *European Journal of Biochemistry* 162.1, pp. 37–43. URL: http://doi.org/10.1111/j.1432-1033.1987.tb10538.x.

Kim, Jong-Ki -K, Dinshaw Patel, and Byong-Seok -S Choi (1995). "Contrasting structural impacts induced by cis-syn cyclobutane and (6-4) adduct in DNA duplex decamers: implication in mutagenesis repair activity". In: *Photochemistry and Photobiology* 62.1, pp. 44–50. URL: http://doi.wiley.com/10.1111/j.1751-1097.1995.tb05236.x.

Kim, Sung Hyun, Mahipal Ganji, et al. (2017). "DNA sequence encodes the position of DNA supercoils". In: *bioRxiv* 112.3, 214a. URL: http://doi.org/10.1101/180414.

King, Graeme A et al. (2013). "Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching using fluorescence microscopy". In: *Proceedings of the National Academy of Sciences* 110.10, pp. 3859–3864. URL: https://doi.org/10.1073/pnas.1213676110.

Knips, Alexander and Martin Zacharias (2015). "Influence of a cis,syn-cyclobutane pyrimidine dimer damage on DNA conformation studied by molecular dynamics simulations". In: *Biopolymers* 103.4, pp. 215–222. URL: http://doi.wiley.com/10.1002/bip.22586.

— (2017). "Both DNA global deformation and repair enzyme contacts mediate flipping of thymine dimer damage". In: *Scientific Reports* 7, p. 41324. URL: http://www.nature.com/articles/srep41324.

Knotts, Thomas A et al. (2007). "A coarse grain model for DNA." In: *The Journal of chemical physics* 126.8, p. 084901. URL: https://doi.org/10.1063/1.2431804.

Kossel, Albrecht (1891). "Ueber die chemische zusammensetzung der zelle". In: *Arch. Physiol* 181-186.

Koster, Daniel A. et al. (2010). "Cellular Strategies for Regulating DNA Supercoiling: A Single-Molecule Perspective". In: *Cell* 142.4, pp. 519–530. URL: https://www.sciencedirect.com/science/article/pii/S0092867410008962?via%3Dihub.

Kratky, O. and G. Porod (1949). "Röntgenuntersuchung gelöster Fadenmoleküle". In: *Recueil des Travaux Chimiques des Pays-Bas* 68.12, pp. 1106–1122. URL: http://doi.wiley.com/10.1002/recl.19490681203.

Kriegel, Franziska, Niklas Ermann, Ruaridh Forbes, et al. (2017). "Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque

tweezers". In: *Nucleic Acids Research* 45.10, pp. 5920–5929. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx280.

Kriegel, Franziska, Niklas Ermann, and Jan Lipfert (2017). "Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers". In: *Journal of Structural Biology* 197.1, pp. 26–36. URL: http://www.sciencedirect.com/science/article/pii/S104784771630137X.

Kriegel, Franziska, Christian Matek, et al. (2018). "The temperature dependence of the helical twist of DNA". In: *Nucleic Acids Research*. URL: https://doi.org/10.1093/nar/gky599.

Lapham, Jon et al. (1997). "Measurement of diffusion constants for nucleic acids by NMR". In: *Journal of Biomolecular NMR* 10.3, pp. 255–262. URL: http://link.springer.com/10.1023/A:1018310702909.

Lavery, R., K. Zakrzewska, and H. Sklenar (1995). "JUMNA (junction minimisation of nucleic acids)". In: *Computer Physics Communications* 91.1-3, pp. 135–158. URL: https://doi.org/10.1016/0010-4655(95)00046-I.

Lavery, Richard, Krystyna Zakrzewska, David Beveridge, et al. (2010). "A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA". In: *Nucleic Acids Research* 38.1, pp. 299–313. URL: https://doi.org/10.1093/nar/gkp834.

Léger, J. F. et al. (1999). "Structural transitions of a twisted and stretched DNA molecule". In: *Physical Review Letters* 83.5, pp. 1066–1069. URL: http://doi.org/10.1103/PhysRevLett.83.1066.

Levene, P a (1919). "The structure of yeast nucleic acid. IV. ammonia hydrolysis". In: *J. Biol. Chem.* 40, pp. 415–424.

Lipfert, Jan, Jacob W J Kerssemakers, et al. (2010). "Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments." In: *Nature methods* 7.12, pp. 977–80. URL: http://dx.doi.org/10.1038/nmeth.1520.

Lipfert, Jan, Matthew Wiggin, et al. (2011). "Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids". In: *Nature Communications* 2.1, p. 439. URL: http://www.nature.com/articles/ncomms1450.

Liu, Zheyun, Lijuan Wang, and Dongping Zhong (2015). "Dynamics and mechanisms of DNA repair by photolyase". In: *Physical Chemistry Chemical Physics* 17.18, pp. 11933–11949. URL: http://xlink.rsc.org/?DOI=C4CP05286B.

Loenhout, M T J van, M V de Grunt, and C Dekker (2012). "Dynamics of DNA supercoils." In: *Science (New York, N.Y.)* 338.6103, pp. 94–7. URL: http://www.sciencemag.org/content/338/6103/94.

Lonie, Iain M (2011). *The Hippocratic Treatises "On Generation", On the Nature of the Child, "Diseases IV": A Commentary*. Ed. by G Baader et al. Ars Medica/Abteilung 2, Griechisch-lateinische Medizin. De Gruyter, p. 446. URL: https://books.google.co.uk/books?id=G6WI7ejsWHoC.

Lucas-Lledó, José Ignacio and Michael Lynch (2009). "Evolution of mutation rates: Phylogenomic analysis of the photolyase/cryptochrome family". In: *Molecular Biology and Evolution* 26.5, pp. 1143–1153. URL: https://doi.org/10.1093/molbev/msp029.

Ludwig, Michael Z. (2002). "Functional evolution of noncoding DNA". In: *Current Opinion in Genetics & Development* 12.6, pp. 634–639. URL: https://www.sciencedirect.com/science/article/pii/S0959437X02003556?via%3Dihub.

Ma, Jie and Michelle D Wang (2016). *DNA supercoiling during transcription*. URL: https://dx.doi.org/10.1007%2Fs12551-016-0215-9.

Machinek, Robert R F et al. (2014). "Programmable energy landscapes for kinetic control of DNA strand displacement." en. In: *Nature communications* 5, p. 5324. URL: http://www.nature.com/ncomms/2014/141110/ncomms6324/abs/ncomms6324.html.

Maffeo, C et al. (2014). "Close encounters with DNA." In: *Journal of physics. Condensed matter : an Institute of Physics journal* 26.41, p. 413101. URL: http://stacks.iop.org/0953-8984/26/i=41/a=413101.

Mahroos, Mona Al et al. (2002). "Effect of Sunscreen Application on UV-Induced Thymine Dimers". In: *Archives of Dermatology* 138.11, pp. 1480–1485. URL: http://doi.org/10.1001/archderm.138.11.1480.

Manosas, Maria et al. (2012). "Collaborative coupling between polymerase and helicase for leading-strand synthesis". In: *Nucleic Acids Research* 40.13, pp. 6187–6198. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks254.

Marantan, Andrew and L. Mahadevan (2018). "Mechanics and statistics of the worm-like chain". In: *American Journal of Physics* 86.2, pp. 86–94. URL: http://aapt.scitation.org/doi/10.1119/1.5003376.

Marko, J. F (1997). "Stretching must twist DNA". In: *Europhysics Letters* 38.3, pp. 183–188. URL: http://doi.org/10.1209/epl/i1997-00223-5.

Marko, J. and E. Siggia (1995). "Statistical mechanics of supercoiled DNA". In: *Physical Review E* 52.3, pp. 2912–2938. URL: http://doi.org/10.1103/PhysRevE.52.2912.

Marko, John F. (2007). "Torque and dynamics of linking number relaxation in stretched supercoiled DNA". In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 76.2, p. 021926. URL: https://doi.org/10.1103/PhysRevE.76.021926.

Marko, John F. and Sébastien Neukirch (2012). "Competition between curls and plectonemes near the buckling transition of stretched supercoiled DNA". In: *Physical Review E* 85.1, p. 011908. URL: http://doi.org/10.1103/PhysRevE.85.011908.

— (2013). "Global force-torque phase diagram for the DNA double helix: Structural transitions, triple points, and collapsed plectonemes". In: *Physical Review E* 88.6, p. 062722. URL: http://doi.org/10.1103/PhysRevE.88.062722.

Marko, John F. and Eric D. Siggia (1994). "Fluctuations and supercoiling of DNA". In: *Science* 265.5171, pp. 506–508. URL: http://doi.org/10.1126/science.8036491.

Marky, Luis A. and Kenneth J. Breslauer (1987). "Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves". In: *Biopolymers* 26.9, pp. 1601–1620. URL: http://doi.wiley.com/10.1002/bip.360260911.

Matek, Christian (2014). "Statistical Mechanics of Nucleic Acids under Mechanical Stress". PhD thesis. University of Oxford. URL: https://ora.ox.ac.uk/objects/uuid:ce44cf50-2001-4f54-8e57-d1757f709fd6.

Matek, Christian et al. (2015). "Plectoneme tip bubbles: coupled denaturation and writhing in supercoiled DNA." en. In: *Scientific reports* 5, p. 7655. URL: http://doi.org/10.1038/srep07655.

McAteer, K et al. (1998). "Solution-state structure of a DNA dodecamer duplex containing a Cis-syn thymine cyclobutane dimer, the major UV photoproduct of DNA." In: *Journal of molecular biology* 282.5, pp. 1013–32. URL: http://www.sciencedirect.com/science/article/pii/S0022283698920629.

Mees, Alexandra et al. (2004). "Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair." In: *Science (New York, N.Y.)* 306.5702, pp. 1789–93. URL: http://www.ncbi.nlm.nih.gov/pubmed/15576622.

Mendel, Gregor (1865). "Versuche über Pflanzen-Hybride". In: *Verhandlungen des Naturforschenden Vereines in Brünn* 4, pp. 3–47.

Meng, He et al. (2014). "Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains". In: *Biophysical Journal* 106.5, pp. 1174–1181. URL: https://doi.org/10.1016/j.bpj.2014.01.017.

Metzler, Ralf et al. (2009). "Single DNA denaturation and bubble dynamics". In: *Journal of Physics: Condensed Matter* 21.3, p. 034111. URL: https://doi.org/10.1088/0953-8984/21/3/034111.

Miaskiewicz, Karol et al. (1996). "Computational simulations of DNA distortions by a cis,syn-cyclobutane thymine dimer lesion". In: *Journal of the American Chemical Society* 118.38, pp. 9156–9163. URL: http://doi.org/10.1021/ja9611304.

Miescher, Friedrich (1897). *Die histochemischen und physiologischen Arbeiten von Friedrich Miescher*. Vol. 2. Vogel.

Milstein, J N and J.- C. Meiners (2011). "On the role of DNA biomechanics in the regulation of gene expression". In: *Journal of The Royal Society Interface* 8.65, pp. 1673–1681. URL: http://doi.org/10.1098/rsif.2011.0371.

Mirny, Leonid et al. (2009). "How a protein searches for its site on DNA: The mechanism of facilitated diffusion". In: *Journal of Physics A: Mathematical and Theoretical* 42.43, p. 434013. URL: https://doi.org/10.1088/1751-8113/42/43/434013.

Modrich, Paul (2016). "Mechanisms in E . coli and Human Mismatch Repair (Nobel Lecture)". In: pp. 8490–8501. URL: https://www.nobelprize.org/prizes/chemistry/2015/modrich/lecture/.

Morita, Rihito et al. (2010). "Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems." In: *Journal of nucleic acids* 2010, p. 179594.

Moroz, J David and Philip Nelson (1997). "Torsional directed walks, entropic elasticity, and DNA twist stiffness". In: *Proceedings of the National Academy of Sciences* 94.26, pp. 14418–14422. URL: http://www.pnas.org/content/94/26/14418.abstract.

Mosconi, Francesco et al. (2009). "Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers". In: *Physical Review Letters* 102.7, pp. 1–4.

Mossa, A et al. (2009). "Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes". In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.2, P02060. URL: https://doi.org/10.1088/1742-5468/2009/02/P02060.

Neidle, Stephen. (2007). *Principles of nucleic acid structure*. Academic press, p. 302.

Neukirch, Sébastien and John F. Marko (2011). "Analytical description of extension, torque, and supercoiling radius of a stretched twisted DNA". In: *Physical Review Letters* 106.13, p. 138104. URL: http://doi.org/10.1103/PhysRevLett.106.138104.

Ngo, Thuy T. M. et al. (2016). "Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability". In: *Nature Communications* 7.1, p. 10813. URL: http://www.nature.com/articles/ncomms10813.

Nguyen, Khanh Thuy et al. (2015). "Self-assembly of mesogenic bent-core DNA nanoduplexes". In: *Soft Matter* 11.15, pp. 2934–2944. URL: http://xlink.rsc.org/?DOI=C4SM01571A.

Noble, Denis (2013). "Physiology is rocking the foundations of evolutionary biology". In: *Experimental Physiology* 98.8, pp. 1235–1243.

Nöllmann, Marcelo et al. (2007). "Multiple modes of Escherichia coli DNA gyrase activity revealed by force and torque". In: *Nature Structural & Molecular Biology* 14.4, pp. 264–271. URL: http://www.nature.com/articles/nsmb1213.

Noy, Agnes and Ramin Golestanian (2010). "The Chirality of DNA: Elasticity Cross-Terms at Base-Pair Level Including A-Tracts and the Influence of Ionic

Strength". In: *The Journal of Physical Chemistry B* 114.23, pp. 8022–8031. URL: http://pubs.acs.org/doi/abs/10.1021/jp104133j.

Noy, Agnes and Ramin Golestanian (2012). "Length scale dependence of DNA mechanical properties". In: *Physical Review Letters* 109.22, pp. 1–5.

Oberstrass, F. C., L. E. Fernandes, and Z. Bryant (2012). "Torque measurements reveal sequence-specific cooperative transitions in supercoiled DNA". In: *Proceedings of the National Academy of Sciences* 109.16, pp. 6106–6111. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1113532109.

Odijk, Theo (1995). "Stiff Chains and Filaments under Tension". In: *Macromolecules* 28.20, pp. 7016–7018. URL: http://pubs.acs.org/doi/abs/10.1021/ma00124a044.

Oijen, Antoine M. van and Joseph J. Loparo (2010). "Single-Molecule Studies of the Replisome". In: *Annual Review of Biophysics* 39.1, pp. 429–448. URL: http://doi.org/10.1146/annurev.biophys.093008.131327.

Ouldridge, Thomas E. (2011). "Coarse-Grained Modelling of DNA and DNA Self-Assembly". PhD thesis. University Of Oxford. URL: https://ora.ox.ac.uk/objects/uuid:b2415bb2-7975-4f59-b5e2-8c022b4a3719.

— (2014). "DNA nanotechnology: understanding and optimisation through simulation". en. In: *Molecular Physics* 113.1, pp. 1–15. URL: http://www.tandfonline.com/doi/abs/10.1080/00268976.2014.975293.

Ouldridge, Thomas E, Ard A Louis, and Jonathan P K Doye (2010). "DNA Nanotweezers Studied with a Coarse-Grained Model of DNA". In: *Physical Review Letters* 104.17, p. 178101. URL: http://link.aps.org/doi/10.1103/PhysRevLett.104.178101.

— (2011). "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model." In: *The Journal of chemical physics* 134.8, p. 085101. URL: http://www.ncbi.nlm.nih.gov/pubmed/21361556.

Ouldridge, Thomas E, Ard A Louis, Jonathan P K Doye, et al. (2014). *oxDNA webpage*. URL: https://dna.physics.ox.ac.uk/index.php/Main_Page.

Park, C J, J H Lee, and B S Choi (2007). "Functional insights gained from structural analyses of DNA duplexes that contain UV-damaged photoproducts". In: *Photochemistry and photobiology* 83.1, pp. 187–195. URL: http://doi.wiley.com/10.1562/2006-02-28-IR-820.

Park, HaJeung, Kaijiang Zhang, et al. (2002). "Crystal structure of a DNA decamer containing a cis-syn thymine dimer". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.25, pp. 15965–70. URL: http://www.pnas.org/content/99/25/15965.full.

Parris, Christopher N. and Michael M. Seidman (1992). "A signature element distinguishes sibling and independent mutations in a shuttle vector plasmid". In: *Gene* 117.1, pp. 1–5. URL: https://doi.org/10.1016/0378-1119(92)90482-5.

Pauling, Linus and Robert B Corey (1952). "A proposed structure for the nucleic acids". In: *Nature* 39.2, pp. 1481–1486. URL: https://doi.org/10.1073/pnas.39.2.84.

Pearlman, D A et al. (1985). "'" In: *Science* 227.4692, pp. 1304–8. URL: http://doi.org/10.1126/science.3975615.

Pennisi, Elizabeth (2012). "ENCODE Project Writes Eulogy for Junk DNA". In: *Science 337* 337.6099, pp. 1159–1161. URL: https://doi.org/10.1126/science.337.6099.1159.

Pérez, Alberto, F. Javier Luque, and Modesto Orozco (2012). "Frontiers in Molecular Dynamics Simulations of DNA". In: *Accounts of Chemical Research* 45.2, pp. 196–205. URL: http://pubs.acs.org/doi/10.1021/ar2001217.

Peyrard, M. and A. R. Bishop (1989). "Statistical mechanics of a nonlinear model for DNA denaturation". In: *Physical Review Letters* 62.23, pp. 2755–2758. URL: https://link.aps.org/doi/10.1103/PhysRevLett.62.2755.

Piana, S. (2005). "Structure and energy of a DNA dodecamer under tensile load". In: *Nucleic Acids Research* 33.22, pp. 7029–7038. URL: https://doi.org/10.1093/nar/gki1010.

Poland, Douglas and Harold A. Scheraga (1966). "Occurrence of a Phase Transition in Nucleic Acid Models". In: *The Journal of Chemical Physics* 45.5, pp. 1464–1469. URL: http://aip.scitation.org/doi/10.1063/1.1727786.

Randall, Graham L., Lynn Zechiedrich, and B. Montgomery Pettitt (2009). "In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form". In: *Nucleic Acids Research* 37.16, pp. 5568–5577. URL: https://doi.org/10.1093/nar/gkp556.

Rao, Shashidhar N., Joe W. Keepers, and Peter Koliman (1984). "The structure of d(CGCGAAT[ ]TCGCG)-d(CGCGAATTCGCG): The incorporation of a thymine photodimer into a B-DNA helix". In: *Nucleic Acids Research* 12.11, pp. 4789–4807. URL: https://www.doi.org/10.1093/nar/12.11.4789.

Riggs, Arthur D., Suzanne Bourgeois, and Melvin Cohn (1970). "The lac represser-operator interaction: III. Kinetic studies". In: *Journal of Molecular Biology* 53.3, pp. 401–417. URL: https://www.sciencedirect.com/science/article/pii/0022283670900744?via%3Dihub.

Romano, Flavio et al. (2013). "Coarse-grained simulations of DNA overstretching". In: *The Journal of Chemical Physics* 138.8, p. 085101. URL: http://aip.scitation.org/doi/10.1063/1.4792252.

Rovigatti, Lorenzo et al. (2015). "A comparison between parallelization approaches in molecular dynamics simulations on GPUs". In: *Journal of Computational Chemistry* 36.1, pp. 1–8. URL: http://doi.wiley.com/10.1002/jcc.23763.

Rumora, Amy (2008). "Exploring DNA destabilization induced by the thymine dimer lesion using base". PhD thesis. Mount Holyoke College. URL: https://ida.mtholyoke.edu/xmlui/handle/10166/610.

Rumora, Amy E et al. (2008). "Thymine dimer-induced structural changes to the DNA duplex examined with reactive probes (†)." In: *Biochemistry* 47.49, pp. 13026–35. URL: http://dx.doi.org/10.1021/bi801417u.

Russo, John, Piero Tartaglia, and Francesco Sciortino (2009). "Reversible gels of patchy particles: Role of the valence". In: *The Journal of Chemical Physics* 131.1, p. 014504. URL: http://aip.scitation.org/doi/10.1063/1.3153843.

Saiz, Leonor and Jose MG Vilar (2006). *DNA looping: the consequences and its control.* URL: https://doi.org/10.1016/j.sbi.2006.05.008.

Salerno, D et al. (2012). "Single-molecule study of the DNA denaturation phase transition in the force-torsion space." In: *Physical review letters* 109.11, p. 118303. URL: http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.109.118303.

Sancar, Aziz (2003). "Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors". In: *Chemical Reviews* 103.6, pp. 2203–2238. URL: https://pubs.acs.org/doi/abs/10.1021/cr0204348.

— (2016). *Mechanisms of DNA Repair by Photolyase and Excision Nuclease (Nobel Lecture).* URL: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2015/sancar-lecture.html.

SantaLucia, John and Donald Hicks (2004). "THE THERMODYNAMICS OF DNA STRUCTURAL MOTIFS". In: *Annual Review of Biophysics and Biomolecular*

Structure 33.1, pp. 415–440. URL: https://doi.org/10.1146/annurev.biophys.32.110601.141800.

Schelvis, Johannes P. M., Xuling Zhu, and Yvonne M. Gindt (2015). "Enzyme-Substrate Binding Kinetics Indicate That Photolyase Recognizes an Extrahelical Cyclobutane Thymidine Dimer". In: *Biochemistry* 54.40, pp. 6176–6185. URL: http://pubs.acs.org/doi/10.1021/acs.biochem.5b00927.

Schiessel, Helmut (2014). *Biophysics for beginners*. Singapore: Panstanford publishing, p. 407.

Schrödinger, Erwin (1944). *What is Life?* Canto. Cambridge University Press.

Seidman, Michael M. et al. (1985). "A shuttle vector plasmid for studying carcinogen-induced point mutations in mammalian cells". In: *Gene* 38.1-3, pp. 233–237. URL: https://doi.org/10.1016/0378-1119(85)90222-7.

Severin, Philip M.D. et al. (2011). "Cytosine methylation alters DNA mechanical properties". In: *Nucleic Acids Research* 39.20, pp. 8740–8751. URL: www.doi.org/10.1093/nar/gkr578.

Sheinin, Maxim Y, Scott Forth, et al. (2011). "Underwound DNA under tension: structure, elasticity, and sequence-dependent behaviors". In: *Physical review letters* 107.10, p. 108102. URL: http://doi.org/10.1103/PhysRevLett.107.108102.

Sheinin, Maxim Y. and Michelle D. Wang (2009). "Twist-stretch coupling and phase transition during DNA supercoiling". In: *Physical Chemistry Chemical Physics* 11.24, pp. 4800–4803. URL: http://xlink.rsc.org/?DOI=b901646e.

Shpigelman, E. S., E. N. Trifonov, and A. Bolshoy (1993). "Curvature: Software for the analysis of curved DNA". In: *Bioinformatics* 9.4, pp. 435–440. URL: https://doi.org/10.1093/bioinformatics/9.4.435.

Sinden, Richard R (1994). *DNA structure and function*. Academic Press.

Skoruppa, Enrico et al. (2017). "DNA elasticity from coarse-grained simulations: The effect of groove asymmetry". In: *The Journal of Chemical Physics* 146.21, p. 214902. URL: http://aip.scitation.org/doi/10.1063/1.4984039.

Smith, S B, Y Cui, and C Bustamante (1996). "Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules." In: *Science (New York, N.Y.)* 271.5250, pp. 795–9. URL: http://www.ncbi.nlm.nih.gov/pubmed/8628994.

Snodin, Benedict E K (2015). "Simulating large DNA nanostructures with a coarse-grained model". In: URL: https://ora.ox.ac.uk/objects/uuid:86b2d03e-3cad-423c-8db0-3d56c65ccc20.

Snodin, Benedict E K et al. (2015). "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA." In: *The Journal of chemical physics* 142.23, p. 234901. URL: http://doi.org/10.1063/1.4921957.

Strick, T R, J F Allemand, D Bensimon, A Bensimon, et al. (1996). "The elasticity of a single supercoiled DNA molecule." In: *Science (New York, N.Y.)* 271.5257, pp. 1835–7. URL: http://www.ncbi.nlm.nih.gov/pubmed/8596951.

Strick, T R, J F Allemand, D Bensimon, and V Croquette (1998). "Behavior of supercoiled DNA". In: *Biophysical Journal* 74.4, pp. 2016–2028. URL: https://doi.org/10.1016/S0006-3495(98)77908-1.

Strick, T R, M-N Dessinges, et al. (2003). "Stretching of macromolecules and proteins". In: *Reports on Progress in Physics* 66.1, pp. 1–45. URL: http://stacks.iop.org/0034-4885/66/i=1/a=201.

Strick, Terence R., Vincent Croquette, and David Bensimon (2000). "Single-molecule analysis of DNA uncoiling by a type II topoisomerase". In: *Nature* 404.6780, pp. 901–904. URL: http://www.nature.com/articles/35009144.

Šulc, Petr et al. (2012). "Sequence-dependent thermodynamics of a coarse-grained DNA model." In: *The Journal of chemical physics* 137.13, p. 135101. URL: http://www.ncbi.nlm.nih.gov/pubmed/23039613.

Sutthibutpong, Thana et al. (2016). "Long-range correlations in the mechanics of small DNA circles under topological stress revealed by multi-scale simulation". In: *Nucleic Acids Research* 44.19, pp. 9121–9130. URL: https://doi.org/10.1093/nar/gkw815.

Swope, William C. et al. (1982). "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". In: *The Journal of Chemical Physics* 76.1, pp. 637–649. URL: http://aip.scitation.org/doi/10.1063/1.442716.

Taylor, John Stephen et al. (1990). "1HNMR Assignment and Melting Temperature Study of Cis-Syn and Trans-Syn Thymine Dimer Containing Duplexes of d(CGTATTATGC) · d(GCATAATACG)". In: *Biochemistry* 29.37, pp. 8858–8866. URL: http://pubs.acs.org/doi/abs/10.1021/bi00489a049.

Tempestini, Alessia et al. (2013). "Magnetic tweezers measurements of the nanomechanical stability of DNA against denaturation at various conditions of pH and ionic strength." In: *Nucleic acids research* 41.3, pp. 2009–19. URL: http://nar.oxfordjournals.org/cgi/content/long/41/3/2009.

Thompson, J. M. T. and A. R. Champneys (1996). "From Helix to Localized Writhing in the Torsional Post-Buckling of Elastic Rods". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 452.1944, pp. 117–138. URL: http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1996.0007.

Valenti, Anna et al. (2011). "Positive supercoiling in thermophiles and mesophiles: of the good and evil". In: *Biochemical Society Transactions* 39.1, pp. 58–63. URL: http://www.doi.org/10.1042/BST0390058.

Vlijm, Rifka et al. (2015). "Experimental phase diagram of negatively supercoiled DNA measured by magnetic tweezers and fluorescence". en. In: *Nanoscale* 7.7, pp. 3205–16. URL: http://pubs.rsc.org/en/content/articlehtml/2015/nr/c4nr04332d.

Wang, C I and J S Taylor (1991). "Site-specific effect of thymine dimer formation on dAn.dTn tract bending and its biological implications." In: *Proceedings of the National Academy of Sciences* 88.20, pp. 9072–9076. URL: http://www.doi.org/10.1073/pnas.88.20.9072.

Wang, M. D., H. Yin, et al. (1997). "Stretching DNA with optical tweezers". In: *Biophysical Journal* 72.3, pp. 1335–1346. URL: https://doi.org/10.1016/S0006-3495(97)78780-0.

Watson, J D and F H C Crick (1953). "Molecular structure of nucleic acids". In: *Nature* 171, pp. 737–738. URL: http://www.nature.com/physics/looking-back/crick/.

Weil, R and J Vinograd (1963). "The Cyclic Helix and Cyclic Coil Forms of Polyoma Viral DNA". In: *Proc Natl Acad Sci U S A* 50.4, pp. 730–738. URL: http://doi.org/10.1073/pnas.50.4.730.

Wenke, Belinda B. et al. (2013). "Base pair opening in a deoxynucleotide duplex containing a cis-syn thymine cyclobutane dimer lesion". In: *Biochemistry* 52.51, pp. 9275–9285. URL: http://pubs.acs.org/doi/abs/10.1021/bi401312r.

Wereszczynski, Jeff and Ioan Andricioaei (2006). "On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension". In: *Proceedings of the National Academy of Sciences* 103.44, pp. 16200–16205. URL: http://doi.org/10.1073/pnas.0603850103.

Wilson, Thomas J. et al. (2011). "Evidence from thermodynamics that DNA photolyase recognizes a solvent-exposed CPD lesion". In: *Journal of Physical Chemistry B* 115.46, pp. 13746–13754. URL: http://doi.org/10.1021/jp208129a.

Yao, Jiaqin, Kathleen Dixon, and Michael P. Carty (2001). "A single (6-4) photoproduct inhibits plasmid DNA replication in xeroderma pigmentosum variant cell extracts". In: *Environmental and Molecular Mutagenesis* 38.1, pp. 19–29. URL: http://doi.wiley.com/10.1002/em.1046.

Yoo, Jejoong et al. (2017). "A Practical Guide to Molecular Dynamics Simulations of DNA Origami Systems". In: *Protocols in DNA Nanotechology*. Ed. by Giampaolo Zuccheri. 2nd ed. Methods in Molecular Biology. Humana Press. Chap. A Practica.

Zadegan, Reza M. et al. (2017). "Twisting of DNA Origami from Intercalators". In: *Scientific Reports* 7.1, p. 7382. URL: https://doi.org/10.1038/s41598-017-07796-3.

Zhabinskaya, Dina and Craig J. Benham (2011). "Theoretical Analysis of the Stress Induced B-Z Transition in Superhelical DNA". In: *PLoS Computational Biology* 7.1. Ed. by Nikolay V. Dokholyan, e1001051. URL: https://doi.org/10.1371/journal.pcbi.1001051.

Zhu, You-Liang, Zhong-Yuan Lu, and Zhao-Yan Sun (2016). "The mechanism of the emergence of distinct overstretched DNA states". In: *The Journal of Chemical Physics* 144.2, p. 024901. URL: http://aip.scitation.org/doi/10.1063/1.4939623.