



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

A double-stranded coarse grained model for DNA: applications to supercoiling and denaturation

Yair Augusto Gutiérrez Fosado



Doctor of Philosophy
The University of Edinburgh
January, 2019

Abstract

DNA supercoiling is the name given to the under or overwinding of the two strands of a DNA double helix. It is of great interest because it is relevant in several crucial biological processes. However, the principles governing its dynamics and its precise role under different circumstances remain elusive. Despite advances in single molecule experimental techniques, measuring supercoiling dynamics persist a challenge; this is where computer simulations are useful. In this thesis, I first introduce a single-nucleotide resolution coarse-grained computational model of DNA, that faithfully reproduces the geometry of the double-stranded helix and also part of its elastic behaviour. The dynamic of the system is implemented using a molecular dynamics scheme, and the results obtained are interpreted through methods of equilibrium and non-equilibrium statistical mechanics.

I then employ this model to specifically study DNA supercoiling. This phenomenon, although topological in nature, is extremely important for the survival of cells because it has a deep impact on the regulation of gene expression, the compaction of DNA inside the cell and DNA replication. In particular, this work finds its motivations in: (i) an experimentally unresolved problem about the effect of supercoiling on DNA melting; (ii) the dynamics of supercoiling under physiological conditions during transcription; and (iii) the relation between supercoiling and DNA-binding proteins. Given that these phenomena may be relevant *in vivo*, they have recently received a great deal of attention. However, until now, no computational model existed to study these kind of process.

The techniques used here have been successful in providing insight into the key elements in the system. This would have been impossible before by using, for example, 1D models. The major achievement of this work is the quantitative characterisation of the role played by DNA supercoiling in a range of situations that are commonly found *in vivo*.

Lay Summary

The discovery of the helical structure of DNA in 1953 revealed the mechanism by which genetic information - with the instructions to accomplish the biological function of all living organisms - is stored. However, its peculiar helical shape would also raise several questions about the organization of DNA inside a cell and its functioning. The problem is easily seen if we think about an old fashioned telephone cord, which has been assembled with a spiral shape similar to that of DNA. Any attempt to over or under twist the cord will cause it to writhe up, a phenomenon known as supercoiling. In the case of DNA, the regulation of supercoiling is important because it permits (or prevents) access to the information encoded along its backbone. To complicate the picture even more, supercoiling takes place in the intracellular environment where DNA interacts with other elements, such as proteins, that continuously change its shape. In this thesis I study the role played by DNA supercoiling in circumstances inspired by biologically relevant scenarios, but still simple enough to obtain insight using computer simulations and analytical theories.

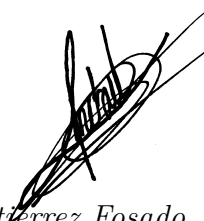
Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in:

- Y A G Fosado, D. Michieletto, J. Allan, C. A. Brackley, O. Henrich, and D. Marenduzzo. A single nucleotide resolution model for large-scale simulations of double stranded DNA. *Soft Matter*, 12:9458–9470, 2016.
- Y A G Fosado, D Michieletto, and D Marenduzzo. Dynamical Scaling and Phase Coexistence in Topologically Constrained DNA Melting. *Phys. Rev. Lett.*, 119:118002, Sep 2017.
- Oliver Henrich, Yair Augusto Gutiérrez Fosado, Tine Curk, and Thomas E. Ouldridge. Coarse-grained simulation of DNA using LAMMPS. *The European Physical Journal E*, 41(5):57, May 2018.

The goals and physical ideas behind the first two papers have been inspired by my supervisor Prof. D. Marenduzzo. The third paper was a collaboration with Dr. O. Henrich. Throughout all the projects I have run the simulations, analysed the data and contributed to the writing. The expertise of Dr. D. Michieletto was key to develop the mean field theory and simulations of the PRL manuscript. I also acknowledge the contributions of Dr. C. A. Brackley in the development of the Coarse-Grained model.



(Yair Augusto Gutiérrez Fosado, January, 2019)

Acknowledgements

The thesis you are holding in your hands would not be possible without the contribution of numerous people. First, I would like to sincerely express my gratitude to my supervisor, Prof. Davide Marenduzzo, who introduced me to such an interesting subject as is the computational study of DNA; and whom I could approach every time I needed help and support. You have set an example of excellence as a researcher, mentor and role model.

I am extremely grateful to Davide Michieletto for his patience and willingness to help in every aspect of my career during my PhD. Without his direction and constant feedback this PhD would not have been achievable. I truly appreciate the assistance, advice and professional guidance provided by Chris Brackley. Thanks also to both of you for the time you generously spent in order to improve this thesis. With your constructive criticism and valuable suggestions, you influenced in a positive way to the content of this manuscript.

I would like to thank Oliver Henrich for his constant enthusiasm, encouragement, guidance, fruitful discussions and more important for his friendship.

I gratefully acknowledge the Mexican National Council for Science and Technology (CONACyT) for the funding received (Ph.D. Grant No. 384582).

Many thanks also to all members of the LL galaxy family, who made my stay in Edinburgh really pleasant and whom I shared many nice experiences and emotions during the last four years. I feel honoured to be part of such group of wonderful and kind people.

I would like to thank the people who have always been there for me, my parents Flor and Benjamín, for their unconditional love, encouragement and unwavering emotional support. Last but not least, I deeply thank my beloved wife Doriscela, not only for her constant support and understanding, but also for being the light that illuminates my life. Thanks for choosing me as your partner to share this amazing journey in life.

Contents

Abstract	i
Lay Summary	iii
Declaration	v
Acknowledgements	vii
Contents	ix
List of Figures	xiii
List of Tables	xvii
List of Symbols	xix
1 Introduction	1
2 Background: physical properties and models of DNA	5
2.1 DNA structure	6
2.1.1 Multiple structures of the double helix	8
2.1.2 Denaturation of DNA	10
2.2 Elastic properties of DNA as a polymer	10
2.2.1 The Freely-Jointed-Chain model	11

2.2.2	The Worm-Like-Chain model	15
2.3	DNA supercoiling	22
2.3.1	DNA torque response	24
2.3.2	Differences between positive and negative supercoiling in DNA	28
2.4	Summary	28
3	A coarse-grained model for large-scale simulations of DNA	31
3.1	Molecular dynamics simulations	32
3.1.1	Brownian dynamics simulations	32
3.1.2	Coarse-Graining	34
3.2	The model	35
3.3	Parameterisation	41
3.3.1	Persistence length	42
3.3.2	Torsional persistence length	43
3.4	Validation through simulated single molecule experiments	44
3.4.1	Response to stretching	44
3.4.2	Response to twisting	46
3.5	Denaturation and supercoiling in the CG-DNA model	47
3.6	Future improvements and limitations	49
3.7	Summary	51
4	Dynamical scaling and phase coexistence in DNA melting	53
4.1	Melting curves	55
4.2	Phase diagram	58
4.2.1	Landau mean field theory	58
4.2.2	Spinodal region and binodal line	64
4.3	Dynamical scaling	68

4.4	Linking within denaturation bubbles	73
4.5	Summary	73
5	Transcription-Driven 3D Supercoiling Dynamics	75
5.1	3D Modelling of transcription	76
5.2	Twin supercoiling domains drive plectoneme formation at a distance.....	77
5.3	Nucleosomes can be unwrapped by transcription at a distance.....	81
5.4	Two modes of supercoiling relaxation.....	84
5.5	Transcription without topological constraints	87
5.6	Summary	88
6	DNA binding proteins	91
6.1	Protein model	92
6.2	Binding curve and the dynamics of the DNA-proteins system.....	95
6.3	Protein binding induces supercoiling.....	97
6.4	Summary	98
7	Conclusions	99
A	DNA model: grooves and closure procedure	101
B	Details on the calculations for the melting of tcDNA	107
C	Details on the transcription of model polymerase	119
	Bibliography	137

List of Figures

(1.1) Comparisson of supercoiling in a telephone cord and DNA micro-graph	2
(2.1) DNA assembly and structure	7
(2.2) FJC model sketch	12
(2.3) Force vs extension experiments	14
(2.4) Sign of writhe	22
(2.5) Supercoiling diagram	24
(2.6) Torque-response plot	25
(2.7) Extension vs supercoiling plot	27
(3.1) Coarse-grained DNA model	36
(3.2) Persistence length and Torsional persistence length	42
(3.3) Force-extension simulations	45
(3.4) Torque-response simulations	47
(3.5) Dynamics of linear DNA denaturation and supercoiling of ring DNA	49
(4.1) Dynamics of melting	56
(4.2) Melting curves	57
(4.3) Contour plot of free energy density	61
(4.4) First order transition for denaturation field ϕ	62

(4.5) Profiles of $f(\phi_0, \sigma)$	63
(4.6) Plot of the spinodal line and of the critical temperature.	64
(4.7) Common tangent construction	66
(4.8) Phase diagram	67
(4.9) Kymographs	70
(4.10) Dynamical scaling	71
(4.11) Evolution of the fields $\phi(x, t)$ and $\sigma(x, t)$ as a function of position x	72
(5.1) Twin-Supercoiling domain panel	78
(5.2) Melting during transcription	81
(5.3) Nucleosome wrapping and unwrapping during DNA transcription	83
(5.4) Diffusion of twist and relaxation of writhe	86
(5.5) Transcription in the absence of a topological barrier	88
(6.1) Representation of a complex between DNA and the <i>Zif268</i> protein	93
(6.2) Coarse-grained protein model	94
(6.3) Protein binding curve and kymographs of DNA binding proteins .	96
(6.4) Protein binding induces supercoiling	97
(A.1) Sketch of a base-pair with the phosphate-group included explicitly	102
(A.2) DNA grooves in the model	103
(A.3) Closure procedure	104
(B.1) Local supercoiling.	108
(B.2) Free energy of the non-conserved parameter.	110
(B.3) Free energy of the conserved parameter.	111
(B.5) Slope of the front of the σ field	115
(B.6) Computing linking number in denatured molecules	117
(B.7) Lk_d grows with the size of the denatured bubble.	118

(C.1) Model polymerase	120
(C.2) Snapshots from simulations of a moving polymerase	121
(C.3) Contact map	125
(C.4) Position and velocity of polIII as a function of the force	126
(C.5) Energies asymmetry in the coarse-grained model	128
(C.6) Kymographs for nucleosome wrapping and unwrapping	129
(C.7) Relaxation of the writhe field ($\chi(n, t)$)	132
(C.8) Plot of the maximum of $\chi(n, t)$ as a function of time	133
(C.9) Relaxation of bending energy	135

List of Tables

(2.1) Structural parameter of the A, B and Z form of DNA	9
(3.1) Parameter values in the model and expressed in simulation units.	39
(3.2) Mapping between simulation and physical units.	40
(C.1) Decay constants during relaxation of Writhe	134

List of Symbols

$\langle \dots \rangle$	Ensemble average.
γ	Friction coefficient acting on a bead.
$\mathbf{\Gamma}$	Torque.
ΔLk	Linking number difference.
$\epsilon_{\text{HB,c}}$	Critical hydrogen bond energy for denaturation in the coarse-grained model.
ζ	Elastic modulus.
η	Stochastic white noise.
θ	Twist angle between two consecutive base-pairs. For a DNA in the relaxed state θ_0 is used instead.
Θ	Polar angle in spherical coordinates.
ϑ	Fraction of denatured base-pairs.
μ	Chemical potential in a certain phase of the system.
μ_0	Mobility.
ν	Viscosity of the solvent.
Π	Pressure of the system.
σ	Super coiling.
σ_s	Diameter of a nucleotide bead.
τ	Inertial time.

τ_{Br}	Brownian time.
ϕ	Denaturation field.
φ	Azimuthal angle in spherical coordinates.
χ	Coupling parameter between the supercoiling (σ) and denaturation (ϕ) fields.
$\chi(n, t)$	local writhe field at position n along the DNA and time t .
ψ	Probability distribution for all the configurations of a polymer.
Ψ	Boltzmann weight.
$\Omega(m)$	Total twist angle between m consecutive segments.
$\Omega_3(s)$	Rate of rotation of a local reference frame along the curve around the tangent $\mathbf{t}(s)$.
a_0	Helical rise in double-stranded DNA.
b	Kuhn length.
C	Torsional stiffness.
D	Diffusion coefficient.
d	Distance from polymerase to the tip of the first plectoneme formed during transcription.
E_{bend}	Bending elastic energy.
E_{tors}	Torsional elastic energy.
\mathbf{F}	Force.
F_p	Magnitude of the force applied by the model polymerase during transcription.
f_+	Fraction of the DNA molecule under melting in the high supercoiling phase.
f_-	Fraction of the DNA molecule under melting in the low supercoiling phase.
$g(s, s')$	Correlation between tangent vectors located at s and s' .
H	Hamiltonian of the WLC model under a pulling force.
\mathcal{H}	Hamiltonian of the EWLC model under a pulling force.

k_B	Boltzmann constant.
l	Size of a denaturation bubble in a DNA molecule under melting.
l'	Length of a plectoneme produced during transcription.
L	Contour length of the DNA. In the state without external pulling force L_0 is used instead.
Lk	Linking number. For DNA in the relaxed state Lk_0 is used instead.
l_p	Persistence length.
l_τ	Torsional persistence length.
Lk_d	Linking number in a denaturated region along DNA.
m	Mass of a bead.
N	Total number of base-pairs.
P_i	Probability of micro-state i with energy E_i .
\mathbf{r}_{COM}	Position of a polymer centre of mass.
\mathbf{r}_n	Position of the centre of the n -th base-pair.
\mathbf{R}	End to end vector. The magnitude of this vector R is the end to end distance.
R_g	Radius of gyration.
s	Arc length.
s_c	Sedimentation coefficient.
\mathbf{t}_n	Tangent vector joining two consecutive base-pairs.
T	Temperature of the system.
T_c	Critical melting temperature of DNA.
Tw	Twist. For DNA in the relaxed state Tw_0 is used instead.
\mathcal{U}	Potential energy.
\mathbf{v}	Velocity.
v_p	Speed of a model polymerase during transcription.
Wr	Writhe. For DNA in the relaxed state Wr_0 is used instead.

- x_p Position of the base-pair closer to a model polymerase.
- Y Young's modulus.
- Z Partition function in the micro-canonical ensemble.

1

Introduction

Deoxyribonucleic acid (DNA) is arguably one of the most studied materials, and yet its properties and behaviour are not fully understood. DNA is usually composed of two strands twisted around each other, that in the relaxed state adopt the well known geometry of a uniform right-handed double helix [4]. Each of the two strands can be described as a polymer chain because they are made of a collection of simple and repetitive units called monomers. Although the detailed structure and the information stored by DNA depends on its exact sequence of monomers [5], the entire functioning of this molecule is more elaborate and depends, among other factors on: its dynamics, spatial organization and the interactions with proteins found in the intracellular environment [5, 6].

All of these aspects are related to DNA topology [7], i.e., to the geometrical properties and relations unaffected by continuous deformations of the polymer. One of the most important and ubiquitous deformations of DNA is due to supercoiling [8]. Supercoiling is an intrinsic property of under or overwound DNA molecules whose ends are not allowed to rotate (for example in bacterial plasmids that are circular [9], see Fig. 1.1(a)). In general, supercoiling results in the creation of higher order structures caused by mechanical deformations of DNA. When the bending or twisting of the polymer is altered from its relaxed state, this tends to lead to the DNA axis coiling up on itself, forming helices of higher order [10]. This is similar to what happens to a telephone cord that has been over- or under-wound while its ends are anchored [4, 8] (see Fig. 1.1(b)).

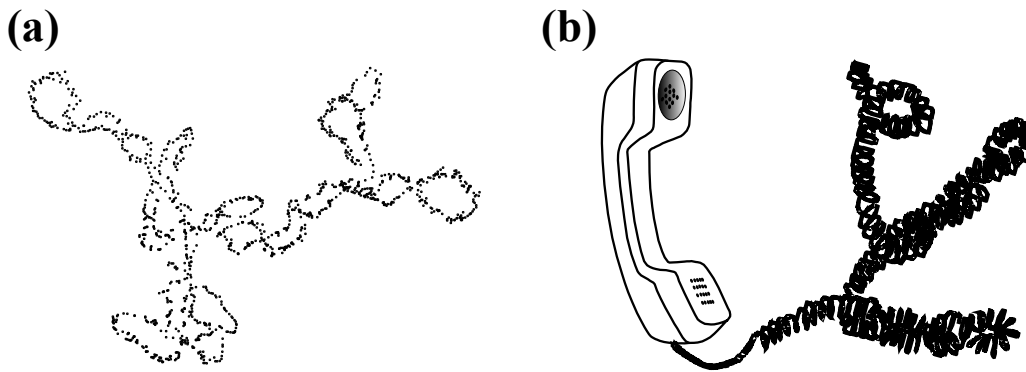


Figure 1.1 (a) Sketch based on a real electron micrograph of supercoiled DNA from *E. coli* bacteria, see for instance reference [11]. (b) Typical example of supercoiling in a telephone cord.

In vivo supercoiling is not only important in compacting DNA [12, 13] and regulating the access to the genetic code [5], but also in replication [14], where two identical replicas are created from one original DNA molecule, and in transcription [15], where a particular section along the DNA is copied into ribonucleic acid (RNA), a polymer with the instructions to create enzymes. On top of that, since the supercoiling is invariant as long as the backbones of both strands are not broken, the superhelical state of a DNA molecule is not altered by environmental factors like temperature or solution condition [7]. This offers an enormous experimental advantage in the study of its properties. This is why supercoiling has been a topic of particular interest and debate since the discovery of DNA.

Soon after their initial description of the helical structure of DNA, Francis Crick and James Watson discussed the mechanical consequence of gene transcription [16]. During transcription a specialized protein called polymerase needs to open up the two strands locally in order to “read” their sequence. As the polymerase moves along the polymer, stress is accumulated in the backbone, and eventually it may cause the entanglement of the DNA [17]. This could stop the transcription process, lead to the loss of genetic information or even cell death. Therefore, nature has developed special enzymes called topoisomerases, capable of removing the excess of stress, which are essential for proliferation and survival of all cells [18]. It is clear from the above, that to achieve a comprehensive understanding of the DNA biological function, we need a precise characterisation

of the role played by supercoiling in different processes.

The dynamic of DNA supercoils has been studied *in vitro* [19], however, current experimental techniques *in vivo* do not allow the determination of supercoiling in real time. In addition, only a small part of the full spectra of supercoiling degrees that may occur during cell life has been explored, focusing mainly on negative supercoiling [20] (corresponding to the underwinding of the helix), with just a few attempts to study the positive case [21, 22] (corresponding to DNA overwinding). To complicate the picture even more, the interplay between supercoiling and other phenomena, such as DNA melting [10] (the process in which a double-stranded DNA helix is heated up, the hydrogen-bonds break and DNA splits into two single strands) or the binding of proteins to DNA, is poorly understood.

The aim of this thesis is to get insight into the behaviour of some of the processes mentioned above and particularly into the function of supercoiling. To this end, in Chapter 2, I will first describe the key relevant features of DNA in its relaxed state and then I will discuss the basic polymer physics models that account for the elastic and topological behaviour of DNA found in single molecule experiments.

In Chapter 3, I will introduce a single nucleotide resolution coarse-grained computational model, that captures several of the DNA features described in Chapter 2. This model is able to simulate large molecules for time-scales which are relevant to physiological processes. Furthermore, I describe the strengths, limitations and possible ways to improve the fidelity of the model to represent real DNA.

In Chapter 4, I will discuss the melting of circular DNA under different topological constraints via simulations and theory. These results rationalise long-debated experimental findings and suggest new single molecule experiments to be performed on topologically-constrained DNA.

In Chapter 5, I will study the dynamics of supercoiling during transcription. Specifically, I will focus on the mechanical action of the polymerase on naked DNA.

In Chapter 6, I will show preliminary results on the binding of proteins to

DNA. I describe how the presence of these proteins alters the structure and level of supercoiling of DNA, and also how the supercoiling affects the binding.

Finally, in Chapter 7 I will summarize the main findings of this thesis, state the conclusions and propose possible future avenues that could be investigated using the model presented in Chapter 3.

2

Background: physical properties and models of DNA

Abstract: I review the basic structural and elastic features of DNA, and the two main polymer physics models that explain its elastic behaviour. I introduce the concept of supercoiling and discuss typical topological problems faced by DNA.

The structure of the DNA double helix has a deep impact on our understanding of the molecular basis of genetics. Thanks to the work carried out, among others, by James Watson, Francis Crick, Maurice Wilkins and Rosalind Franklin, it was revealed that DNA structure fits its function in the most remarkable way. We can see now that all organisms share the same machinery to store their hereditary information in the form of a double-stranded DNA (dsDNA), made by a pair of hybridized polymer chains which encode the specifications of the organism using four types of monomers [5].

This discovery was certainly one of the triumphs of 20th century, allowing scientists to read, interpret and even change this common code to achieve a coherent understanding of all the forms of life. However, it is becoming more and more evident that not only is the genetic information encoded in the DNA sequence of primary importance, but also how DNA folds in the three-dimensional space and its topology can alter crucial biological functions, such as gene expression and replication [5, 23–27]. That is why in this chapter I first focus on the structural and elastic properties displayed by DNA and then I study

the basis for higher levels of DNA organization such as supercoils.

2.1 DNA structure

Understanding the structure of DNA has been fundamental to describe the mechanisms that make life possible. However, DNA is a complex molecule which exhibits an elaborated architecture. The building blocks of DNA are called nucleotides. Any nucleotide is made of about 20 atoms, grouped into three main structures: a phosphate group, a five-carbon sugar (ribose or deoxyribose) and a nitrogen base (Adenine(A), Guanine(G), Cytosine(C) and Thymine(T)). The three parts are coupled through a chemical process known as condensation reaction, in which two molecules combine into a larger molecule with the loss of water, as shown in Fig. 2.1(a).

Individual nucleotides can be assembled into a single-stranded DNA (ssDNA) by chemical bonds linking the sugar of one nucleotide with the phosphate group of the adjacent one. With this process, a polymer chain with a directionality (known as 5' to 3') and composed of a repetitive sugar-phosphate backbone with bases protruding from it, is created (see Fig. 2.1(b)). Due to the chemical affinity between nucleotides, the bases of this ssDNA serve as a template to synthesize a complementary chain that runs in the opposite direction. The new chain is coupled to the original one via specific hydrogen bonds (HB) between bases. This unit of coupled monomers is known as base-pair (bp) and it is built following a local-pairing rule, which imposes the formation of hydrogen bonds such that A pairs with T, and G pairs with C. This reveals the basic mechanism of DNA replication: by unzipping the two complementary strands, each one can then be used as a template for the assembly of a new complementary strand, establishing the process of inheritance [28].

While the phosphates in a base-pair are 20 Å apart, the distance between adjacent sugars or phosphates in the backbone is 6 Å in the usual case (known as B-DNA, see Fig. 2.1(b)) and the thickness of one base is 3.4 Å. This distance is fixed because the bases are built with strong inflexible bonds between the atoms [4]. If DNA was assembled into a ladder made by the cross-chain pairing of bases, there would be a 2.6 Å hole between consecutive bases in the backbone.

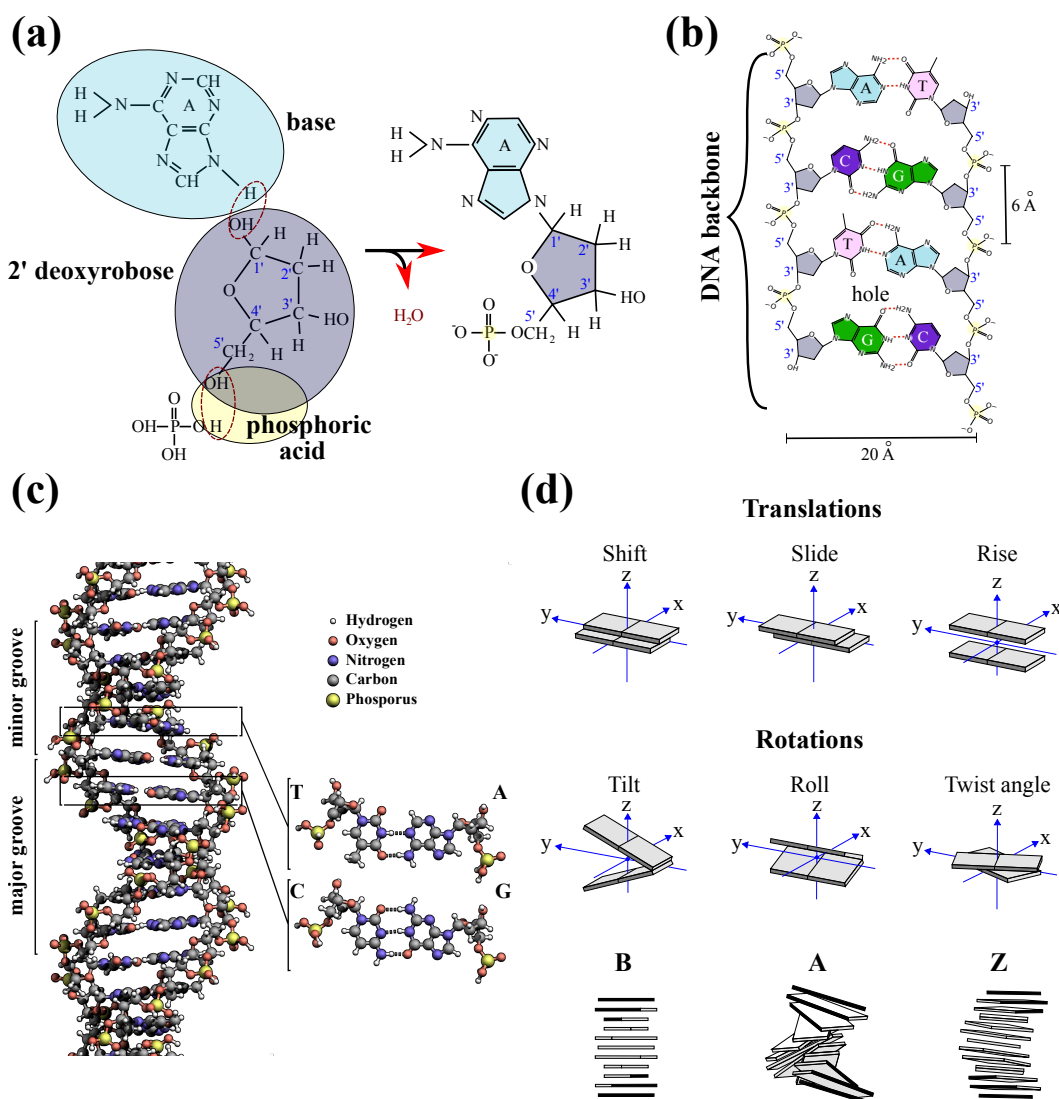


Figure 2.1 DNA assembly and structure. **(a)** Formation of a nucleotide through two condensation reactions: one between the phosphate group and the sugar (deoxyribose), and the second one between the sugar and the base (adenine). The carbons in the sugar are labelled in blue (from 1' to 5') starting from the one closer to the base. For the nucleotide shown the phosphate is attached to the 5' site. When there are more nucleotides in the system, the phosphate of the adjacent nucleotide is connected to the 3' site of the current one. That is why the directionality of the chain is known as 5' to 3'. **(b)** Schematic view of the DNA backbone and the base pairing interactions. Covalent bonds are shown as solid lines and hydrogen bonds as dashed lines. There are two HB between A and T and three between G and C (image modified from [29]). **(c)** 3D-Helical structure of B-DNA (modified from [30]). **(d)** Schematic representation of the six degrees of freedom between consecutive base-pairs. Translation along the x , y and z axes are called shift, slide and rise respectively. The tilt, roll and twist angle, are the rotations about the same axes. Using the parameters of Table 2.1 the A, B and Z forms of DNA are drawn.

This would permit water to get inside DNA, but as the bases in the nucleotides are hydrophobic, the DNA double strand adopts a form that minimizes the distance between consecutive base-pairs, a double helix. This molecule is stable in water at neutral pH.

2.1.1 Multiple structures of the double helix

There are different kinds of DNA helices whose structure is determined, among other things, by their sequence of bases and the solution conditions (pH, salt concentration, temperature, etc). The most common form of DNA in cells is the so-called B-form. Its structural parameters were derived from X-ray diffraction experiments of DNA at high humidity (92%) [31, 32]. In B-DNA each base-pair is twisted from the previous one by about 36° in the azimuthal direction, producing a right-handed double helix with the bases oriented perpendicular to the helix axis, and distinct major and minor grooves defined by the location of phosphates [33] (see Fig. 2.1(c)). The number of base-pairs necessary to complete a full helical turn of DNA is the pitch, which accounts to about 10 bp (34 Å) in B-DNA.

The A-form of DNA can be observed under conditions of lower humidity (75%) than the B-form [32, 34]. The structure of DNA in this form is the one of a right-handed helix with a diameter of 23 Å [33]. The rise between consecutive base-pairs (2.55 Å) is shorter than the one in B-DNA and they are not perpendicular to the helix-axis. The pitch of A-DNA is 11 bp (28.2 Å) per helical turn, related to a local twist-angle of 32.7° . In addition, the major groove is thin and deep while the minor groove is shallow.

The most intriguing structure of DNA is known as Z-DNA. This is characterized by a diameter of 18 Å, its rise and pitch are the largest of all the forms of DNA [33], 3.7 Å and 12 bp (44.4 Å) per helical turn respectively. The local twist-angle is approximately -30° (producing a left-handed helix). Since the physiological conditions inside a cell favour the structure of B-DNA, in the following I will refer mainly to this form unless otherwise indicated. However, we have to bear in mind that although the A/Z forms are rare, some experiments suggest that these forms might play an important role in biology. In reference [35] for example, it was shown that certain proteins bind to Z-DNA, protecting it

against viral intruders. It has also been shown that dehydration of DNA protects A-DNA under conditions such as the extreme desiccation of bacteria [36].

How the base-pairs stack on one another is an important feature of the different kinds of helices. In a simplified view, we can think of each base-pair as a rectangular solid block with a local reference frame (see Fig. 2.1(d)). In this representation, the relative orientation between two neighbouring base-pairs can be described in general using six degrees of freedom: three translations and three rotations (see Fig. 2.1(d)). Since the interaction at the level of nucleotides limits the range of motion of a base-pair, to a first approximation we only require three degrees of freedom to characterize the real DNA helices [4]: twist angle, roll and slide. The local twist angle is a rotation about the axis perpendicular to the plane of the base-pair, that runs along (or nearly along) the line joining the centre of two consecutive base-pairs. The roll describes a rotation about the longest axis in a base-pair. The slide refers to a translation in the direction of the longest axis of the base-pair. The values of the parameters that correspond to each of the forms of DNA are summarised in Table 2.1. Using this information, the schematic representation of the A, B and Z forms of DNA are depicted in Fig. 2.1(d).

Parameter	DNA form		
	A	B	Z
Helix Handedness	Right handed	Right handed	Left handed
bp per helical turn	11	10	12
Diameter	23 Å	20 Å	18 Å
Pitch	28 Å	34 Å	44.4 Å
Axial rise per bp	2.5 Å	3.4 Å	3.7 Å
Slide	2 Å	0 Å	1.92 Å
Roll	12°	0°	0.2°
Twist angle	32.7°	36°	-30°

Table 2.1 Structural parameter of the A, B and Z form of DNA. The values of the bp per helical turn, diameter, pitch and axial rise were obtained from reference [33] for the three forms of DNA. The values of slide, roll and twist angle can be found in [4, 33] for A-DNA, in references [4, 37] for B-DNA and [38] for Z-DNA.

2.1.2 Denaturation of DNA

Changes in the conditions of the system, such as temperature, pH, or decreasing the salt concentration, can cause the disruption of the relatively weak (noncovalent) forces between base-pairs [39–41]. Some of these changes may lead to the dissociation of the the dsDNA helix into its single-strand components: this process is known as DNA denaturation. A critical “melting” temperature T_c can be defined as the temperature at which 50% of base-pairs in solution are denatured. This process is reversible in the sense that by taking the system back to physiological conditions one can observe the formation of a regular double helix from ssDNA chains, a phenomenon known as hybridization or renaturation. Because G-C pairs form three hydrogen bonds, while A-T pairs form only two (see Fig. 2.1(b)), the content of GC bases in DNA has an effect on its melting temperature: the higher the GC content, the higher the value of T_c .

Experimentally, the denaturation can be monitored by measuring the absorbance of ultra-violet light passing through a solution of DNA [6]. Since the bases are the main objects responsible for the absorption of light at this wavelength, the absorbance at 260 nm is markedly higher for ssDNA with accessible bases. This process is screened when the bases are arranged into a DNA duplex, due to them being less accessible in this structure. The result is that the dsDNA absorbs less ultraviolet light by 40% than do individual ssDNA chains.

It is worth mentioning that experiments [42] from the 1960’s showed that early melting occurs in closed circular DNA (where the rotation of the ends is restricted) compared to the linear DNA, and it was also found that the critical melting temperature is higher for circular DNA. These observations imply, as we will see in Chapter 4, that topology plays an important role in DNA melting.

2.2 Elastic properties of DNA as a polymer

The description of DNA at the level of its main components, the nucleotides, is a good starting point to study its function. However, we will find that the relevant ingredients necessary to understand different aspects of DNA behaviour,

will occur at very different spatial and time scales. Depending on the scale of interest, a molecule of DNA will display different features. At the microscopic scale, DNA is a molecule 2 nm in diameter characterized by a double helical structure, with phosphates on the exterior and base-pairs on the interior. If looked at length scales ranging from 10-100 nm, DNA can be seen as a very rigid polymer. Above 100 nm, DNA looks like a semi-flexible polymer made by a collection of rods of about 50 nm (which is the persistence length, discussed more in detail in section 2.2.2). In the following I will adopt this view.

In addition, a DNA molecule is constantly subject to random collisions with the solvent. Such fluctuations affect the end-to-end distance of the molecule, and a competition between entropic and elastic forces determines the typical size of a DNA molecule in solution. Two models are often used to describe the entropic elasticity of DNA: the freely-jointed-chain and the worm-like-chain [43]. Here I introduce these polymer physics models.

2.2.1 The Freely-Jointed-Chain model

One of the simplest models in polymer physics is the freely jointed chain (FJC). In this model [44] a polymer is approximated as a series of N straight segments, all of the same length b , which we will call the Kuhn length (see Fig. 2.2(a)) and hence with total length $L = Nb$. The value of b represents the minimum length between two points on a polymer that are uncorrelated. Therefore, in the FJC model each segment is fully uncorrelated with respect to the others. This is unrealistic, because it means that the chain can overlap with itself. Yet, the FJC can still provide important insight into generic polymer behaviour.

The configuration of the chain is represented by the set of vectors $\{\mathbf{r}_n\}$ with $n = 0, 1, \dots, N$ or alternative by the segments:

$$\mathbf{t}_n = \mathbf{r}_n - \mathbf{r}_{n-1} \quad n = 1, 2, \dots, N. \quad (2.1)$$

Since these segments are independent, the probability distribution for all possible conformations is

$$\psi(\{\mathbf{t}_n\}) = \prod_{n=1}^N \psi(\mathbf{t}_n), \quad (2.2)$$

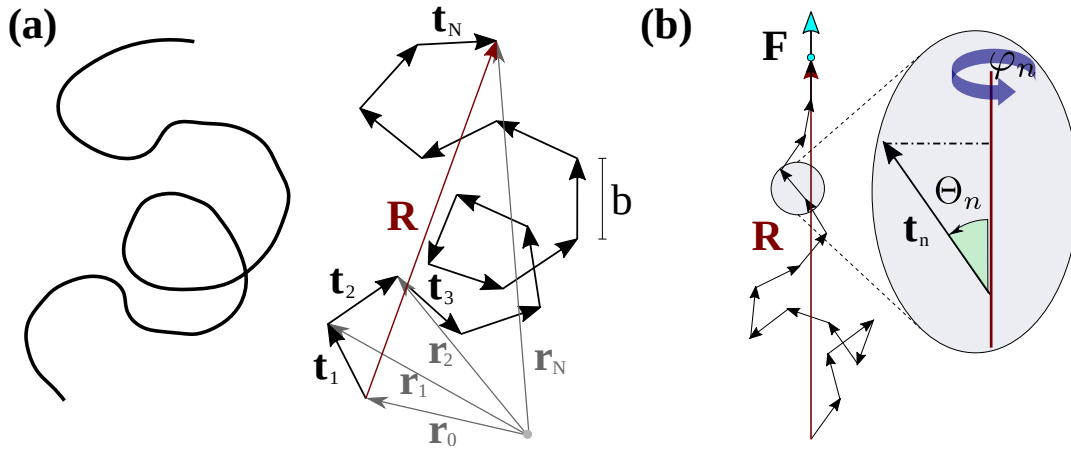


Figure 2.2 (a) Sketch of a polymer and its representation as a FJC. (b) Configuration of a polymer when a stretching force is applied to one end, and the other is fixed. The zoom-in to one of the Kuhn segments is also shown. Its orientation with respect the end-to-end vector \mathbf{R} is completely determined by the polar and azimuthal angles (Θ_n and φ_n).

where $\psi(\mathbf{t}_n)$ represents the random distribution of a vector with constant length b :

$$\psi(\mathbf{t}_n) = \frac{1}{4\pi b^2} \delta(|\mathbf{t}_n| - b). \quad (2.3)$$

The factor $1/4\pi b^2$ ensures the normalization of the function:

$$\int d\mathbf{t}_n \psi(\mathbf{t}_n) = 1. \quad (2.4)$$

One characteristic quantity related to the size of the polymer is the end to end vector (shown in red in Fig. 2.2):

$$\mathbf{R} = \sum_{n=1}^N \mathbf{t}_n. \quad (2.5)$$

Since all the segments are uncorrelated $\langle \mathbf{t}_n \rangle = 0$ and $\langle \mathbf{R} \rangle = 0$. However, $\langle R^2 \rangle$ has a finite value and can then be used to define a characteristic length of the chain. This is known as the mean-square end-to-end distance, which scales with the total length of the polymer as

$$\langle R^2 \rangle = \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{t}_n \cdot \mathbf{t}_m \rangle = \sum_{n=1}^N \underbrace{\langle \mathbf{t}_n \cdot \mathbf{t}_n \rangle}_{b^2} + 2 \sum_{n>m} \underbrace{\langle \mathbf{t}_n \cdot \mathbf{t}_m \rangle}_0 = Nb^2. \quad (2.6)$$

The FJC under an external pulling force

The behaviour of the expected end-to-end distance $\langle R \rangle$ of the FJC under a pulling force is important, because it can be compared directly with experiments [45, 46] (explained below). Here I study the case in which an external force pointing along the z-direction is applied to one end of the FJC, while the other end is fixed (see Fig. 2.2(b)). To this end, we introduce a spherical coordinate system where Θ_n denotes the angle of the segment \mathbf{t}_n with the z-axis and φ_n represents its azimuthal direction. The energy due to the external force is given as:

$$E = \sum_{n=1}^N E_n = -Fb \sum_{n=1}^N \cos(\Theta_n). \quad (2.7)$$

The probability of observing a micro-state i , with energy E_i associated to an end-to-end distance R_i , is given by its Boltzmann factor, properly normalized:

$$P_i = \frac{\exp(-E_i/k_B T)}{Z}, \quad (2.8)$$

where k_B is the Boltzmann's constant, T is the temperature of the system, $Z = \sum_i \exp(-E_i/k_B T)$ is the partition function in the canonical ensemble and the sum is over all possible micro-states i . For a system that takes on continuous values, we replace the sum by an integral

$$Z = \int d\mathfrak{t} \exp(-E(\mathfrak{t})/k_B T), \quad (2.9)$$

where \mathfrak{t} represents the integration over all configurations. Each segment can point in any direction on a sphere with radius b . Therefore, the integral has the form

$$Z = \int_0^{2\pi} d\varphi_1 \int_0^\pi \sin \Theta_1 d\Theta_1 \dots \int_0^{2\pi} d\varphi_N \int_0^\pi \sin \Theta_N d\Theta_N e^{(\frac{Fb}{k_B T} \sum_{n=1}^N \cos \Theta_n)}, \quad (2.10)$$

that can be rewritten as

$$\begin{aligned} Z &= \prod_{n=1}^N \int_0^{2\pi} \int_0^\pi d\varphi_n d\Theta_n \sin \Theta_n \exp\left(\frac{Fb}{k_B T} \cos \Theta_n\right) \\ &= \left[4\pi \frac{\sinh(Fb/k_B T)}{Fb/k_B T}\right]^N. \end{aligned} \quad (2.11)$$

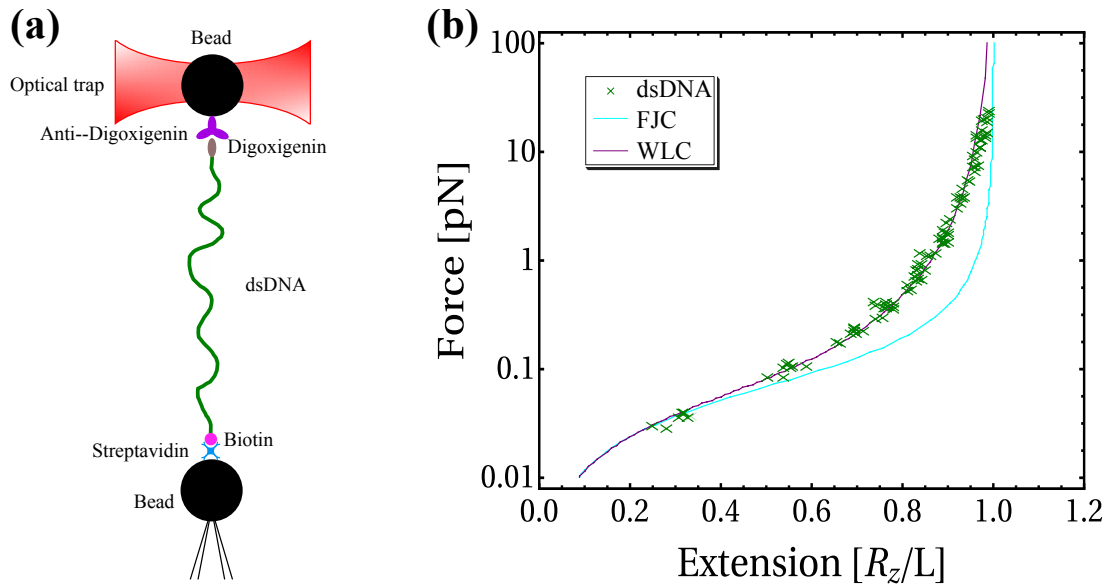


Figure 2.3 (a) Schematic representation of the dsDNA pulling experiment. The ends of the dsDNA are tethered to microbeads through organic molecules complexes (biotin-streptavidin and digoxigenin-antidigoxigenin). While one of the beads is fixed to a micropipette, the other is under the action of an external pulling force. (b) Force vs extension data for λ -phage dsDNA (48,502 bp long) pulled by magnetic beads [46]. The data (green crosses) are fit to the WLC model (Eq. (2.22)) returning a persistence length $l_p = 53$ nm. The FJC assuming $b = 2l_p = 106$ nm in Eq. (2.13) is also shown.

Since the force is only in the z -direction, the chain displays a random walk in the x and y -direction and consequently $\langle \mathbf{R} \rangle = \langle \mathbf{R}_z \rangle$. Therefore, with the partition function (Eq. (2.11)) the expected value of the end-to-end distance can be computed by using the well known statistical mechanics relation:

$$\langle R_z \rangle = k_B T \frac{\partial}{\partial F} \ln(Z). \quad (2.12)$$

By substituting Eq. (2.11) into Eq. (2.12) and taking the derivative, we find the force-extension relation.

$$\langle R_z \rangle = Nb \left[\coth \left(\frac{Fb}{k_B T} \right) - \frac{k_B T}{Fb} \right]. \quad (2.13)$$

A strict test of the applicability of the FJC model to DNA has been provided by extension experiments [43, 46], where the ends of a double-stranded DNA molecule are attached to a pair of protein coated micron-sized beads (see Fig. 2.3(a)). One of the beads is connected to an optical trap, while the other

is placed on a micropipette (or sometimes to a glass surface). Then, the two strands are pulled apart by known forces acting on the top bead, and the end-to-end distance is recorded as a function of the pulling force \mathbf{F} . The results in Fig. 2.3(b) indicate that while the FJC model can describe the behaviour of dsDNA in the limit of low stretching forces ($F < 0.1$ pN), it fails at higher forces.

The reason why the FJC model fails to predict the behaviour of real DNA, is because it only considers the entropic forces tending to counteract the external stretching force, and misses the energetic cost of deforming the polymer. To account for this, each configuration has to be appropriately weighted by its corresponding bending energy cost. Both the bending energy and the entropic contribution are taken into account by the worm-like-chain model.

2.2.2 The Worm-Like-Chain model

At mesoscopic scales it is possible to think of DNA as an elastic rod characterized by three physical quantities: *stretching*, which measures the increase in length of a material when is subjected to a force; *bending*, related to the rate of change of the tangent vector along the rod; and *twist*, which measures how each consecutive infinitesimal element of the rod changes orientation about the axis relative to its neighbour. The bending is accurately described by the Worm-Like Chain model (WLC). In this model DNA is considered as an inextensible rod, represented by a curve parametrized by arc length (s), and embedded in the three dimensional space \mathbb{R}^3 , according to the mapping: $s \in [0, L] \rightarrow \mathbf{r}(s)$, with L the fixed total length of the rod. Unlike the FJC, the WLC model is a continuous model where the bending energy is determined by the geometry of the curve. Following the description of an elastic rod by Moroz and Nelson [47] one can define the bending rigidity via the elastic energy functional

$$\frac{E_{\text{bend}}}{k_B T} = \frac{l_p}{2} \int_0^L \left(\frac{d\mathbf{t}}{ds} \right)^2 ds, \quad (2.14)$$

where $\mathbf{t}(s) = \frac{d}{ds}\mathbf{r}(s)$ is the tangent vector at location $\mathbf{r}(s)$ along the curve, and l_p is the *persistence length*, a local quantity that gives a measure of the length-scale along the polymer contour over which tangent-tangent correlations decay. The

distribution of polymer conformations is thus given by the Boltzmann weight

$$\Psi[\mathbf{t}] \propto \exp\left(-\frac{E_{\text{bend}}}{k_B T}\right) = \exp\left[-\frac{l_p}{2} \int_0^L \left(\frac{d\mathbf{t}}{ds}\right)^2 ds\right]. \quad (2.15)$$

Given Eq. (2.15), it can be shown [44] that the correlation $g(s, s')$ between tangent vectors located at s and s' has the form

$$g(s, s') = \langle \mathbf{t}(s) \cdot \mathbf{t}(s') \rangle = e^{-|s-s'|/l_p}, \quad (2.16)$$

where $\langle \dots \rangle$ denotes ensemble average. This exponential relation comes naturally as a consequence of the fact that to compute l_p we average over all configurations of the polymer, with each one weighted by a Boltzmann factor. In addition, we can argue that the correlator $g(s, s')$ should be 1 for $s' = s$ and 0 for $|s - s'| \gg l_p$. Furthermore, two infinitesimal segments of the polymer (located at $s + s_1$ and $s + s_2$) are subjected to independent thermal forces and therefore, the tangent-tangent correlation over both segments is the product of the individual ones: $g(s, s + s_1 + s_2) = g(s, s + s_1) g(s, s + s_2)$. In an aqueous solution, the average persistence length of DNA is 46-50 nm or 140-150 bp. This makes DNA a very stiff molecule, given that it is only 2 nm thick.

To characterize the size of a polymer, the mean square end-to-end distance and the radius of gyration are useful quantities. They can both be obtained from the definition of the end-to-end vector in the WLC.

$$\mathbf{R} \equiv \mathbf{r}(L) - \mathbf{r}(0) = \int_0^L \mathbf{t}(s) ds. \quad (2.17)$$

The magnitude of the end-to-end vector is called the end-to-end distance, a scalar measure of the Euclidean distance between the two end points of the molecule at any instant. By using Eq. (2.16) the mean square end-to-end distance is easily calculated:

$$\begin{aligned} \langle R^2 \rangle &= \int_0^L ds \int_0^L ds' \langle \mathbf{t}(s) \cdot \mathbf{t}(s') \rangle \\ &= 2Ll_p - 2l_p^2 \left[1 - e^{-L/l_p}\right]. \end{aligned} \quad (2.18)$$

Depending on the relation between the total length of the polymer and

its persistence length, we can make a general classification of the stiffness by identifying two limiting cases in the last equation

- If $l_p \gg L$ the polymer behaves like a rigid rod and $\langle R^2 \rangle = L^2$.
- If $l_p \ll L$ the polymer exhibits a random walk behaviour and $\langle R^2 \rangle = 2Ll_p$. By comparing this result with Eq. (2.6) of the FJC model, we find that $b = 2l_p$.

The radius of gyration R_g of a DNA molecule is an average measure of the Euclidean distance between any point on the DNA backbone and its center of mass, \mathbf{r}_{COM} . In addition, it is the most easily measured DNA property, because it can be directly related to light scattering experiments [48], where the time-averaged intensity of scattered light by DNA, as a function of the scattering angle is obtained. This yields information on the structure of the macromolecule (including R_g). The mathematical expression for the average of the square radius of gyration is

$$\begin{aligned} \langle R_g^2 \rangle &\equiv \frac{1}{L} \int_0^L ds \langle |\mathbf{r}(s) - \mathbf{r}_{\text{COM}}|^2 \rangle \\ &= \frac{1}{L^2} \int_0^L ds' \int_0^{s'} ds \langle |\mathbf{r}(s) - \mathbf{r}(s')|^2 \rangle. \end{aligned} \quad (2.19)$$

For a polymer in the limit of random walk behaviour ($l_p \ll L$), the radius of gyration is related to the mean square end-to-end distance by the equation

$$\langle R_g^2 \rangle \simeq \frac{1}{6} \langle R^2 \rangle = \frac{Ll_p}{3}. \quad (2.20)$$

It is interesting to note that for a bacterial chromosome, such as that of *E. coli* with a total length of 4.7 million base pairs ($L \simeq 1.5$ mm) and a persistence length of approximately 53 nm, the radius of gyration given by Eq. (2.20) is around 5 μm [49], which is more than double the size of the longest dimension of the cell (2 μm). This implies that the DNA *in vivo* is not a Gaussian coil and that bacteria must package their DNA. As we will see this is related to a higher level of DNA organization.

Stretching an inextensible WLC

When the WLC is under a stretching force, the system is described by a Hamiltonian composed of two parts: the bending energy given by Eq. (2.14) and the tensile energy expressed in terms of the displacement of the two ends:

$$H = \frac{k_B T}{2} \int_0^L ds l_p \left(\frac{d\mathbf{t}}{ds} \right)^2 - \mathbf{F} \cdot [\mathbf{r}(L) - \mathbf{r}(0)], \quad (2.21)$$

where the last term can be identified with the Lagrange multiplier \mathbf{F} to fix the end-to-end extension $[\mathbf{r}(L) - \mathbf{r}(0)]$. If the tension is applied along the z direction, $\mathbf{F} = F\mathbf{e}_z$, and the z component of the end-to-end vector represents the extension of the chain $R_z = [\mathbf{r}(L) - \mathbf{r}(0)] \cdot \mathbf{e}_z$.

The numerical integration of Eq. (2.21) that gives the force versus extension relation is found in reference [50], where also, an useful expression that is asymptotically exact in the large- and small-force limits is derived. This expression takes the form

$$\frac{Fl_p}{k_B T} = \frac{R_z}{L} + \frac{1}{4(1 - \frac{R_z}{L})^2} - \frac{1}{4}, \quad (2.22)$$

when $L \gg l_p$ is assumed. This equation is derived following the same steps presented in section 2.2.1 for the FJC. In practical terms, the partition function has to be computed, but this time we replace the energy in Eq. (2.8) by the Hamiltonian of Eq. (2.21) and then we have to use the Feynman path integral to sum over micro-states. The force extension Eq. (2.22) is obtained by taking the derivative with respect to the force of the logarithm of this partition function.

The results predicted by Eq. (2.22) have been successful at accounting for the elastic behaviour of DNA at forces up to 10 pN, as shown in Fig. 2.3(b). However, the model failed to reproduce the force extension data from DNA experiments [46, 51] in the over-stretching regime (when the length of the polymer is slightly longer than L_0 , the contour length in the state without external force $\mathbf{F} = 0$), showing that also the intrinsic structure of the DNA has to be taken into account.

Extensible WLC

The WLC model can be extended to incorporate the elongation of polymers under tension. In this case it is assumed that the contour length (L) of the polymer is not longer fixed, and in particular it is a function of \mathbf{F} , the tension exerted at both ends of the chain. The Hamiltonian of the system is composed by the two terms in Eq. (2.21) and an additional term for the energy of material elongation:

$$\mathcal{H} = \frac{k_B T}{2} \int_0^L ds l_p(L, s) \left(\frac{\partial \mathbf{t}}{\partial s} \right)^2 + \frac{1}{2} \int_0^L ds \zeta(L, s) \left(\frac{s}{s_0(s)} - 1 \right)^2 - \mathbf{F} \cdot [\mathbf{r}(L) - \mathbf{r}(0)]. \quad (2.23)$$

In the first term of Eq. (2.23) the bending energy has exactly the same form as for a WLC (Eq. (2.14)), but in this case the persistence length will generally depend on the material elongation and hence implicitly on the tension. In the second term of Eq. (2.23), $\zeta(L, s)$ is the *elastic modulus*, a quantity that measures the resistance of an object to being deformed (not permanently) when a force is applied to it, and s/s_0 is the *extension ratio*, which is a measure of how much an object has been extended. Finally, the last term of Eq. (2.23) represents the energy due to the tension.

When the force is applied along the z direction and the limit case of small elongations ($\Delta L = L - L_0 \ll L_0$) is considered, the force-extension relation (derived in reference [52]) is given by

$$R_z = L_0 \left[1 + \frac{F}{\zeta} - \frac{k_B T}{2FL_0} \left(1 + \sqrt{\frac{FL_0^2}{l_p k_B T}} \coth \sqrt{\frac{FL_0^2}{l_p k_B T}} \right) \right]. \quad (2.24)$$

For chains longer than the deflection length [53] (quantity that determines the rigidity of a semiflexible polymer when confined to a cylindrical tube), this reduces [52, 54] to

$$R_z = L_0 \left[1 - \frac{1}{2} \left(\frac{k_B T}{Fl_p} \right)^{1/2} + \frac{F}{\zeta} \right]. \quad (2.25)$$

The force-extension curves from experiments [55, 56] at high salt concentration

have shown that Eq. (2.25) is accurate in the force regime between 5 and 50 pN. The estimations for ζ obtained there, give $\zeta \sim 1050$ pN. This value can be related to the Young's modulus for an elastic rod by $Y = \zeta / (\pi r_{\text{DNA}}^2) \sim 0.3$ GPa (considering that the radius of DNA is $r_{\text{DNA}} = 10$ Å), which is about three times larger than the one of rubber.

Torsion in the WLC

The extensible worm-like chain takes into account both, bending and stretching, however, a complete description of DNA mechanics should also consider the effects of torque. Modelling the coupling of all these interactions is non trivial. The twistable WLC model (TWLC) correctly describes the general behaviour of dsDNA to applied torques and pulling forces below the over-stretching limit ($F < 10$ pN), when the polymer is treated as an inextensible elastic rod. In this model the Hamiltonian of the system comprises the bending energy of Eq. (2.14) and a torsional elastic energy, but does not include a twist-bending coupling term [57].

Following Moroz and Nelson [47], the torsional elastic energy functional of the TWLC model is given by

$$\frac{E_{\text{tors}}}{k_B T} = \frac{C}{2} \int_0^L \Omega_3(s)^2 ds, \quad (2.26)$$

where C is the intrinsic torsional stiffness and $\Omega_3(s)$ is the rate of rotation of a local reference frame along the curve around the tangent $\mathbf{t}(s)$ (defined as in section 2.2.2). Analogous to the measurement of the bending persistence length via the tangent-tangent correlation, we can measure the torsional persistence length by computing the decorrelation of the twist angle. As described in reference [58], we stress here the fact that the TWLC model, by neglecting the twist-bending coupling, fails to explain the pulling force dependence of the torsional stiffness seen in experiments [59, 60]. However it is a good first approximation to modelling the elastic behaviour of DNA.

To understand the torsional rigidity in DNA we consider a discrete rod version of an elastic rod, composed of segments of constant length a_0 . With

this assumption the correlation of the twist angle can be quantified by defining a local reference frame at each segment, and tracking the rotation of the frames via their Euler angles. As specified in reference [61], Eq. (2.26) is an integral over the rate of rotation of the Darboux frame (or material frame) of reference with respect to the distance along the rod. The discrete approximation of this frame can be given in terms of the Euler angles $\alpha_n, \beta_n, \gamma_n$ which describe the rotation of the frame at segment $n + 1$ with respect to the one at segment n . Using this discrete approximation one can find that

$$\frac{E_{\text{tors}}}{k_B T} = \frac{C}{a_0} [1 - \cos(\alpha_n + \gamma_n)]. \quad (2.27)$$

The twist angle between frames is given by $\alpha_n + \gamma_n$, so the total angle between m consecutive segments is given by $\Omega(m) = \sum_{n=1}^m (\alpha_n + \gamma_n)$. An appropriate measure of the thermal fluctuations about the equilibrium twist is given by the mean of the cosine of this angle; since this quantity will decrease with m , we identify the decay constant as the torsional persistence length l_τ via

$$\langle \cos \Omega(m) \rangle = e^{-ma_0/l_\tau}. \quad (2.28)$$

The ensemble average is obtained in the usual way by taking the integral over the phase space of the system; in the small a_0 limit this gives $l_\tau = 2C$. For a DNA molecule, the non-zero equilibrium twist angle ($\theta_0 = 36^\circ$) between each base-pair will appear in the energy functional and it must be subtracted from $\Omega(m)$ so that the ensemble average is a simple exponential decay.

Since the resolution of experiments does not give information on the positions of single nucleotides, it is not possible to directly compute either the persistence length from the tangent-tangent correlation (Eq. (2.16)), or the torsional persistence length from the discrete frame correlation (Eq. (2.28)). Instead, as we have already seen, the persistence length may be found from force-extension experiments [46]. In a similar way, the torsional persistence length is found from torque-response experiments [62].

One important feature of DNA is that an applied torque can cause a loop in which two distant segments along its axis come into contact. In other words, the 3D conformation of DNA is the result of a competition between bending and

torsion. In the following section we will show how this physical aspect, when combined with DNA topology, can be used as a way of measuring the torsional stiffness C .

2.3 DNA supercoiling

The DNA double helix has a preferred configuration where the two strands wrap around each other approximately once every 10.5 base pairs. Mechanical manipulation which over or under winds the helix leads to torsional strain that can be relieved by distorting the shape of the DNA, or in other words the DNA writhes up on itself. This phenomenon is known as *supercoiling*. If the DNA is in a closed loop (such as a plasmid in bacteria), or is linear with fixed ends, the number of times the two strands wind around each other is a topological invariant known as linking number (Lk). This number, which must be an integer for chains whose ends are not allowed to rotate, is related to two geometrical properties of the molecule: the twist (Tw), i.e. the extent of rotation of the two strands about the helix axis, and the writhe (Wr), that accounts for the number of self-crossings of the dsDNA centreline. The three of these quantities are connected by the well-known White-Fuller-Calugareanu theorem [63] which states that the sum of the twist and writhe of DNA at time t , is equal to its linking number

$$Lk(t) = Tw(t) + Wr(t). \quad (2.29)$$

In a DNA molecule, most of the excess in twist (when is overwound) is converted into positive writhe. The opposite is also true, for an underwound molecule, the deficit in turns is converted into negative writhe. The way of distinguish these two cases is by the orientation of the crossings of the centre line of the helix (see Fig 2.4).

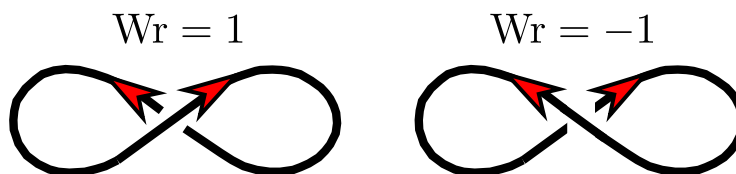


Figure 2.4 Diagram of positive (left) and negative writhe (right).

A distinctive feature of circular DNA is that the value of Lk can not be altered by any conformational rearrangement that does not involve the nicking of one or both strands. However, changes in the configuration of the molecule that redistribute the twist and writhe are allowed as long as Eq. (2.29) is satisfied. For DNA $Tw_0 = N/p$ (with N the length of the molecule in base-pairs and p representing its pitch) is the relaxed twist. This means that for a circular dsDNA initialized in a configuration with $Lk(t = 0) = Lk_0 = Tw_0$ and $Wr(t = 0) = Wr_0 = 0$, its temporal evolution will show either, no self-crossings of the DNA axis (see Fig. 2.5(a)) or crossings with opposite sign (as it will be seen in Chapter 5). If, instead, the circular dsDNA is initialized in a over-wound state ($Lk(t = 0) > Lk_0$) (see Fig. 2.5(b)), the excess of linking number will redistribute in twist and writhe, and the DNA will cross itself. This phenomenon, topological in nature, plays an important role in many biological processes [64].

A DNA molecule whose linking number differs from the one in the relaxed state is said to be supercoiled. Therefore, the linking difference $\Delta Lk = Lk - Lk_0$ is a quantitative measure of DNA supercoiling. If $\Delta Lk < 0$ the DNA is said to be negatively supercoiled. Conversely if $\Delta Lk > 0$ the DNA is positively supercoiled. In *vivo* the only way to change Lk is to introduce a break in one or both DNA strands, rotate the two DNA strands relative to each other and seal the break. This is precisely the role of DNA topoisomerases type I (one strand break) and II (two strands break) respectively. Both prokaryotes and eukaryotes, have type I and type II topoisomerases capable of removing supercoils from DNA. While Type I topoisomerase is responsible of altering the twist of DNA, the type II topoisomerase changes the writhe [65, 66]. In addition, prokaryotes have a particular topoisomerase known as gyrase that generates negative supercoils [65].

Supercoiling introduces torsional stress in a DNA molecule and this stress is proportional to the linking number difference ΔLk compared to the relaxed Lk_0 , i.e. is proportional to the superhelical density:

$$\sigma = \frac{\Delta Lk}{Lk_0}. \quad (2.30)$$

Circular DNA from bacteria are usually negatively supercoiled, having values of σ that vary around -0.06 [6]. The negative supercoiling is due to the untwisting of the DNA and it is useful in processes that require strand separation such

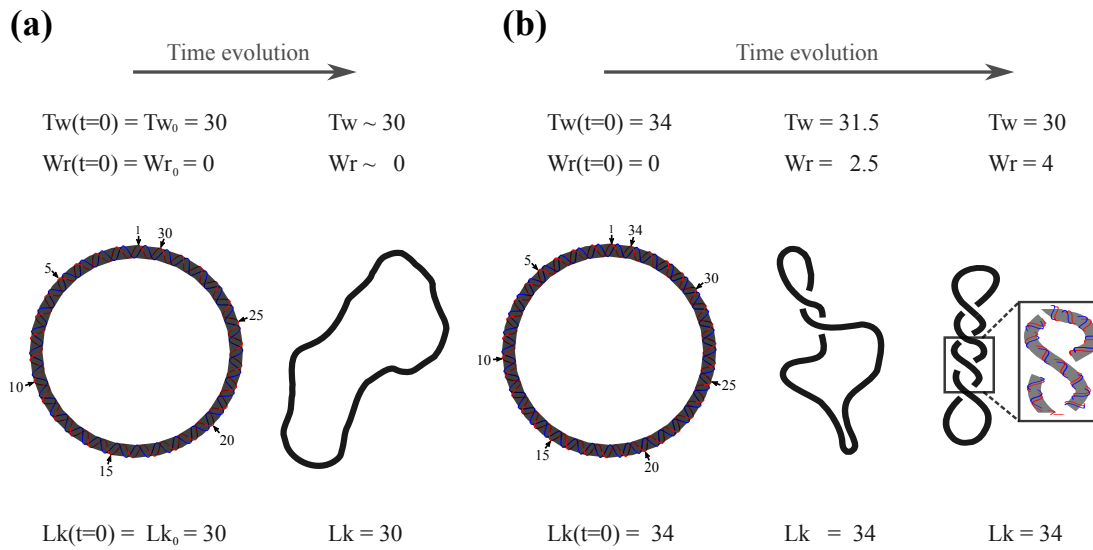


Figure 2.5 (a) Evolution of a circular dsDNA molecule 315 bp long, initialized with a relaxed linking number Lk_0 . (b) Temporal evolution of twist and writhe when a circular dsDNA molecule is initialized with a excess of linking number $Lk = 34$ (positively supercoiled). The helical turns are labelled at $t = 0$ for both cases. Both (a) and (b) diagrams are based in results from the simulations using the model introduced in Chapter 3.

as replication and transcription [10, 67]. Opposite to this behaviour, positive supercoiling helps to prevent the denaturation of a dsDNA, even under conditions of high temperature. Remarkably, this topological property of dsDNA has been selected by evolution in organisms that live in environments with extremely high temperatures - the thermophiles. These microorganisms possess (in addition to gyrase), a special DNA topoisomerase called reverse gyrase, which introduces the positive supercoiling [68]. The presence of these two types of topoisomerases suggests that thermophiles prevent DNA melting by positive supercoiling, but when needed they favour general linking excess and local melting, for example during DNA transcription [69].

2.3.1 DNA torque response

The number of times a DNA molecule crosses over itself will depend on the competition between its persistence length (l_p) and intrinsic torsional stiffness (C). As mentioned before, C can be extracted from torque-response curves [62]. Briefly, the idea is to study the behaviour of a dsDNA molecule which is fixed

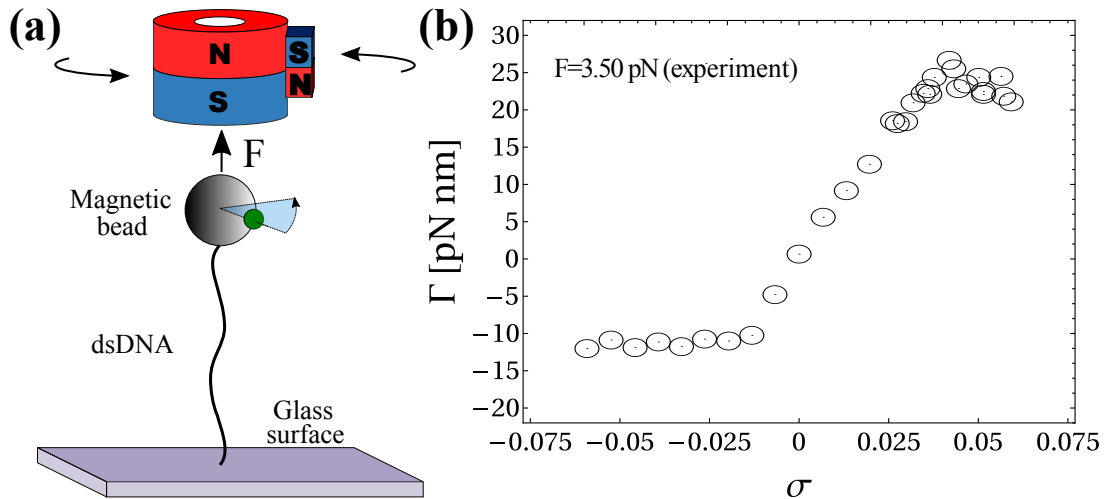


Figure 2.6 (a) Schematic representation of the dsDNA magnetic tweezers experiment [70]. The dsDNA molecule is tethered to a glass surface from one end, and to a magnetic bead (shown in black) to the other. A cylindrical magnet is used to exert a pulling force \mathbf{F} on the magnetic bead. The torque is applied via an additional magnet placed on the side of the cylindrical one. The marker bead (shown in green and attached to the magnetic bead) allows to track the number of turns added or removed to the dsDNA. (b) Torque-response curve: experimental data obtained from [62, 71] at $[Na^+] = 550$ mM and under a tension $F = 3.5$ pN.

at one end, while the other is under the action of both a stretching force ($\mathbf{F} = F\mathbf{e}_z$) and a torque ($\mathbf{\Gamma} = \Gamma\mathbf{e}_z$), the latter required to impose a given twist (see Fig. 2.6(a)). For large stretching forces, the change in twist cannot be converted into writhe and therefore $Lk = Tw$.

The torque needed to impose a certain twist is plotted as a function of the super helical density (see Fig. 2.6(b)), and exhibits a linear regime at small σ , related to the value of C_{eff} , the effective torsional stiffness [72, 73] (in contrast to the intrinsic stiffness C):

$$C_{\text{eff}} = \frac{a_0 \Delta\Gamma}{\theta_0 \Delta\sigma}. \quad (2.31)$$

Here $a_0 = 0.34$ nm is the helical rise for a relaxed dsDNA, θ_0 is the equilibrium twist angle across a base-pair step in the relaxed case and C_{eff} expresses the resistance of the DNA to a global torsional deformation, applied at its ends. Since at low stretching forces the bending fluctuations can absorb a significant part of the applied torsional stress: $C_{\text{eff}} < C$. However, for sufficiently large stretching forces, bending fluctuations are mostly suppressed, and hence the

effective torsional stiffness tends to approach the intrinsic one ($C_{\text{eff}} \approx C$). This latter case will be explored later in section 3.4.2. Finally, by fitting the data in Fig. 2.6(b) with Eq. (2.31) the value $C_{\text{eff}} \sim 115$ nm is obtained.

It is interesting to notice that the torque-supercoiling relationship for DNA shows two constant-torque regions, reflecting structural transitions at $\Gamma \sim -10$ pN nm and $\Gamma = 25$ pN nm (see Fig. 2.6(b)). This behaviour has been confirmed in several experiments [74–76]. While the plateau in the underwinding regime (at $\Gamma \sim -10$ pN nm) has been related to the denaturation of the helix to single strands [74, 77], the plateau in the overwinding regime was postulated [74, 77] to reflect the conversion of canonical B-DNA into the extremely overwound helical form called Pauling-DNA [75].

Similar experiments [78, 79] (which use the same technique depicted in Fig. 2.6(a)) have been done to extract the end-to-end distance of a DNA molecule as a function of the superhelical density, for different stretching forces applied to the molecule. In these experiments the tension resists the contraction of the DNA and as we have seen, if high enough it could inhibit the writhing. However, under low applied forces ($|\mathbf{F}| < 0.5$ pN) the elastic behaviour of DNA is basically symmetrical for positive and negative supercoiling. Both the excess and deficit of turns promote the formation of plectonemes (supercoiling in the form of writhe) and therefore the end-to-end distance reduces.

At higher forces the chiral nature of DNA becomes evident. Under pulling forces in the range ($0.5 \text{ pN} < |\mathbf{F}| < 3 \text{ pN}$) the extension of the DNA remains constant as it is underwound, whereas it reduces when overwound. This is explained in the following way: when the molecule is underwound, the pulling force hinders the formation of negative writhe and the stress accumulated by further under twisting it, leads to its partial denaturation. The separated strands thus absorb the molecule's linking number deficit and allow for the remainder of the DNA to adopt its natural B-form. On the other hand, positive supercoiling inhibits denaturation and the DNA is therefore enforced to writhe to relieve torsional stress.

At high forces ($|\mathbf{F}| > 3 \text{ pN}$) writhing is suppressed for both positive and negative supercoiling, and the extension of the molecule changes very little with the supercoiling. In this case undertwisting the molecule has the same effect as

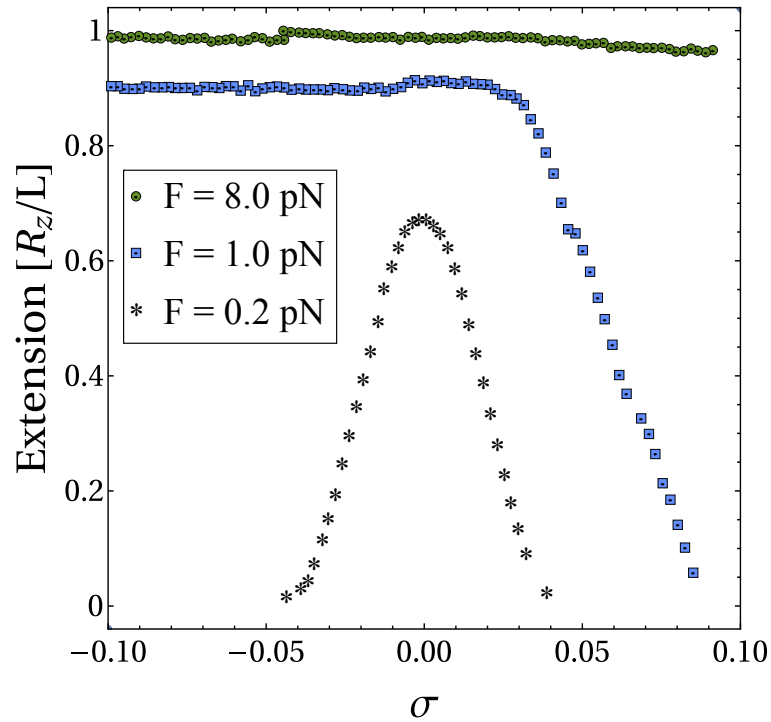


Figure 2.7 *Extension vs supercoiling curve: experimental data obtained from reference [78] for dsDNA under three different pulling forces (0.2 pN, 1 pN and 8 pN).*

in the intermediate forces (with a longer section of DNA denature and a short segment in its B-form). This time however, the excess of linking number obtained when overwinding the molecule, is stored in local regions with a high helical pitch (DNA adopts the Pauling form) instead of in the plectonemes. The three cases are depicted in Fig.2.7.

Simulations [80] and experiments [81] performed with DNA minicircles (90-178 bp long in simulations and 339 bp in experiments) have found a similar behaviour (to the one above for intermediate levels of applied forces and torques): positive supercoiled DNA has a much higher propensity to writhe. In contrast, the negatively supercoiled minicircles has a tendency to denature regions of the DNA instead of writhing. Writhing of negatively supercoiled minicircles was only observed under conditions of high salt concentration (which inhibits denaturation).

2.3.2 Differences between positive and negative supercoiling in DNA

From the previous section it is clear that there is an asymmetry between overwound and underwound DNA. Therefore, when comparing the properties of positive and negative supercoiling, it is important to consider how supercoiling is partitioned into ΔTw and ΔWr , and also to keep in mind that while writhe reduces the torsional strain and thus the twisting energy, there is also a bending energy cost. Several models of the elasticity of DNA ignore these facts. Thus they do not consider that the torque required to overtwist DNA differs from the torque required to undertwist it. One of the first attempts to model this particular behaviour of DNA was made by Marko and Siggia [82]. By considering DNA as a right-handed helix with major and minor grooves, instead of an isotropic elastic rod, they postulated a coupling between twist and bend. In this work they neglect the abrupt structural changes of DNA such as denaturation or conversion to the Pauling-DNA form. Under these assumptions they found that DNA rings with the same magnitude of supercoiling but opposite sign will have different average structures. Furthermore, an imposed twist would change the effective persistence length: overwinding stiffens the DNA molecule, while underwinding makes the resulting structure easier to bend. As we will see in Chapter 5, our simulations of a model polymerase moving along DNA during transcription are in agreement with that result.

2.4 Summary

This chapter introduced the basic structural properties of dsDNA in equilibrium such as rise, pitch, helicity, base-pair and backbone interactions. In addition, the parameters related to the elastic response of DNA found in experiments and explained through polymer physics models were introduced and their physical meaning discussed. These intrinsic parameters and the chirality of the molecule make DNA respond differently to over- and under-twisting. We also defined the fundamental topological concept of supercoiling in terms of linking number, writhe and twist of the DNA. The structural changes due to supercoiling produce higher levels of DNA organization which are important not only for

the confinement of DNA but also its function. Furthermore, supercoiling has an important influence on many DNA processes *in vivo* such as denaturation, replication or transcription. It is therefore desirable that a dsDNA model takes into account all these physical aspects. In the next section I will present a computational model for dsDNA which accomplishes this and that could therefore be used as a starting point for the study of the dynamics of DNA supercoiling.

3

A coarse-grained model for large-scale simulations of DNA

Abstract: I implement a mesoscopic model for dsDNA at the single nucleotide level. This model retains the structural properties mentioned in the previous chapter and compares favourably with single molecule experiments where dsDNA is manipulated by external forces or torques. I also present some results on the kinetics of denaturation of linear DNA and supercoiling of closed dsDNA molecules.

A good proportion of the classical chemical and physical experimental studies of DNA have been carried out by X-ray diffraction [34, 83, 84], or via experiments which probing the behaviour of a solution of DNA molecules [85]. X-ray diffraction gives structural information of DNA which is sparse in comparison to the numerous configurations accessible to DNA during biological processes. In bulk experiments (probing the behaviour of DNA in solutions), several molecules are sampled simultaneously and therefore there is inherent averaging in the measurements. In the past two decades this situation has changed dramatically. Single molecule experiments (SME) [46, 51, 55, 59, 86–90] are being used to exert and measure mechanical forces and torques on single molecules, or to passively track the motion of individual molecules using fluorescent tags [91].

Notwithstanding these noticeable improvements, the direct experimental examination of DNA molecules still suffer from many limitations. Especially when trying to resolve time-dependent processes or the resolution of more complicated

DNA structures. In light of this, the formulation of accurate theoretical and computational models that can efficiently capture the behaviour of DNA, either *in vivo* or *in vitro*, is of great importance in order to understand a number of outstanding biological problems.

Here I describe the principles of a particular computational method known as Molecular Dynamics (MD) simulations following reference [92]. Based on it, I explain the details of the implementation of a single nucleotide resolution coarse-grained model for dsDNA.

3.1 Molecular dynamics simulations

In MD simulations, given an initial configuration (positions and velocities) of a set of atoms representing the system, their temporal evolution is computed by integrating their equations of motion. In practice, the MD method follows the laws of classical mechanics:

$$\begin{aligned}\frac{d\mathbf{r}_i}{dt} &= \mathbf{v}_i \\ \mathbf{F}_i &= m_i \frac{d\mathbf{v}_i}{dt} = -\nabla \mathcal{U}_i,\end{aligned}\tag{3.1}$$

where $\nabla = \frac{\partial}{\partial \mathbf{r}}$. The mass, position and velocity of the i -th particle are m_i , \mathbf{r}_i and \mathbf{v}_i respectively. \mathcal{U}_i is the potential representing all the internal and external interactions on particle i . Due to this, the i -th atom feels a force \mathbf{F}_i that governs its dynamics.

This means that some potentials that describe the interactions between the particles must be set. If chosen properly, these potentials will reflect accurately some properties of the system. That is why it has to be clear which aspects of the system are of particular interest. In this thesis I am focusing on the geometrical and elastic properties of dsDNA in the B-form.

3.1.1 Brownian dynamics simulations

In a MD simulation the solvent can be modelled explicitly as a set of additional particles. Although more realistic, this method is computationally expensive.

The Langevin formalism is an efficient way of incorporating an implicit solvent and dissipative forces into the deterministic Newton's equations (3.1). In this formalism, the force acting on the solute (Eq. (3.2)) is written as the sum of a viscous force proportional to the particle's velocity, and a noise term representing the effect of the collisions with the molecules of the fluid. The method that considers MD in addition to stochastic and thermostat is often referred as to Brownian Dynamics (BD). Therefore, in this type of simulations, the motion of the particles is no longer deterministic but stochastic and can be written as

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\gamma_i \frac{\partial \mathbf{r}_i}{\partial t} - \nabla \mathcal{U}_i + \sqrt{2k_B T \gamma_i} \eta_i. \quad (3.2)$$

In this equation γ_i is the friction coefficient of the i -th particle and η_i is the stochastic white noise term which satisfies $\langle \eta_\alpha(t) \eta_\beta(t') \rangle = \delta_{\alpha\beta} \delta(t - t')$ along each Cartesian coordinate represented by the Greek letters. The form of the last term in Eq. (3.2) ensures that the equipartition theorem (which states that in thermal equilibrium any degree of freedom of the system which appears quadratically in the energy has an average energy of $1/2k_B T$) is satisfied.

In the over-damped limit, for which the left-hand side of Eq. (3.2) is zero, the Langevin equation can be rewritten as

$$\frac{\partial \mathbf{r}_i}{\partial t} = -\frac{1}{\gamma_i} \nabla \mathcal{U}_i + \sqrt{2D_i} \eta_i, \quad (3.3)$$

where the diffusion coefficient is $D_i = k_B T / \gamma_i$. Eq. (3.3) is a good approximation of Eq. (3.2) only after an inertial time $t > \tau_i = m_i / \gamma_i$. This time is identified as the necessary time for the temperature to equilibrate across the system, or for the kinetic energy to be redistributed.

One of the limits of the Langevin formalism is that the hydrodynamic effects are neglected. These are important in systems with high viscosity, in which the solvent flow induced by the motion of one solute particle affects the motion of nearby solute particles. We will neglect this effects for computational efficiency.

3.1.2 Coarse-Graining

Several fully atomistic models for double-stranded DNA are available in the literature [93–95]. While these give an accurate description of the dynamics of DNA molecules and their interaction with single proteins, the complexity of the all-atom approach places severe limits on the size (up to about a hundred base-pairs) and time scales (of the order of μs) which can be probed [96]. Coarse-graining, where large collections of atoms or molecules are represented by single units, allows larger systems to be simulated for longer at the expense of molecular detail. One of the most challenging aspects in designing a computational model is to retain the key microscopic details necessary to answer a given question while “trimming” the rest.

At the large scale limit, entire eukaryotic chromosomes can be modelled using simple bead-and-spring polymer models [24, 97], where each monomer can represent up to 3000 bp and the simulated time can reach time-scales spanning minutes [97] or even several hours [98]; similar chains of beads can also be used to model naked DNA, though clearly such an approach neglects microscopic details such as the base-pair specificity or the double-stranded structure. While in some cases these models can still capture the essential physics [99], in others they are only a crude approximation of the real systems.

Several mesoscopic models have recently been proposed which aim at bridging the gap between the “all-atom” and “bead-spring” limits [100, 101]. Very successful examples of this are the oxDNA code introduced in Ref. [102], and the 3SPN.1 and 3SPN.2 codes [103]. While the latter has three spheres per nucleotide (making it computationally expensive), the former in his original version uses non-standard potentials and is not highly efficient and parallel. In addition, the recent implementation of oxDNA [3] into the LAMMPS engine lacks the ability to account for interactions between non-DNA particles and specific sites of nucleotides (phosphate or base sites): it is only possible to set interactions between non-DNA particles and the centre of mass of the nucleotides.

Therefore, a coarse grained model able to retain the necessary physical microscopic details while allowing simulations of several tens or hundreds of kilo-base pairs (kbp) that would be useful to address many biologically relevant questions, is still currently needed. Examples of biological processes for which

such a mesoscopic approach would be highly valuable can be classified in two broad categories: processes where DNA is mechanically manipulated by enzymatic machines (for example during replication or transcription which require opening of the double-helical structure), or processes where interactions between DNA and proteins depend more subtly on the topological and geometrical properties of the double-helix.

In the following section I present the construction of a coarse grained model which retains several of the B-DNA features mentioned in the previous chapter, while being able of delivering large scale simulations. The model is implemented in the Large-scale Atomic Molecular Massively Parallel Simulator (LAMMPS) engine [104] and it is fully parallel, highly scalable, portable to supercomputers and able to reach the length and time scales needed to study the under explore processes mentioned above, some of which will be discuss later in this thesis.

3.2 The model

We start by considering a complex made of two spherical monomers (see Fig. 3.1(a)), one of which represents the sugar-phosphate backbone of the DNA (this is referred to as a “bead” hereafter, and shown in blue), while the other represents the nitrogenous base (“patch” hereafter, shown in pink), and is placed at a distance of 0.5 nm from the bead centre. Beads have an excluded volume so that they cannot overlap, whereas patches have no associated excluded volume. In order to see the resolution of the system, a nucleotide structure lying inside the bead is shown in black in Fig. 3.1(a). Depending on the relevant features of the system to be modelled, a second patch representing the phosphate group may also be included explicitly to more accurately represent the DNA steric hindrance. When this second patch is not included, we imply that the phosphate is sitting 0.5 nm from the bead centre but slightly away from the antipodal point to the patch, marked in yellow in Fig. 3.1(a) for clarity.

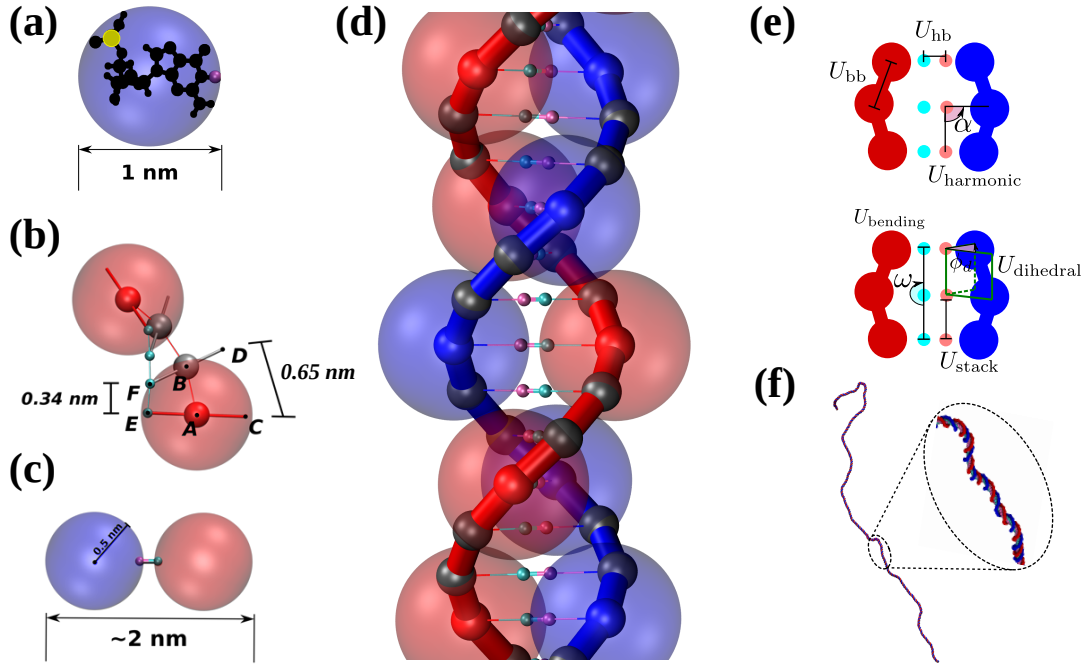


Figure 3.1 (a) The level of coarse-graining of the model is here summarised by encapsulating the atoms forming one nucleotide into one bead-patch complex. The small yellow sphere represents the position of the phosphate with respect to the complex, while the pink sphere denotes the position of the hydrogen bond between bases. The blue sphere approximates the excluded volume of the nucleotide. (b) This panel shows the main interaction sites between consecutive beads in the same strand. The equilibrium distance between patches (E-F) is set to 0.34 nm while the one between beads centres (A-B) to 0.46 nm. This leads to an equilibrium distance of 0.65 nm between the external edge of the backbone (C-D). These distances are set so that the correct pitch of 10 bp is recovered. (c) Two nucleotides are bonded via a breakable harmonic spring. Their distance is set so that the full chain thickness is around 2 nm, as that of B-DNA. (d) Representation of the double-stranded DNA model. Each ssDNA chain shows the beads which interact sterically (solid red and blue) as well as the phantom beads (solid grey) which we refer to in the main text. The faded red and blue spheres represent the steric interaction volume (set through the potential in Eq. (3.4)) of the interacting beads with the same colour. (e) The diagrams indicate the particles involved in each of the interactions in the model. (f) An example of an equilibrated configuration of a 1000 bp double-stranded DNA molecule.

The excluded volume between all possible pairs of beads (we anticipate to the reader that not all the beads have the same excluded volume. The reason is explained below) is modelled via a truncated and shifted Lennard-Jones (LJ) potential, which has the following form:

$$U_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma_s}{r} \right)^{12} - \left(\frac{\sigma_s}{r} \right)^6 + \frac{1}{4} \right], \quad (3.4)$$

for $r < r_c$ and 0 otherwise. Here r_c is the interaction cut-off, σ_s represents the diameter of a spherical bead, ϵ parametrises the strength of the repulsion and r is the Euclidean distance between the beads. In order to model purely repulsive interactions we set $r_c = 2^{1/6}\sigma_s$.

Each bead-patch complex represents a single nucleotide, and acts as a rigid body; we connect a chain of these bodies via finite extensible non-linear elastic (FENE) bonds of length $d_{bp} = 0.46$ nm between the beads (A-B in Fig. 3.1(b)) to represent one strand of DNA. This interaction is set by the potential:

$$U_{bb}(r) = \begin{cases} -\frac{\epsilon_{bb}R_0^2}{2} \ln \left[1 - \left(\frac{r}{R_0} \right)^2 \right] & \text{if } r < R_0 \\ \infty, & \text{if } r \geq R_0, \end{cases} \quad (3.5)$$

where R_0 is the maximum bond length, ϵ_{bb} is the spring constant and r is the Euclidean distance between bead i and bead $i+1$. When summed to the Lennard-Jones potential, the minimum of this potential is located at $r_{\min} = 0.96 \sigma_s$.

We set the distance between two consecutive patches along the strand (E-F in Fig. 3.1(b)) at 0.34 nm by means of a Morse potential:

$$U_{\text{stack}}(r) = \epsilon_m [1 - e^{-\lambda(r-r_{0,\text{stack}})}]^2, \quad (3.6)$$

where $r_{0,\text{stack}}$ is the equilibrium distance. The depth and width of this potential are represented by ϵ_m and λ respectively. With this we mimic the base-stacking distance of DNA.

The difference between the lengths A-B and E-F in Fig. 3.1(b) implies that the distance between the implicit phosphates at the external edge of the beads (C-D in Fig. 3.1(b)) is $d_{ph} = 0.65$ nm as accurate for B-DNA [4] (considering that the backbone distance fluctuates from its equilibrium value 0.6 ± 0.05 nm). While the ratio between d_{bp} and d_{ph} is well known to crucially regulate the correct pitch of the chain [4], the right-handedness is imposed using a dihedral potential between the quadruplets of monomers forming two consecutive nucleotides (A, E, F and B in Fig. 3.1(b)), and it has the form

$$U_{\text{dihedral}}(\phi_d) = \epsilon_d [1 + \cos(\phi_d - d)]. \quad (3.7)$$

This potential regulates the angle ϕ_d between the planes A-E-F and E-F-B. The phase d is related to the equilibrium helical pitch and ϵ_d is the energy of the interaction. The minimum of this potential is set equal to 36° , so as to match the geometry of a regular B-DNA helix.

Nucleotides belonging to different strands are bonded together via breakable harmonic springs acting between two patches and representing hydrogen bonds (see Fig. 3.1(c)). The equilibrium bond distance $r_{0,\text{hb}}$ is set to zero; if the extent of the bond increases beyond a critical value $r_{c,\text{hb}} = 0.3$ nm, the bond breaks, modelling the denaturation of the chain. Therefore this potential reads:

$$U_{\text{hb}}(r) = \frac{\epsilon_{HB}}{2(r_{0,\text{hb}} - r_{c,\text{hb}})^2} \left[(r - r_{0,\text{hb}})^2 - (r_{c,\text{hb}} - r_{0,\text{hb}})^2 \right] \quad (3.8)$$

if $r \leq r_{c,\text{hb}}$, and 0 otherwise. Here ϵ_{HB} represents the spring constant and r is the distance between patches i and i' .

In order to limit the splay of consecutive nucleotides and set the roll and slide to the values found in experiments for B-DNA (see Table 2.1), we used a stiff harmonic potential so as to keep the angle α between particles E, F and B (two patches and one bead in Fig. 3.1(b)) at 90° . This interaction imposes the planarity between consecutive bases in the same strand and has the form:

$$U_{\text{harmonic}}(\alpha) = \frac{\epsilon_h}{2} (\alpha - \alpha_0)^2, \quad (3.9)$$

with strength ϵ_h and minimum at α_0 . Finally, the last ingredient of this model is a Kratky-Porod potential regulating the angle ω between three consecutive patches along one strand.

$$U_{\text{bending}}(\omega) = \epsilon_b [1 + \cos(\omega)], \quad (3.10)$$

This allows us to finely regulate the chain stiffness by tuning the energy ϵ_b .

The excluded volume around each bead depicted in Fig. 3.1(d) (transparent red and blue spheres) has diameter 1 nm. Since we use spherical beads rather than asymmetrically shaped ones (this is important for the speed of the algorithm), the geometry of the double-strand depicted in Fig. 3.1(b) and (d) would involve a large degree of overlapping which would lead to a large steric repulsion. To

avoid this we consider two types of beads in each strand: sterically interacting beads (shown as small solid red and blue spheres for each strand in Fig. 3.1(d)) are intercalated by two ghost beads (depicted as small grey spheres) which do not interact sterically along the same strand but they do interact with all the beads on the complementary strand with an excluded volume of 0.5 nm. This choice ensures that only non-overlapping beads sterically interact with one another. In addition, this allows us to preserve the correct thickness of the chain (2 nm for B-DNA), to maintain the desired distance between contiguous nucleotides and avoid the strands crossing through one another.

The diagram in Fig. 3.1(e) indicates all the interactions set through the potentials in the model. The lines connect the particles involved in each interaction. In addition, the angles α , ω and ϕ_d , corresponding to the potentials U_{harmonic} , U_{bending} and U_{dihedral} are shown.

The parameters for each potential are reported in simulation units in the following table.

Interaction	Parameters
Backbone: U_{bb}	$\epsilon_{\text{bb}} = 30$, $R_0 = 0.6825$, $\epsilon = 1$ and $\sigma_s = 0.4430$
Hydrogen bond: U_{hb}	$\epsilon_{\text{HB}} = 6$, $r_{0,\text{hb}} = 0$ and $r_{\text{c,hb}} = 0.3$
Steric: U_{LJ}	$\epsilon = 1$ and $\sigma_s = 1$
Dihedral: U_{dihedral}	$\epsilon_d = 50$ and $d = -144^\circ$
Stacking: U_{stack}	$\epsilon_m = 30$, $\lambda = 8$ and $r_{0,\text{stack}} = 0.34$
Planarity: U_{harmonic}	$\epsilon_h = 200$ and $\alpha_0 = 90^\circ$
Bending: U_{bending}	$\epsilon_b = 52$

Table 3.1 Parameter values in the model and expressed in simulation units.

Mapping the simulation units to physical ones can be done by setting the fundamental units: distance, energy and time. These are shown in Table 3.2. The chosen system of reference is a bath at room temperature $T = 300$ K and with the viscosity of water $\nu = 1$ cP.

Parameter	Experimental units
Distance (σ_s)	1 nm \simeq 3 bp
Energy ($\epsilon = k_B T$)	4.1419×10^{-21} J
Force ($F = \epsilon/\sigma_s$)	4.1419×10^{-12} N
Mobility ($\mu_0 = 1/(3\pi\nu\sigma_s)$)	1.06×10^{11} m/Ns
Diffusion ($D = \mu_0 k_B T$)	4.39×10^{-10} m ² /s
Time ($\tau_{Br} = \sigma_s^2/D$)	2.28×10^{-9} s

Table 3.2 Mapping between simulation and physical units.

Finally, the numerical integration of Eq. (3.2) is performed by LAMMPS in an NVT ensemble, using a standard velocity-Verlet algorithm with integration time-step

$$\Delta t = 0.005\tau_{Br}. \quad (3.11)$$

We should stress here that the model as presented in this section should be thought of as a simple, starting point, which is based on some crucial geometric constraints of double-stranded DNA. One of the main strength of the model is the ability to deliver large-scale simulations, which is achieved by using spherical monomers that interact via standard potentials that have already been implemented and optimized in LAMMPS. These potentials ensure a highly scalable performance in large scale parallel simulations. As an example, in Fig. 3.1(f) we show a typical equilibrated configuration using the presented model for a 1000 bp molecule.

From now on, I will only show results for the model without the explicit presence of the phosphates (yellow bead in Fig. 3.1(a)) unless otherwise stated. At the same time, there are several features of dsDNA that the model (as presented up to now) does not include. Some examples are: (i) the distinction between minor and major groove; (ii) the description of electrostatic effects due to charges on the DNA, and to the variation of the density of counter-ions in solution (our parameters are tuned assuming room temperature and a physiological buffer, 0.15 M NaCl); and (iii) the effect of sequence specificity (in this simple description, our dsDNA is viewed as a homopolymer). Such effects will be important, for instance, when one needs to more faithfully describe interactions between DNA and DNA-binding proteins. It is in principle possible to include these effects in a modular fashion in our framework, and in section 3.6 (see also Appendix A.1) we

describe how we can account for (i) in a simple way, and how (ii) and (iii) might be implemented in the future.

It is worth mentioning that in the current version of the model, hydrogen bonding is the only interaction responsible for holding the two DNA strands together. The rest of the potentials are defined independently for each strand. As a consequence, when the model was tested for single stranded DNA (ssDNA), and the persistence length was computed, its value (30 nm) was higher than expected from experiments [105] ($\approx 1 - 2$ nm). To model ssDNA and reproduce this dramatic change in flexibility, we impose that, once the hydrogen bond keeping the two strands together breaks, the energies of the harmonic (ϵ_h), dihedral (ϵ_d), stacking (ϵ_m) and bending (ϵ_b) interactions acting on each individual strand, should decrease from the values shown in table 3.1 to 20, 0, 3 and 2 $k_B T$ respectively. In this way, we effectively take into account of the larger flexibility of single DNA strands and, in particular, we observe a persistence length of about 1 nm (see section 3.5). This denaturation protocol in the simulations implies changes in the energy of the system. However, as mentioned in section 3.1.1 we performed thermostat simulations in which fluctuations of the energy are allowed, providing that the temperature of the system is kept constant.

3.3 Parameterisation

Our model has several parameters which can be varied to control the pitch, bending and torsional properties of the simulated DNA molecule. Nonetheless, we are interested in modelling the B-form of dsDNA, of which two main physical properties are: the persistence length $l_p = 53$ nm $\simeq 156$ bp [46], and the torsional rigidity $C/k_B T \simeq 60 - 80$ nm $\simeq 177 - 235$ bp [59, 87, 89]. Due to the interplay between the potentials presented in the previous section, there is no simple mapping between individual simulation parameters and the resulting physical properties; instead we obtain a simulated molecule with the correct values of l_p and C via a systematic tuning of the parameters. In this section we measure these properties from the microscopic positions of the beads in equilibrated DNA molecule configurations. Then in the following section, we use the parametrised force field to simulate single-molecule experiments and show that the DNA molecules display the correct macroscopic response to mechanical manipulations.

3.3.1 Persistence length

The persistence length of dsDNA is a well-studied physical property that plays an important role in many biological processes. This quantity can be readily measured by computing the tangent-tangent correlation (Eq. (2.16)) for a discrete rod. In our model, we use the position of the patches to extract the centreline of the dsDNA molecule, where the tangent at the n th patch at position $\mathbf{r}(n)$ is $\mathbf{t}(n) \equiv (\mathbf{r}(n+1) - \mathbf{r}(n))/|\mathbf{r}(n+1) - \mathbf{r}(n)|$. One can compute the tangent-tangent correlator along this curve and obtain the persistence length by extracting the length scale associated with the exponential decay. In order to avoid finite-size effects due to the presence of ends, we neglect the two terminal segments (~ 5 bp at each end). The resulting curve (shown in Fig. 3.2(a)) is adjusted by tuning the parameters of the model until the exponential fit returns a persistence length of $l_p \simeq 49 \pm 2.3$ nm = 143 ± 7 bp, in agreement with the experimentally observed values ($l_p = 53$ nm [46]).

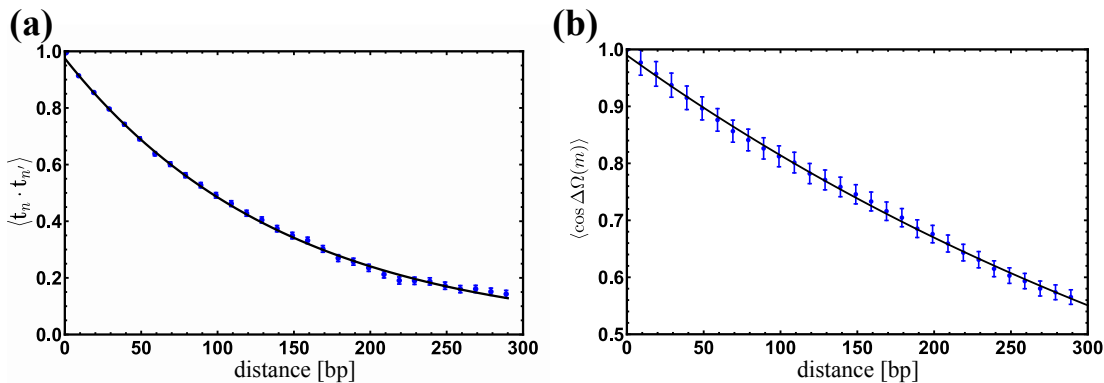


Figure 3.2 (a) The tangent-tangent correlator $\langle \mathbf{t}(n) \cdot \mathbf{t}(n') \rangle$ computed for a chain 300 bp long; it shows an exponential decay as in Eq. (2.16) with a correlation length $l_p = 143 \pm 7$ bp. Points show correlations measured from the simulations (average over time and ensemble), and the line shows a fit to Eq. (2.16). Error bars give the standard error of the mean. (b) The average of the cosine of the total twist angle $\Delta\Omega(m)$ is computed for a chain 300 bp long; in this figure we show the correlator to decay exponentially as in Eq. (3.12) with a characteristic length $l_\tau = 512 \pm 18$ bp. Data points are obtained from simulations while the line is an exponential fit. Error bars give standard errors of the mean.

3.3.2 Torsional persistence length

The behaviour of DNA when twisted is regulated by its torsional rigidity. There are several well known examples in which this property is crucial for important biological processes, such as transcription and gene expression [106, 107]. Furthermore, the high torsional stiffness of DNA molecules implies that, when placed under torsion, they preferentially bend, thereby creating writhe and plectonemes [4]. In order to correctly take this feature into account, it is therefore crucial to account for bending and torsional rigidities [87].

Here we measure the torsional persistence length by computing the decorrelation of the twist angle (see section 2.2.2) between frames along the DNA molecule. Each local frame is specified by the tangent vector $\mathbf{t}(n)$ as defined above, a normal vector $\mathbf{f}(n)$, defined as the projection of the vector connecting two beads in a base-pair onto the plane perpendicular to $\mathbf{t}(n)$, and a third vector $\mathbf{v}(n) = \mathbf{t}(n) \times \mathbf{f}(n)$, the binormal, which is perpendicular to both $\mathbf{t}(n)$ and $\mathbf{f}(n)$.

The Euler angles between the frames at n and $n + 1$ can be used to obtain the twist increment between those base-pairs, and the correlation function for the total twist between m consecutive base-pairs $\Omega(m)$. Since dsDNA has an equilibrium twist angle $\theta_0 = 36^\circ$ per bp, we subtract this out, and calculate the correlation for the residual twist $\Delta\Omega(m) = \Omega(m) - m\theta_0$. As in section 2.2.2, the average cosine of the residual total twist between any two reference frames separated by m bases exhibits an exponential decay analogous to Eq. (2.28):

$$\langle \cos \Delta\Omega(m) \rangle = e^{-ma_0/2C}, \quad (3.12)$$

where $a_0 = 0.34$ nm and we define $l_\tau = 2C$ the characteristic torsional correlation length. We obtain $\langle \cos \Delta\Omega(m) \rangle$ by simulating a 300 bp long dsDNA molecule and by averaging the value over time and ensemble (meaning independent configurations). The curve obtained is shown in Fig. 3.2(b) on top of which we show the fitted exponential. Again, we tune the parameters of the model so that the curve displays a characteristic decay length $l_\tau = 512 \pm 18$ bp $\simeq 174 \pm 6$ nm, which is consistent with experimental estimates valid for the B-form of dsDNA [59, 87, 89].

3.4 Validation through simulated single molecule experiments

In this section we reproduce the conditions of two different experiments, in order to test the response of our model DNA to stretching and twisting. This also provides us with an independent method to evaluate its persistence length and torsional rigidity at the macroscopic scale. In the following, we therefore keep the parameters of the model fixed at the values used in the previous section, and do not further tune them to achieve the experimentally known behaviours but simply validate the model as it is through its response to mechanical stress.

3.4.1 Response to stretching

The classic elastic response of DNA to an external stretching force \mathbf{F} is known to be well described by the inextensible WLC model (Eq. (2.22)) for forces up to 10 pN. In order to test if our model can reproduce this result, we performed simulations in which a constant pulling force directed along \mathbf{e}_z and acting on the last base pair of the dsDNA was applied, while the other end of the molecule was anchored at a surface (see Fig. 3.3(a)).

The force-extension curve measured for a chain 300 bp long is reported in the inset of Fig. 3.3(b) as data points, while the solid curve is the fit to the WLC result of Eq. (2.22). The fitting results in values for both L and l_p , that we can compare with the values obtained in the previous sections and set by the parameters of our model. In particular for a 300 bp chain we obtain $L = 100.3 \pm 1.7$ nm (which gives a bp step size of 0.33 ± 0.01 nm) and $l_p = 47 \pm 2$ nm $\simeq 140 \pm 7$ bp. The results are therefore in good agreement with the calculation and the tuning of the persistence length performed in the previous section which gave $l_p = 49$ nm.

When the applied force is greater than 10 pN the existence of a finite stretching modulus (ζ) has to be taken into account. This is done by the Extensible WLC model described by Eq. (2.25). The force-extension curve from the simulations of a 300 bp chain in this regime, corresponds to the data points above 10 pN shown in blue in Fig. 3.3(b). Fitting these points with Eq. (2.25) gives $L = 99.7 \pm 0.5$ nm, $l_p = 60.2 \pm 2$ nm and $\zeta = 2086 \pm 23$ pN. This value of L is the one expected

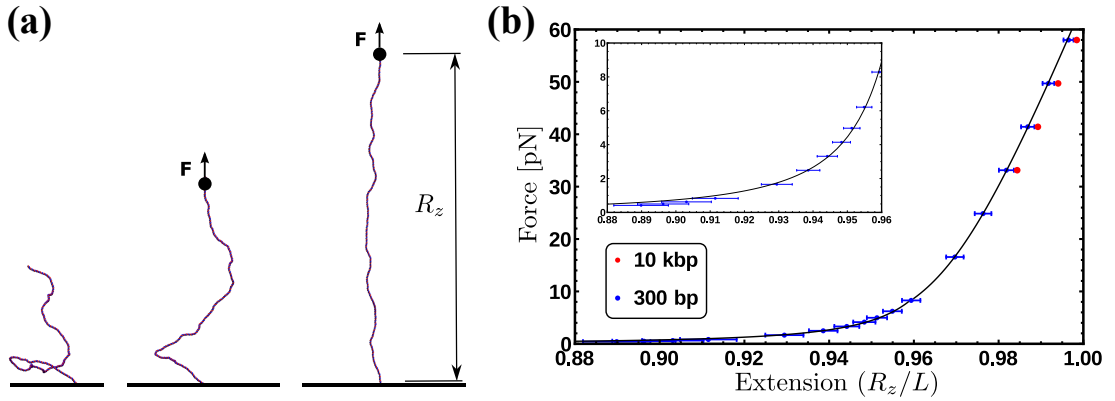


Figure 3.3 (a) I simulate single-molecule experiments by anchoring the dsDNA to a surface at the bottom end while stretching it with a constant force \mathbf{F} from the top end. I then monitor the end-to-end elongation along the z -direction, R_z , and report its equilibrium value for a given force. (b) Force-extension curve from the simulation (data points) of two different length chains: 300 bp (blue) and 10 kbp (red). The inset data (low force regime) is fitted by the function in Eq. (2.22) (solid line) and the data above 10 pN is fitted with Eq. (2.25). For the WLC the free parameters for the fitting are the total polymer length L and the persistence length l_p , both of which are in agreement with the fixed parameters of the model (see text). For the EWLC in addition to the previous parameters the stretching modulus ζ is found.

for a chain made by 300 bp. The value for the persistence length l_p is slightly bigger than the one observed in the WLC regime; on the other hand, this apparent increase of l_p is also observed in experiments [105]. Finally, the stretching modulus ζ is found to be twice the one expected for real dsDNA (~ 1000 pN), although this discrepancy will not be critical to the following.

We also performed simulations of a stretching experiment on a chain ten thousand base-pairs long (comparable to viral P4 DNA). The results for this case are reported as red data-points in Fig. 3.3(b). These measurements are in agreement with the behaviour of the 300 bp-long chain within error-bars, although they systematically show a slightly larger extension, possibly due to finite-size effects.

At forces of 65 pN or more, dsDNA changes its form dramatically [105] and it has been observed to stretch up to 70% beyond the canonical contour length shown by its B-form. This is not currently reproduced in our model and it would require a change in the structure of how the base-pairs are arranged and stacked together (i.e. the distance between base-pairs would no longer be 0.34 nm).

However, such an extreme structural transition has not been observed either *in vivo* or *in vitro*, and probably is not the relevant scenario for biological functions.

3.4.2 Response to twisting

The torsional stiffness of DNA can be calculated by computing the twist response of dsDNA to an imposed external torque, for instance applied by a magnetically controlled macroscopic bead [71, 87] (see Fig. 3.4(a)). For different magnitudes of the applied torque, $|\mathbf{\Gamma}|$, we compute the superhelical density, σ . Although the chain we use is not closed into a loop, and therefore it is not possible to formally define a linking between the strands, it is possible to compute the linking number between two “artificially” closed strands [108, 109] which follow the paths of the DNA strands along the chain backbone and then join the respective ends far away from the molecule (see Appendix A.2). Furthermore, the molecule is kept straight (writhe-free) by applying a constant stretching force which ensures that the twist is very close to the computed linking number.

By measuring the deviation of linking number ΔLk from the equilibrium value $\text{Lk}_0 = \text{Tw}_0 = N/p$ (see section 2.3), we can then readily obtain σ . With this information it is possible to map out the response curve of the molecule to an external torque. A feature of this is a linear regime for small $|\sigma|$ which we recover (see Fig. 3.4(b)). The torsional rigidity, C , can finally be calculated (in the limit of large stretching forces [101]) using Eq. (2.31). The data points shown in Fig. 3.4(b) are obtained from simulations of a 600 bp long chain anchored at a surface to one end, while the other end was pulled by a constant force of 16 pN and different applied torques, $\Gamma = \mathbf{\Gamma} \cdot \mathbf{e}_z$. From the fit we get the value of torsional persistence length $C \sim 88 \text{ nm} \simeq 260 \text{ bp}$ in good agreement with experimental results [41, 110, 111].

One can finally use the relation between the torsional persistence length l_τ and the torsional stiffness C obtained from the twistable worm-like chain theory (see section 2.2.2), which gives $l_\tau = 2C \sim 176 \text{ nm}$, in agreement with the measurement performed in the previous section (yielding $l_\tau = 174 \pm 6 \text{ nm}$).

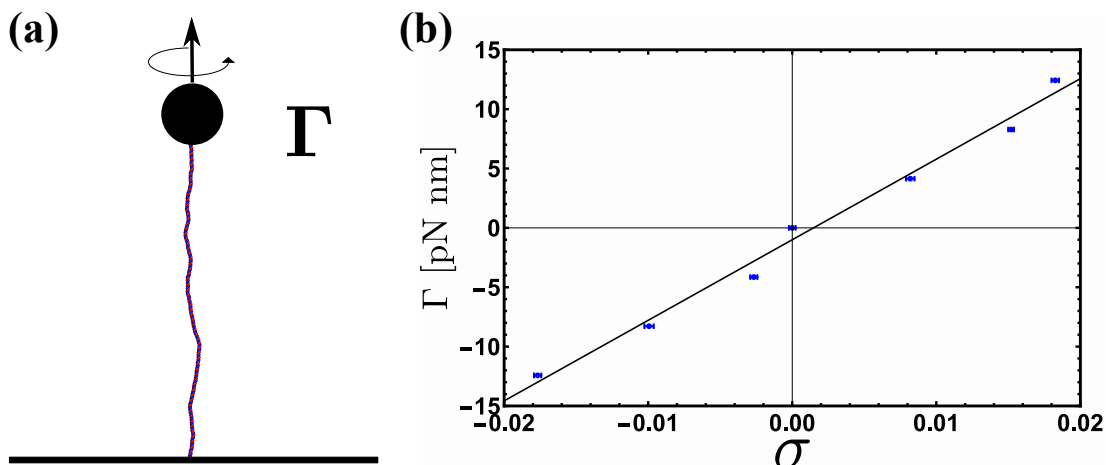


Figure 3.4 (a) The DNA is anchored to a surface at the bottom end while being stretched with a constant force \mathbf{F} , and a torque Γ is applied at the top end. I then compute the linking number and report its equilibrium value for a given torque. With this information it is possible to compute the superhelical density. (b) Response to torque experiment for a chain 600 bp long pulled with an external stretching force of 16 pN. Here we show the linear regime for small $|\sigma|$ which gives the torsional rigidity C by a linear fit with Eq. (2.31).

3.5 Denaturation and supercoiling in the CG-DNA model

It is well known that local denaturation has several biological implications such as favouring transcription initiation, DNA repair or recombination [106, 112, 113], and that the dynamics of these denaturated segments of DNA (often referred to as “bubbles”), can be affected by torsional stress, which is itself often regulated by enzymes, such as RNA polymerases [114–116]. This fascinating interplay between the elasticity and biology of DNA has received much theoretical and experimental attention [41, 113, 116–120], but there have been remarkably few attempts to address it from a computational point of view [121, 122].

Although theoretical models can capture the thermodynamics of a “stress-induced DNA-duplex destabilisation” (SIDD) [123], elucidating the kinetics of such a process, under both equilibrium and out-of-equilibrium conditions, is an important question that can be addressed using computer simulations. In this section we show that our model can readily display DNA denaturation upon

decreasing the stiffness, ϵ_{HB} , of the spring connecting patches in the two strands (U_{hb}). While the most common experimental strategy to denature DNA consists in increasing the solution temperature, this pathway instead mimics a change in solution pH [124].

In Fig. 3.5(a) we show the fraction (ϑ) of denatured base-pairs for a linear DNA molecule as a function of time for three different choices of ϵ_{HB} . As the energy of the bond is decreased, we observe the unbinding of the two strands, which starts from the ends of the chain, as observed experimentally [125]. We then observed that the denatured region spreads to the middle of the molecule, finally melting the whole chain when $\epsilon_{HB} \leq 1.2k_B T$ and producing two single strands.

Single stranded DNA is much more flexible than its intact counterpart (experiments found $l_p \sim 1 - 2$ nm in ssDNA [105]). In order to mimic this behaviour in our model, we follow the denaturation protocol mentioned at the end of section 3.2 by changing the energies of the interactions between nucleotides which are part of a “bubble” larger than two base-pairs. This results in single strands with a persistence length of around 2 bp which are extremely flexible, as one can appreciate from the snapshots in Fig. 3.5(a).

As previously mentioned, another way to denature DNA is by increasing the temperature of the system. We performed simulations where this pathway was adopted to denature a linear DNA molecule and we found that the melting temperature is approximately 70°C which is somewhat close but below the experimentally observed melting temperature (87° – 93° [126]).

As a preliminary step to show that our model can readily take into account supercoiling, in Fig. 3.5(b) we present an example of simulated closed (ring) dsDNA. In particular, we consider a molecule 500 bp long and we initialise it with a linking number deficit of $\Delta Lk = Lk_0 - Lk = -3$ (47 turns instead of the usual 50 for a pitch of 10 bp). In a linear molecule this deficit would dissipate quickly by the free motion of the ends, whereas in a closed molecule the difference creates a negative supercoiling $\sigma = \Delta Lk / Lk_0 \simeq -0.06$ which is conserved throughout the dynamics. The supercoiling can then be distributed into the torsional or bending degrees of freedom as long as the White-Fuller theorem is satisfied (see Eq. (2.29)). Since the torsional stiffness of DNA is larger than the bending rigidity, a good

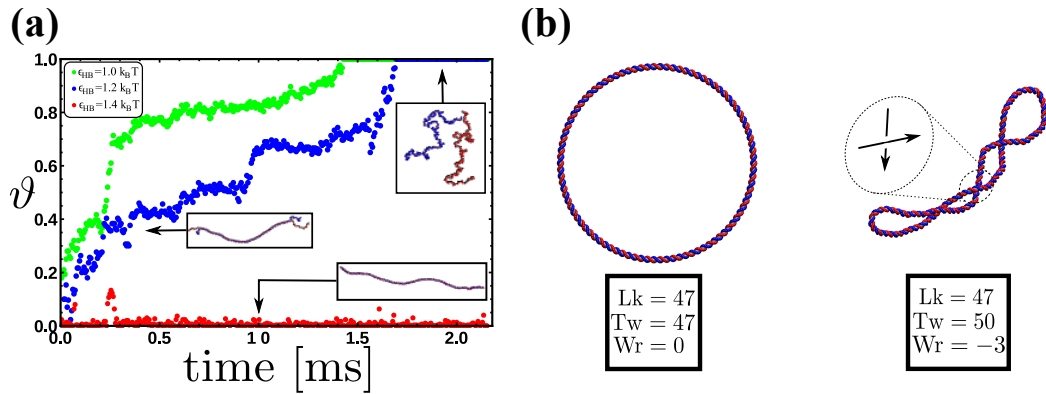


Figure 3.5 (a) This figure shows the fraction of denatured base pairs ϑ as a function of time and for different bond energies connecting the patches of paired bases, for a linear chain 300 bp long. Snapshots from simulations are also shown. The energies used range between $\epsilon_{HB} = 1.0 k_B T$ and $\epsilon_{HB} = 1.4 k_B T$. We always observe that in linear dsDNA the denaturation process nucleates from the ends, as suggested by experiments [125]. (b) This figure shows the relaxation of a negatively supercoiled circular dsDNA. (left) The molecule (500 bp long) is initialised as a perfect ring from which three full turns are removed. (right) As the system evolves, the linking deficit is converted into writhe, and the molecule assumes stable buckled configurations. This behaviour is expected for a real dsDNA molecule because the torsional stiffness is larger than the bending rigidity.

proportion of the twist deficit is quickly converted to writhe, as can be readily seen in Fig. 3.5(b) (as a first approximation, the number of crossings displayed in the figure is the magnitude of the writhe).

3.6 Future improvements and limitations

Our model aims at bridging the gap between all-atoms and coarse-grained models for dsDNA; while it is currently less sophisticated than other mesoscopic models [102, 103], most notably in the treatment of sequence specificity or hybridisation, this model exploits the scalability of LAMMPS, and is ideally suited to study problems such as DNA-protein interactions, or the denaturation of supercoiled DNA, where it is essential to consider long molecules, as well as to simultaneously model double-stranded and denatured regions.

We should highlight that the presented model has some limitations which arise from the compromise between accuracy and scalability. For instance, while our model can be used to study the reverse of partial denaturation by, non-

monotonically tuning the value of ϵ_{HB} or temperature, it cannot create hybridised molecules in which nucleotides partner up with nucleotides other than those to which they were bonded to start with. In other words, “secondary” structures and hairpins cannot be formed. Furthermore, as previously mentioned, the model does not include sequence specificity, which is known to affect the local dynamics of denaturation. We aim to address both these aspects in the future. In regard to sequence specificity, this can be accounted for straightforwardly by defining two types of harmonic bonds connecting patches in the complementary strands and by using springs with different stiffness such that $\epsilon_{HB}(AT) < \epsilon_{HB}(CG)$. Since stacking is also sequence-dependent we could, in a similar way, define different types of stacking (Morse) potentials with distinct parameters which can depend on the local sequence.

One of the improvements that can be readily made to the model is to account for the presence of major and minor grooves. Distinguishing between the two grooves may be important, for example, to capture the correct interaction of the chain with DNA-binding proteins, since these often bind selectively to one of the grooves. To address this issue, the model can be extended to include explicitly a phosphate group by means of a third monomer per nucleotide (see Appendix A.1 for details). We note that without additional parametrization, the model is able to display the presence of asymmetric grooves with a total length of 1.22 nm (for the minor groove) and 2.18 nm (for the major groove).

Another important aspect worth considering in the future is the electrostatic interaction of DNA, either with itself or the environment. This is neglected in the current formulation of the model for the sake of efficient scalability of the parallelised code. Therefore the parameter tuning in section 3.2 was carried out by implicitly assuming that the salt concentration corresponds to physiological conditions (0.15 M NaCl) and that the system is at room temperature (27° C); under these conditions as previously mentioned the persistence length is $l_p \approx 50$ nm. This, of course, will limit the range of applications of our model to systems where electrostatic properties are screened. Different approaches could be tested to address this aspect where needed. In Ref. [72], for example, a Debye-Hückel potential is used to model DNA-DNA interactions, with an effective charge located at the backbone sites and an interaction radius depending on the salt concentration. This additional force field can be easily added to our model. A

slightly different approach could be to capture the effects of screened electrostatic repulsion by modulating the effective thickness of the chain in a similar way to that proposed in Refs. [127, 128]. For the model we present here, it is possible to moderately modify the thickness of the chain by adjusting the excluded volume interaction between phosphates, when these are explicitly considered.

3.7 Summary

The interplay between the physics and biology of DNA is one of the most intriguing topics in biophysics. While computational models can strongly aid the understanding of this fascinating open problem, the computational resources for such an expensive task have traditionally been limited. Researchers often use either very detailed and accurate all-atoms models, which can only cover short time and length scales, or coarse-grained models, which can follow the evolution of the system for much longer times at the expense of neglecting key physical properties of dsDNA. Here, we have proposed a mesoscopic coarse grained model that can be readily implemented at minimal cost into LAMMPS, one of the most popular molecular dynamics engines for atomistic and mesoscopic simulations.

In summary, in this chapter we tuned the model in order to reproduce crucial physical features of dsDNA such as bending and torsional rigidities. We then tested it by simulating single-molecule experiments so as to independently check the parameterisation and the response of our model to external manipulation. Finally, we provided examples of denaturation and supercoiled DNA. We have shown that this implementation can comfortably reach length and time scales that are relevant to both single molecule and biological experiments, therefore making our model interesting for applications. In the next chapter, we use this model to investigate a biologically-inspired problem: the interplay between denaturation and supercoiling, especially in light of its connection to gene expression [106, 107].

4

Dynamical scaling and phase coexistence in DNA melting

Abstract: There is a long-standing experimental observation that the melting of circular DNA occurs in a larger range of temperatures than for linear DNA. The reason behind this observation has been extensively studied and debated in the past but a general consensus has not been reached yet. Here I combine Brownian Dynamics simulations (see section 3.1.1) and mean field theory, to show that the difference between the two cases is that the linking number between strands is conserved for circular, but not linear DNA. This topological constraint leads to phase coexistence between intact and denatured phases, hiding the first-order melting transition, so that it appears much smoother. I also find new topology-dependent scaling laws describing the growth of denaturation bubbles, which in principle could be tested via single molecule experiments.

In the 1960s two different groups in the same department (one led by Renato Dulbecco and the other by Jerome Vinograd), were independently investigating the properties of DNA from the tumour virus, polyoma [129]. They performed sets of sedimentation velocity (following those by Lionel Crawford [130]) and buoyant density experiments which resulted in a model for this particular system [131, 132]. The main result from the experiments is that polyoma DNA could exist into two forms which have different sedimentation coefficients (s_c , a quantity related to the size of a molecule, equal to the terminal outward velocity of the molecule

when centrifuged in a fluid medium divided by the centrifugal force acting on it. Sedimentation occurs faster for molecules with higher s_c): the fast ($21s_c$) form and the slow ($14s_c$) form. At neutral pH both forms had the same molecular weight and therefore the difference was related to their conformation, with the fastest expected to be more compact. In addition, when the pH was increased, the fast form was more difficult to melt than the slow form. To complicate the picture even more, later experiments by Dulbecco [131] showed that using DNase I to introduce a single break in the backbone of the DNA molecule, caused the fast form to convert into the slow form.

Earlier experiments carried on ssDNA rings from bacteriophage $\phi X174$, had shown that DNase I converts the ssDNA rings into ssDNA linear molecules [133]. The similarities with Dulbecco's results on polyoma, led him to think that the fast form had to be circular and covalently-closed with no ends, while the slow form was linear. However, electron microscopy experiments on polyoma [134] showed that there were no linear molecules to be seen, only dsDNA circular forms were present in the images of both the $21s_c$ and $14s_c$ systems. Since the main difference between the slow form (linear in the bacteriophage $\phi X174$ case and nicked-circular for the polyoma virus) and the fast form (covalent-closed DNA in both cases), is that the former does not preserve the linking number, while the latter does, it has been argued that topology plays a fundamental role in the behaviour of DNA.

The melting transition of DNA from one double-stranded helix to two single-stranded coils is a remarkable example of one of the processes governed by DNA topology. Experiments [42, 135], theories [136, 137] and simulations [138] have shown that the "helix-coil" transition in linear or nicked (topologically free (tf)) DNA molecules, is abrupt and bears the signature of a first-order-like transition. In this case, the fraction of unpaired base-pairs jumps from 0 to 1 over a very short range of temperatures (see Fig. 4.2(a)). On the other hand, the same transition is much smoother, i.e. that occurs over a larger range of temperatures, for circular-closed (topologically constrained (tc)) DNA [42, 139]. Intriguingly, and in evident contrast with the behaviour of tfDNA, the width of the melting transition of tcDNA is relatively insensitive to the precise nucleotide sequence [139] thereby suggesting that a universal physical mechanism, rather than biochemical details, may underlie the above mentioned broadening. While

biophysical theories of tcDNA melting do exist, they do not reach a consensus as to whether the transition should weaken or disappear altogether [113, 118, 140–142]; more importantly, the physics underlying the broadening of the transition remains unclear.

In this chapter I shed new light in this issue by employing a combination of complementary methods. First, I perform large-scale coarse-grained BD simulations of 1000 base-pairs long tf and tc double-stranded DNA molecules during melting, using the model introduced in the previous chapter. These simulations predict topology-dependent melting curves in quantitative agreement with experiments. Second, I propose and study a phenomenological Landau mean field theory which couples a critical “denaturation” field (ϕ) with a non-critical “supercoiling” field (σ). This captures the interplay between local DNA melting and topological constraints, and predicts the emergence of phase coexistence within a wide temperature range (for the tcDNA), in line with the simulations. This provides a generic mechanism to explain the observed experimental broadening of the melting transition in tcDNA.

4.1 Melting curves

A melting curve reports the dissociation of the dsDNA into ssDNA after a step change in the solution conditions (temperature (T), pH or salt concentration) that effectively reduce the strength of the H-bonds (ϵ_{HB} in our model, see section 3.5) in between nucleotides.

An observable from our simulations that can be directly compared with experiments is the fraction of denatured base-pairs at equilibrium, $\langle\vartheta\rangle$. To measure it in practice, we start from equilibrated dsDNA molecules in non-denatured states, such that $\vartheta(t=0) = 0$. All these samples are either tf (linear) or tc with different levels of supercoiling. After a sudden quench of ϵ_{HB} , corresponding to a sudden increase in pH, we record the time evolution of the system until a new steady state, characterized by a constant value of $\vartheta(t)$ is reached (examples are shown in Fig. 4.1).

For topologically constrained molecules, $\vartheta(t)$ reaches a steady state value between 0 and 1 (see red, green and blue curves in Fig. 4.1). On the other

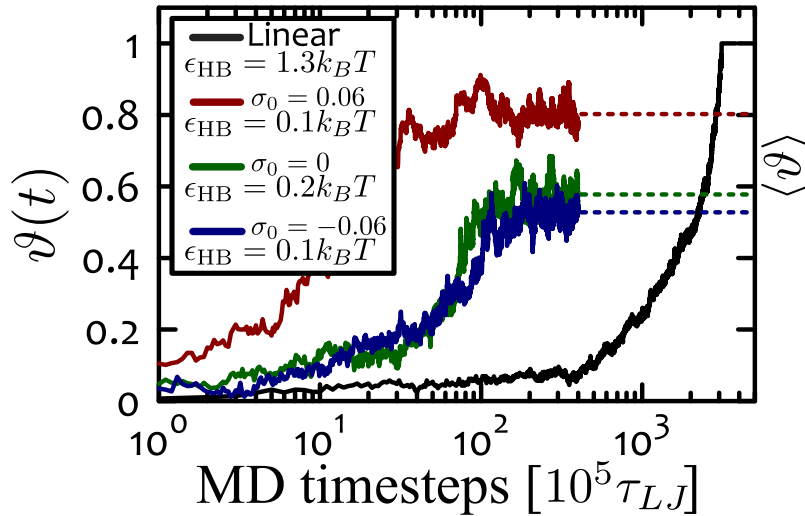


Figure 4.1 *Dynamics of melting.* Examples of trajectories obtained during melting of the model dsDNA in BD simulations. The curves represent the fraction of denatured base-pairs at a given time t for linear (black line) and circular closed molecules with supercoiling $\sigma_0 = 0.06, 0, -0.06$ plotted in red, green and blue, respectively. At $t = 0$ the value of the effective bond energy was changed from $6k_B T$ (see section 3.2) to $0.1, 0.2$ and $0.3k_B T$, respectively. The dashed lines represent the equilibrium values $\langle \vartheta \rangle$ shown in Fig. 4.2(b). It can be seen that, while circular closed molecules equilibrate at a value of ϑ in between 0 and 1, the linear molecule eventually fully denatures.

hand, linear molecules initially denature very little but eventually fully denature. Finally, the value of stably, partially denatured bp $\langle \vartheta \rangle$ is obtained by averaging the last $2 \times 10^7 \tau_{LJ}$ timesteps of the trajectories obtained for $\vartheta(t)$, for five different independent simulations.

The plot of $\langle \vartheta \rangle$ as a function of temperature or bond strength is called the melting curve for DNA. Typical profiles obtained from experiments [42] and BD simulations performed in this work, are shown in Fig. 4.2(a-b): the qualitative agreement is remarkable. The sharpness of the melting transition can be quantified in terms of the maximum value attained by the differential melting curve as $\Delta\kappa = |d\langle \vartheta \rangle / d\kappa|^{-1}$, where κ can either be temperature, T , or effective hydrogen bond strength, ϵ_{HB} , depending on the denaturation protocol. Quantitatively, Figures 4.2(a)-(b) show that experiments and simulations agree in predicting melting curves for tcDNA about three times broader than for tfDNA, i.e. $\Delta\kappa_{tc} / \Delta\kappa_{tf} \simeq 3$. Finally, it is important to notice that the critical Hydrogen bond energy ($\epsilon_{HB,c}$) at which half of the base-pairs are melt, changes

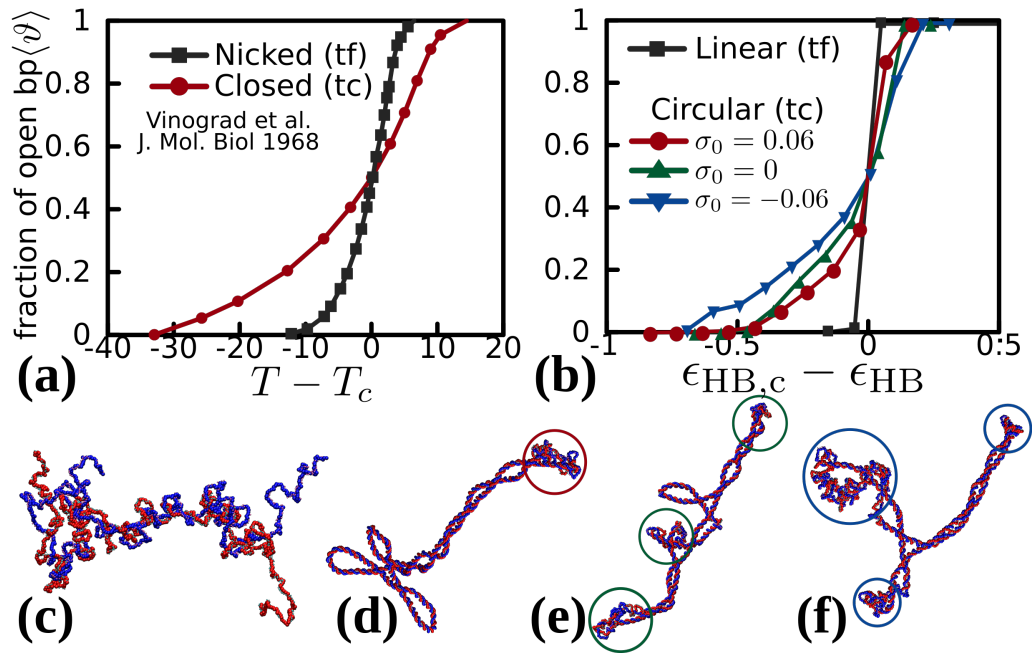


Figure 4.2 Melting curves. (a) shows the melting curves (fraction of denatured bp, $\langle \vartheta \rangle$) for nicked and intact circular DNA of polyoma as a function of $T - T_c$ (data from [42]). The nature of the transition has been investigated in the past. Simple theoretical models predict the transition to be continuous (second order) in two and three dimensions. However, when the effects of excluded volume between unbound loops and the rest of the chain are considered, the phase transition is found to be first order in dimensions two and above [137]. (b) shows the melting curves obtained here via BD simulations of *tf* and *tc*DNA molecules, with length 1000 bp and different levels of supercoiling, as a function of the (shifted) effective hydrogen-bond strength $\epsilon_{\text{HB},c} - \epsilon_{\text{HB}}$ (averaged over 5 replicas and 10^6 BD timesteps). In both experiments (a) and simulations (b), the transition appears smoother for *tc*DNA and the relative broadening $\Delta\kappa_{\text{tc}}/\Delta\kappa_{\text{tf}} \sim 3$ is in quantitative agreement. The critical bond energies for which half of the base-pairs melt are $\epsilon_{\text{HB},c}/k_B T = 1.35$ for linear DNA and 0.309, 0.238, 0.168 for $\sigma_0 = -0.06, 0$ and 0.06, respectively. These values show that the critical bond energy decreases (linearly) with supercoiling. (c)-(f) show snapshots of typical configurations for $\epsilon_{\text{HB}} = 0.3 k_B T$ for *tf* (linear) and *tc*DNA with $\sigma_0 = 0.06, 0, -0.06$, respectively. Stably denatured bubbles localise at regions of high curvature (tips of plectonemes [143], highlighted by circles). In (c) the linear DNA molecule is fully denatured.

linearly with the level of supercoiling ($\epsilon_{\text{HB},c} = 0.309, 0.238$ and $0.168 k_B T$ for $\sigma_0 = -0.06, 0$ and 0.06 , respectively). It should be observed that increasing the temperature is analogous to decrease the hydrogen bond energy, and that is why the melting curves in figures 4.2(a) and (b) are plotted as function of $T - T_c$ and $\epsilon_{\text{HB},c} - \epsilon_{\text{HB}}$ respectively.

4.2 Phase diagram

From the observations in the previous section, it is clear that DNA melting is affected by global topology. On the other hand, melting occurs through local opening of the double-helical structure. Therefore, the challenge faced by a theory of melting in tcDNA is to capture local effects due to the global topological invariance. It is useful to define an effective local supercoiling field $\sigma(x, t) \equiv (\text{Lk}(x, t) - \text{Lk}_0) / \text{Lk}_0$, where Lk_0 is the linking number between the two strands in the relaxed B-form state, i.e. 1 every 10.5 bp, and $\text{Lk}(x, t)$ is the effective linking number at position x and time t (see Appendix B.1 for more details). For a circular closed molecule of length L

$$\frac{1}{L} \int_0^L \sigma(x, t) dx = \sigma_0 \forall t, \quad (4.1)$$

where σ_0 is the initial supercoiling deficit, which can be introduced into the chain by, for instance, the action of topological enzymes [5]. To the contrary, circular nicked or linear (tf) dsDNA molecules need not satisfy Eq. (4.1), since any deviation from the torsionally relaxed state can be expelled through the chain ends or the nick. Subjecting a tcDNA to denaturation-promoting factors causes a competition between entropy and torsional stress: the former associated with the denatured coiled regions [136, 137], the latter arising in the intact helical segments [113]. In the following section we propose a phenomenological mean field theory in order to qualitatively understand these observations.

4.2.1 Landau mean field theory

Matter exists in different thermodynamic “phases”, which are different states of aggregation with qualitatively different properties. The most widespread example is the one of water, that depending on the temperature can be found in the familiar forms of ice, liquid or vapour. The drastic change in the thermodynamic behaviour of a physical system (for example from ice to vapour) is known as phase transition [144]. In the mean-field theory each of the phases in the system is characterized by a physical property that essentially depends on the temperature, called the order parameter ϕ [145] and each system has its own order parameter

(the density of a liquid-gas transition, the magnetization in a ferromagnet, among others). In general, ϕ is a quantity that changes from zero to a non-zero value when going from the disordered state (at high temperatures) to the ordered state (at low temperatures). The critical temperature at which the transition occurs is T_c . If the order parameter jumps discontinuously at T_c it is said that the transition is first-order, but if the order parameter changes continuously between the two phases the transition is second-order [144, 146].

Since at a phase transition the system undergoes a qualitative change and develops some order where there was none before, the free energy does not vary smoothly as a function of the order parameter. This means that the free energy is a non-analytic function of the order parameter, for which some derivatives are undefined at certain points (the singularities) [147]. The idea behind the Landau mean field theory is that the equilibrium thermodynamics is completely determined by a local free energy density ($f[T, \langle \phi(\mathbf{x}) \rangle]$, with $\langle \phi(\mathbf{x}) \rangle$ the local order parameter at position \mathbf{x}) that can be constructed in such a way that it exhibits the non-analyticity of the phase-transition [145]. In this approximation, the free energy is assumed to inherit all possible symmetries of the order parameter. In general f is a complicated function of ϕ , but since the order parameter is zero at the critical temperature, it is possible to expand f in a power series of ϕ in the vicinity of the critical point. Therefore, the local free energy density can be approximated to a polynomial function of the order parameter as

$$\beta f = a(T)\phi^2 + b(T)\phi^3 + \dots, \quad (4.2)$$

where $\beta = 1/(k_B T)$ and it is assumed that all the temperature dependence resides in the coefficients of the expansion.

In the case of the melting of *tf* and *tcDNA*, there are two important order parameters. A denaturation field, $\phi(x, t)$ (which is a non-conserved parameter), describing the state of base-pair x at time t (e.g., taking the value 0 if intact or > 0 if denatured), coupled to the conserved field, $\sigma(x, t)$ (it satisfies Eq. (4.1)), tracking the local supercoiling. A Landau free energy f can be constructed with these parameters, by noticing that: (i) the denaturation field ϕ should undergo a first-order phase transition when decoupled from σ [136]; (ii) the energy generated by small torsional strains should be symmetric with respect to σ if the DNA

is intact ($\phi = 0$). In other words, over and under twisting with respect to the equilibrium helical state should cost the same energy if there is no helix opening [118]. We stress that this does not mean that in the presence of local denaturation positive and negative supercoiled DNA are equivalent: clearly they are not, and this is taken care of by a term in the free energy that couples the two parameters (linear in σ and proportional to ϕ^2); (iii) the coupling between σ and ϕ should be such that the system displays an intact dsDNA phase at sufficiently low T , i.e. $\phi = 0$ for any σ_0 at $T < T_c$.

Based on these considerations we can write an effective free energy density as:

$$\beta f(\phi, \sigma) = \left(\frac{b^2}{4c} + 1 - a \right) \phi^2 + b\phi^3 + c\phi^4 + a_\sigma \sigma^2 + b_\sigma \sigma^4 + \chi \sigma \phi^2. \quad (4.3)$$

In this equation a is the only temperature-dependent parameter (it satisfies $a(T) \sim T/T_c$) and χ is the coupling between ϕ and σ . To get some insight into the behaviour of the free energy in the previous equation, in Appendix B.2 the isolated terms associated to the supercoiling and denaturation fields are studied (although is not necessary for the following calculations). The free energy density in Eq. (4.3) describes a first-order transition between a closed (double-stranded, ds, $\phi = \phi_{\text{ds}} = 0$) and an open (single stranded, ss, $\phi = \phi_{\text{ss}} > 1$) DNA molecule. We choose the parameter $c = -b/2$ so that the critical temperature depends only on $\chi\sigma$ (see below) and so that the free energy density displays a minimum located at around $\sigma = -1$ when ϕ attains its free-energy minimising value of ϕ_0 (see below). The corresponding choices are $c = -b/2$, $a_\sigma = 1$, $b_\sigma = 1/2$ and $b = -4$ (see Appendix B.2). The free energy density then reduces to

$$\beta f(\phi, \sigma) = (3 - a(T)) \phi^2 - 4\phi^3 + 2\phi^4 + \sigma^2 + \frac{\sigma^4}{2} + \chi \sigma \phi^2. \quad (4.4)$$

As a function of ϕ , Eq. (4.4) can display two local minima that represent the two phases (closed ϕ_{ds} and open ϕ_{ss}). At the critical (melting) temperature $a_c = a(T_c)$, the two phases coexist and therefore the free energy evaluated at the two minima is equal. We can find a_c by solving the equation $f(\phi_{\text{ds}}, \sigma) = f(\phi_{\text{ss}}, \sigma)$. For $\chi = 0$ molecules $a_c = 1$ independent of the initial level of supercoiling. In this situation (see left panel of Fig. 4.3) the minima are located at $\phi_{\text{ds}} = 0$ and at $\phi_{\text{ss}} = 1$ and mark the equal probability coexistence point. In the right

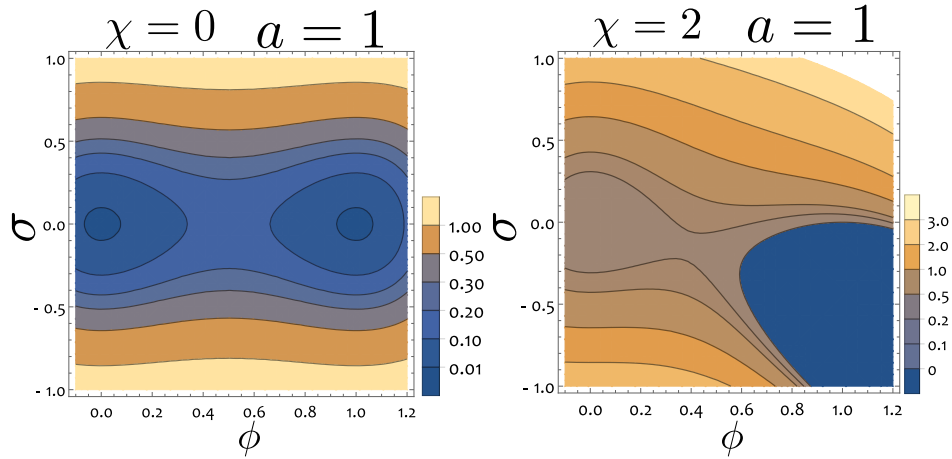


Figure 4.3 *Contour plot of free energy density.* Left shows $f(\phi, \sigma)$ for the *tf* case ($\chi = 0$) at the critical temperature $a(T_c) = a_c = 1$. The two minima (whose energy is equal) as function of ϕ are located at $\phi_{ds} = 0$ and $\phi_{ss} = 1$. Right shows the same information for the *tc* case ($\chi = 2$) and $a = 1$. The two minima have not the same energy. There is a global minimum at $\phi > 1$.

of Fig. 4.3 we show that by including the coupling parameter, the value of the critical temperature is shifted (this time $a = 1$ does not result in the two minima with the same free energy).

In general, the critical temperature is supercoiling-dependent through the coupling parameter χ (see Fig. 4.4(a-c)). The function $a_c(\sigma, \chi)$ is simply found by solving $f(\phi_{ds}, \sigma) = f(\phi_{ss}, \sigma)$, i.e.

$$a_c(\sigma, \chi) = \frac{b^2 + 4c(1 + b + c)}{4c} + \chi\sigma, \quad (4.5)$$

or

$$a_c(\sigma, \chi) = 1 + \chi\sigma, \quad (4.6)$$

by substituting the values of parameters. We report the values of a_c in Fig. 4.4(d) for $\chi = 0$ and $\chi = 2$. It can be readily appreciated that for topologically constrained ($\chi > 0$) molecules, the critical temperature a_c depends linearly on the supercoiling. The form of the coupling term ($\chi\sigma\phi^2$) in the free energy was intentionally introduced to produce this linear dependence which is in agreement with experimental observations [148] and the simulations reported in the previous sections ($\epsilon_{HB,c}$ changes linearly with σ).

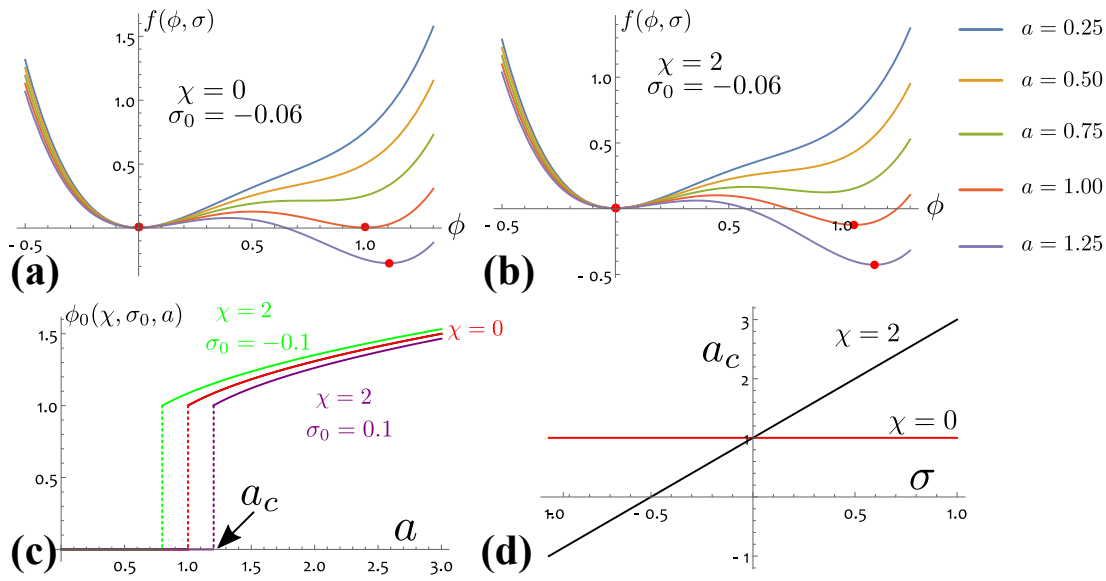


Figure 4.4 *First order transition for denaturation field ϕ .* (a)-(b) These plots show the free energy density as a function of ϕ . Red dots mark the location of the global minimum, i.e. $\phi = \phi_0$. (c) This plot shows the free-energy minimising value $\phi_0 = \phi_0(\chi, \sigma_0, a)$ that discontinuously jumps at the critical temperature a_c . (d) This plot shows that the critical temperature $a_c = 1$ for $\chi = 0$ and $a_c \sim \sigma$ for $\chi > 0$.

The equilibrium value of the denaturation field ϕ_0 is a discontinuous function of the temperature. It identifies the global minimum of the free energy with respect to the non-conserved order parameter. Therefore, if $\{\phi\}$ represents the set of local minima that satisfies $\frac{\partial f(\phi, \sigma)}{\partial \phi} = 0$, and the free energy is evaluated at these values ($f(\{\phi\}, \sigma)$), ϕ_0 is the one that minimises the free energy.

This condition leads to (see also curves in Fig. 4.4(c))

$$\phi_0(\chi, \sigma, a) = \begin{cases} 0 & \text{if } a < a_c(\sigma, \chi), \\ \frac{(3 + \sqrt{-3 + 4a - 4\chi\sigma})}{4} & \text{if } a \geq a_c(\sigma, \chi). \end{cases} \quad (4.7)$$

Because ϕ is a non-conserved order parameter, it will always attain the value $\phi_0(\chi, \sigma, a)$ for any given set of χ , σ and a . The free energy density of the system can therefore be written as $f(\phi_0(\chi, \sigma, a), \sigma) = f(\phi_0, \sigma)$. The behaviour of this function is plotted in Fig. 4.5.

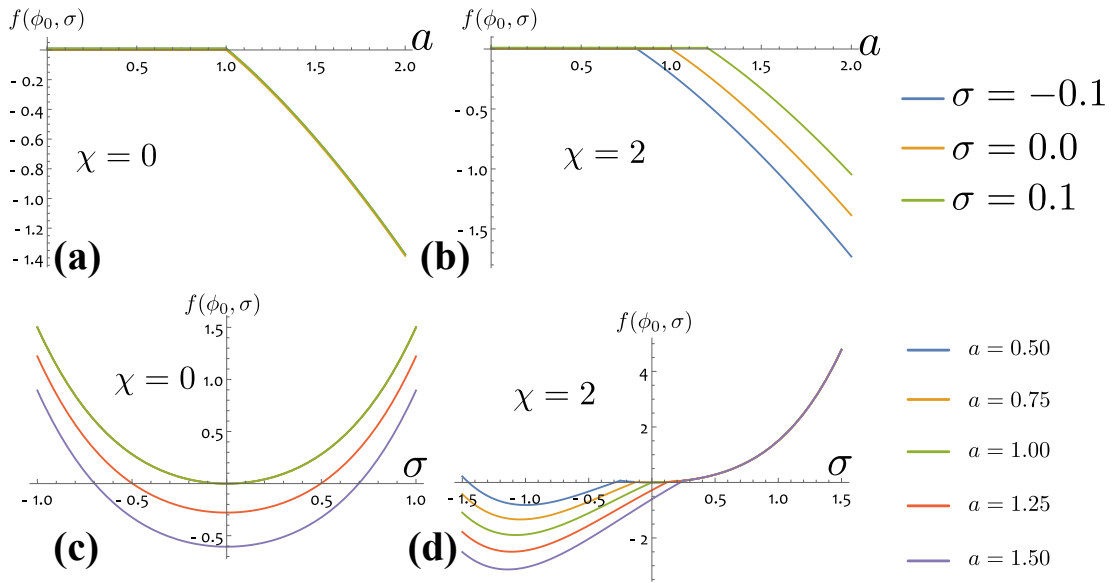


Figure 4.5 *Profiles of $f(\phi_0, \sigma)$.* (a) and (b) show the free energy density as a function of a for $\chi = 0$ and $\chi = 2$. (c)-(d) Same function plotted as a function of σ for topologically free ($\chi = 0$) and topologically constrained ($\chi = 2$) molecules.

One can readily check that for topologically constrained ($\chi > 0$) molecules, the critical temperature is supercoiling-dependent, $a_c = a_c(\sigma)$, and the profile as a function of σ is more complex. The term σ^4 in the free energy and the parameters chosen ensure that $f(\phi_0, \sigma)$ displays a minimum around $\sigma = -1$ for $\phi_0 = \phi_{ss}$ which corresponds to a fully denatured state with no supercoiling ($Lk = 0$) within the melted region. To the contrary, for $\chi = 0$ the discontinuity is always located at $a_c = 1$, irrespective of σ , and that the profile of the free energy is symmetric with respect to σ .

Figure 4.5(d) can be understood as follows: for fixed temperature (a) the critical supercoiling σ_c can be found through Eq. (4.6) as

$$\sigma_c(a) = (a - 1) / \chi. \quad (4.8)$$

At $\sigma_c(a)$ the system undergoes a first-order transition (corresponding to the discontinuity of the curves in Fig. 4.5(d)), and the system switches equilibrium state from ϕ_{ds} to ϕ_{ss} . This corresponds to a cusp in the behaviour of $f(\phi_0, \sigma)$. In order to find the coexisting phases we now need to perform the common tangent construction on the function $f(\phi_0, \sigma)$ represented in Fig. 4.5(d).

4.2.2 Spinodal region and binodal line

The supercoiling field σ is a globally conserved field, and the system may therefore undergo phase separation into regions of high and low σ in order to lower the overall free energy while preserving the initially set $\sigma = \sigma_0$. The region in the parameter space (a, σ_0) for which this occurs is known as coexistence region and is delimited by the binodal line.

The “spinodal” region, where small fluctuations favour spontaneous separation of the uniform phase $\sigma(x, t) = \sigma(t)$ into two phases, can be found by identifying the location of the inflexion points of $f(\phi_0, \sigma)$. This means that one needs to solve the equation,

$$\frac{\partial^2 f(\phi_0, \sigma)}{\partial \sigma^2} = 0. \quad (4.9)$$

This condition gives an analytical expression for the spinodal lines (for $\chi = 2$) as (see Fig. 4.6)

$$a_s(\sigma) = \frac{72\sigma^5 + 27\sigma^4 + 24\sigma^3 + 9\sigma^2 + 2\sigma + 3}{36\sigma^4 + 12\sigma^2 + 1} \quad (4.10)$$

valid if $-0.57735 < \sigma_0 < 0.57735$. The range of values of χ for which the requirement in Eq. (4.9) returns a solution $a_s(\sigma)$ is $1.72 \leq \chi \leq 2.67$. For this reason we chose a value in this range so $a_s(\sigma)$ is defined, i.e. $\chi = 2$.

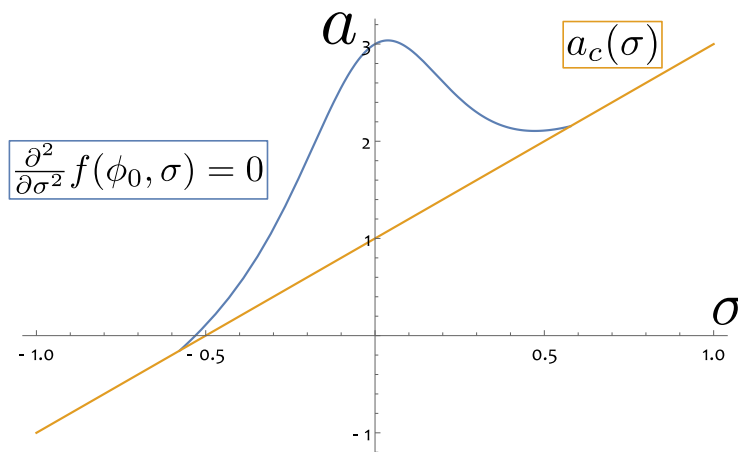


Figure 4.6 Plot of the spinodal line $a_s(\sigma)$ (blue) from Eq. (4.10) and of the critical temperature $a_c(\sigma)$ (yellow) from Eq. (4.6).

When the system undergoes phase separation in the conserved field σ , two coexisting solutions (or phases) appear, with distinct values of supercoiling: we label σ_+ the higher supercoiling solution and σ_- the lower supercoiling solution (in general σ_+ and σ_- need not have opposite sign). The values of σ_+ and σ_- are found by the common tangent construction, i.e. by equating the chemical potentials

$$\mu(s) = \left. \frac{\partial f(\phi_0, \sigma)}{\partial \sigma} \right|_s \quad (4.11)$$

and the pressures $\Pi(s)$

$$\Pi(s) = f(\phi_0, s) - \left. \frac{\partial f(\phi_0, \sigma)}{\partial \sigma} \right|_s \quad (4.12)$$

in the two phases, thereby leading to a system of two equations:

$$\mu(\sigma_-) = \mu(\sigma_+) \quad (4.13)$$

$$\Pi(\sigma_-) = \Pi(\sigma_+). \quad (4.14)$$

Equation (4.13) states that the chemical potentials of the two phases must balance in order for the system to be in equilibrium. Equation (4.14) identifies the only pair of points of $f(\phi_0, \sigma)$ that have equal tangent $\mu(\sigma_+) = \mu(\sigma_-)$ and that can be joined by a straight line.

Graphically, this procedure is summarised in Figure 4.7. The blue curve is the free energy $f(\phi_0, \sigma)$ resulting from a choice of $a = 1.1$. The green line corresponds to $\mu(\sigma_-)\sigma = \mu(\sigma_+)\sigma$ and the orange curve to the pressure $\Pi(\sigma)$. The common tangent construction identifies the points σ_- and σ_+ with same tangent μ and with equal pressure. The red dot placed along the (blue) curve $f(\phi_0, \sigma)$ represents the value of the free energy density $f(\phi_0, \sigma_0)$ for $\sigma_0 = -0.1$, which is higher than the free energy per unit volume of the phase-separated system:

$$f_+ f(\phi_0, \sigma_+) + f_- f(\phi_0, \sigma_-) \quad (4.15)$$

represented by the black dot. In Eq. (4.15), f_+ and f_- are the fractions of the system in the high and low supercoil phases, respectively.

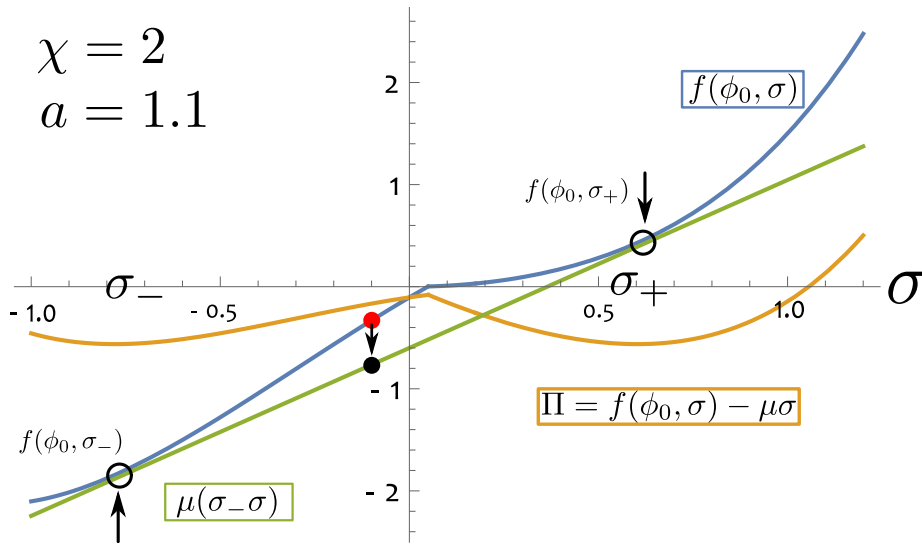


Figure 4.7 Graphical common tangent construction. The blue curve is the free energy $f(\phi_0, \sigma)$. The green curve is the term $\mu(\sigma_+)\sigma$ where $\mu(\sigma_+) = \mu(\sigma_-) = df/d\sigma|_{\sigma_-}$. The orange curve is the pressure $\Pi(\sigma) = f(\phi_0, \sigma) - \mu(\sigma_+)\sigma$. The red dot is the value of the free energy density in the uniform phase $f(\phi_0, \sigma_0 = -0.1)$, while the black dot gives the value of the free energy per unit volume in the phase separated state, i.e. $f = f_+f(\phi_0, \sigma_+) + f_-f(\phi_0, \sigma_-)$ (see text).

For a general value of the initial supercoiling field σ_0 , one can find f_+ and f_- through the so-called lever-rule [149]:

$$f_+ + f_- = 1 \quad (4.16)$$

$$f_- \sigma_- + f_+ \sigma_+ = \sigma_0 \quad (4.17)$$

which gives

$$f_- = \frac{\sigma_+ - \sigma_0}{\sigma_+ - \sigma_-} \quad (4.18)$$

and

$$f_+ = \frac{\sigma_0 - \sigma_-}{\sigma_+ - \sigma_-}. \quad (4.19)$$

Equations (4.13) and (4.14) can be solved numerically and give the binodal lines $\sigma_+(a)$ and $\sigma_-(a)$. These are plotted in Fig. 4.8 along with the critical line $\sigma_c(a)$ and the spinodal line from Eq. (4.10). Together they form the phase diagram in the (a, σ_0) space. As seen before, the critical supercoiling $\sigma_c(a)$ is a linear function of the temperature (see Eq. (4.6)). Therefore, at any initial σ_0 one

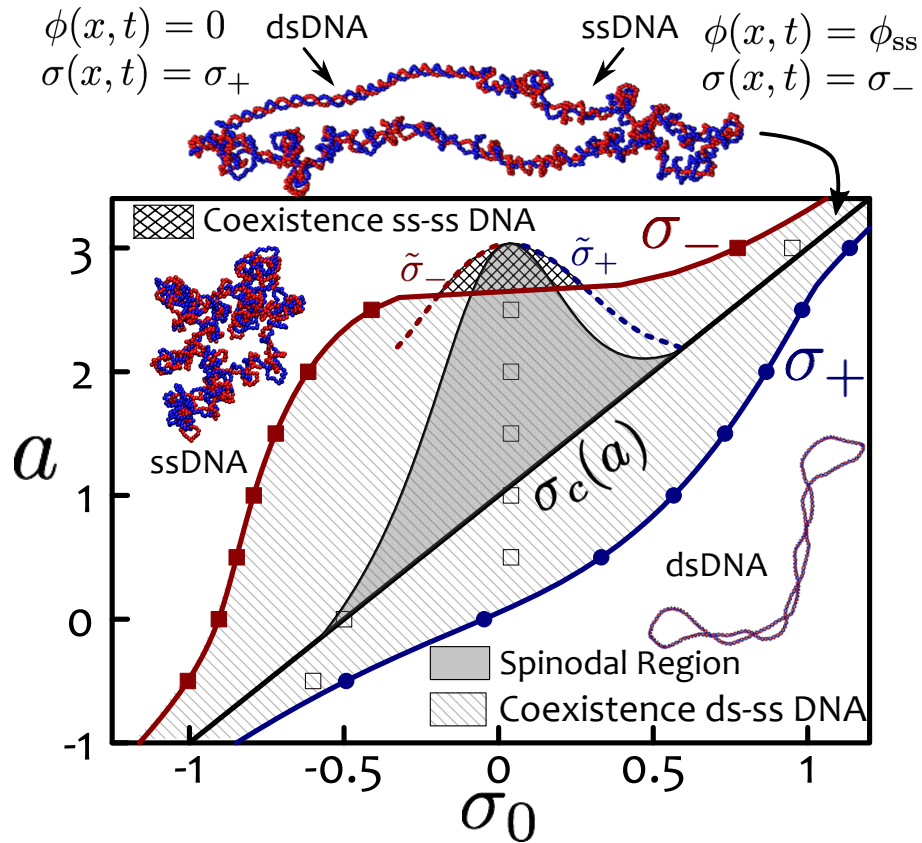


Figure 4.8 Phase Diagram. The thick solid black line represents the hidden first order transition line $\sigma_c(a)$. The shadow gray region represents the spinodal. Binodal lines are denoted as σ_- (red) and σ_+ (blue). Inside the binodal lines, the hatched area highlights the region of absolute instability of the uniform phase. Cross-hatched area highlights the region of coexistence of two denatured (ss) phases. Filled symbols denote the values obtained from numerical integration of Eq. (4.22), with initial σ_0 as indicated by the empty squares. Snapshots of ssDNA, dsDNA and ds-ss DNA coexistence observed in BD simulations are also shown.

can always find a temperature above the critical line for which the DNA molecule completely melts (the ssDNA region). The opposite is also true, it is always possible to find a temperature below a_c at which the DNA remains closed (the dsDNA region). However, the coupling χ in the free energy results into both, the spinodal described by Eq. (4.10) and the binodal lines $\sigma_+(a)$ and $\sigma_-(a)$. These curves mark the region of two coexisting phases with different values of supercoiling.

The binodal line $\sigma_-(a)$ crosses the spinodal leaving a reentrance where the uniform solution is linearly unstable, yet the binodals $\sigma_+(a)$ and $\sigma_-(a)$ are not

the stable phases. In this region two stable open (ssDNA) phases are observed, whose supercoiling take values on additional binodals $\tilde{\sigma}_-$ and $\tilde{\sigma}_+$, shown as red and blue dashed lines respectively (obtained also as solutions of equations (4.13) and (4.14)). The presence of this coexistence region between denatured DNA is a prediction of the Landau theory. It arises because the effective free energy as a function of σ , $f(\phi_0, \sigma)$, is not convex. Therefore, while this coexistence is physically plausible in principle, as it decreases linking in some of the denatured regions, we cannot exclude its appearance might be an artefact due to our specific choice of the stabilising terms in the free energy. To explore this issue further, we compute the local linking number directly in the BD simulations as described in Appendix B.1. We observe that $\text{Lk}(x, t)$ of configurations near the transition fluctuates around the relaxed Lk_0 . However, we are currently unable to unambiguously determine whether the uniform supercoiling state in the denatured region is stable or not.

Finally, in Figure 4.8 we show that the coexistence lines $\sigma_-(a)$ and $\sigma_+(a)$ wrap around the first-order transition line at the critical supercoiling $\sigma_c(a) = (a - 1)/\chi$ which therefore becomes “hidden” [150]. In light of this we argue that the smoother transition observed for tcDNA [42, 139] can be understood as a consequence of the emergence of a coexistence region in the phase space which blurs the underlying first-order transition. This argument also explains the “early melting” of closed circular DNA [42], resulting from the entry into the coexistence region from low temperatures.

4.3 Dynamical scaling

The dynamic properties of a system are quantities that depend on the equations of motion, and are not simply determined by the equilibrium distribution of the particles at a given instant of time [151]. When a system is close to a critical point, its dynamics can be understood from the evolution of the order parameters. The Glauber and Cahn-Hilliard prescriptions [151, 152] are used to describe the dynamics of the non-conserved and conserved order parameters, respectively. Here we are interested in solving the case when two order parameter (one conserved σ and the other non-conserved ϕ) are coupled. Consequently, the

system can be described by the following “model C” equations [145]

$$\frac{\partial\phi(x,t)}{\partial t} = -\Gamma_\phi \frac{\delta\mathcal{H}(\phi,\sigma)}{\delta\phi}, \quad \frac{\partial\sigma(x,t)}{\partial t} = \Gamma_\sigma \nabla^2 \frac{\delta\mathcal{H}(\phi,\sigma)}{\delta\sigma} \quad (4.20)$$

with $\Gamma_{\sigma,\phi}$ the relaxation constants of σ and ϕ respectively, and

$$\mathcal{H}(\phi,\sigma) = \int \left(f(\phi,\sigma) + \gamma_\phi (\nabla\phi)^2 + \gamma_\sigma (\nabla\sigma)^2 \right) dx, \quad (4.21)$$

where $\gamma_{\phi,\sigma}$ determine the effective surface tension of bubbles and supercoiling domains, respectively. From Eqs. (4.3) and (4.20) we get

$$\begin{aligned} \frac{\partial\phi}{\partial t} &= -\Gamma_\phi \left[2 \left(\frac{b^2}{2c} + 1 - a \right) \phi + 3b\phi^2 + 4c\phi^3 + 2\chi\phi\sigma - \gamma_\phi \nabla^2 \phi \right] \\ \frac{\partial\sigma}{\partial t} &= \Gamma_\sigma \nabla^2 \left[2a_\sigma\sigma + 4b_\sigma\sigma^3 + \chi\phi^2 - \gamma_\sigma \nabla^2 \sigma \right]. \end{aligned} \quad (4.22)$$

We numerically solve this set of partial differential equations (PDE) on a 1D lattice of size L for fixed a and σ_0 (see Appendix B.3) and compare the evolution of denaturation bubbles with that observed in BD simulations. These equations disregard thermal noise, hence constitute a mean field theory.

In Figure 4.9(a-b) we show “kymographs” from BD simulations, capturing the state of each base-pair (intact or denatured) against time for tf and tcDNA. After the energy quench at $t = 0$, the linear (tfDNA) molecule starts to denature from the ends and eventually fully melts. On the other hand, in the closed circular (tcDNA) molecule, bubbles appear randomly over the whole length, and the steady state entails a stable fraction, $0 < \vartheta < 1$, of denatured bp (Fig. 4.1).

We observe similar behaviour when ϕ and σ are evolved via Eqs. (4.22), starting from a single small bubble at temperature a within the coexistence region (see Fig. 4.9(c)). While the bubble grows, the supercoiling field is forced outside the denatured regions and accumulates in the double-stranded segments. The increasing positive supercoiling in the helical domains slows down and finally arrests denaturation, resulting in phase coexistence in steady state, between a denatured phase with $\sigma = \sigma_-$ and $\phi = \phi_{\text{ss}} > 1$, and an intact phase with $\sigma = \sigma_+$ and $\phi = \phi_{\text{ds}} = 0$.

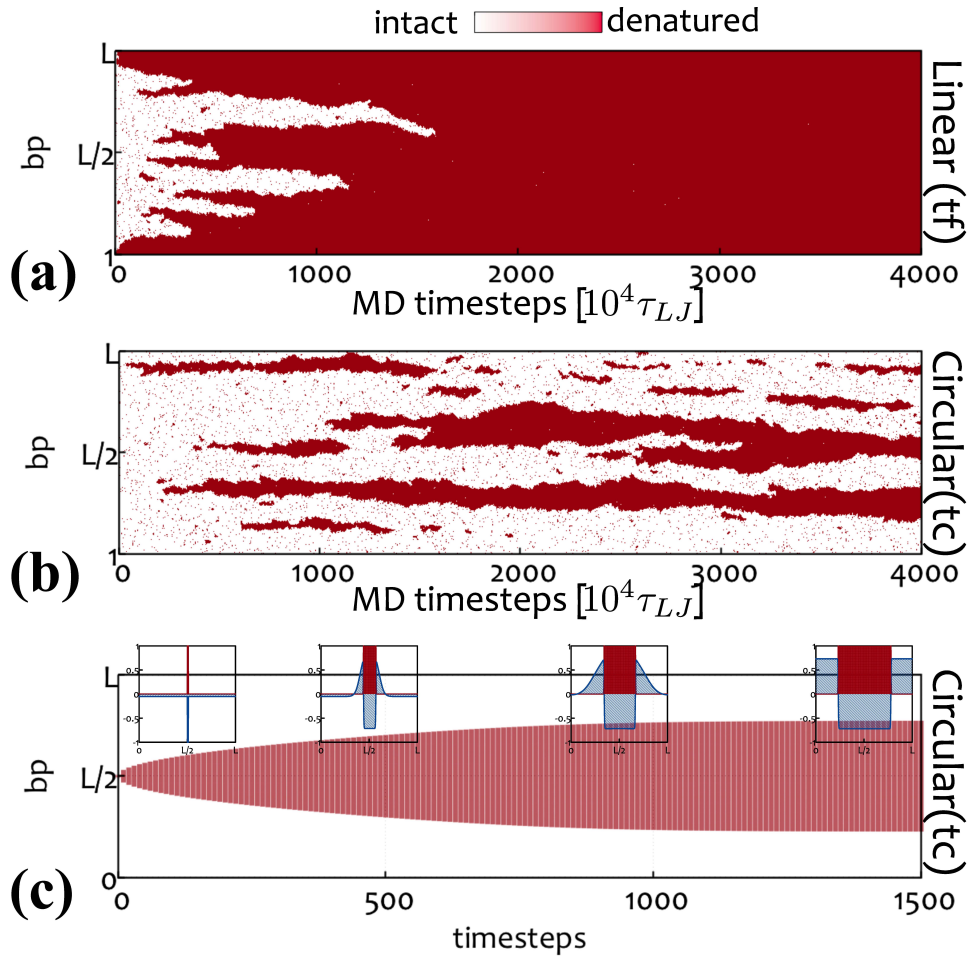


Figure 4.9 Kymographs. (a)-(b) report results from BD simulations. At time $t = 0$, the H bond strength is quenched to $\epsilon_{HB} = 0.3k_B T$ and the local state of the chain (red for denatured and white for intact dsDNA) is recorded as a “kymograph”. (a) and (b) show the case of a tf and $tcDNA$ ($\sigma_0 = 0$), respectively. (c) shows the kymograph of the system during integration of Eqs. (4.22) starting from a small bubble (see Appendix B.3). Insets show instantaneous profiles of denaturation field (red) and supercoiling field (blue).

The growth, or coarsening, of a denaturation bubble, l , can be quantified within our mean field theory and BD simulations. As shown in Fig. 4.10(a-c) we find that in both cases,

$$l(t) \sim \begin{cases} t^1 & \text{for topologically free DNA,} \\ t^{1/2} & \text{for topologically constrained DNA.} \end{cases} \quad (4.23)$$

In other words, the exponent α governing the *local* growth of a denaturation bubble depends on the *global* topology of the molecule.

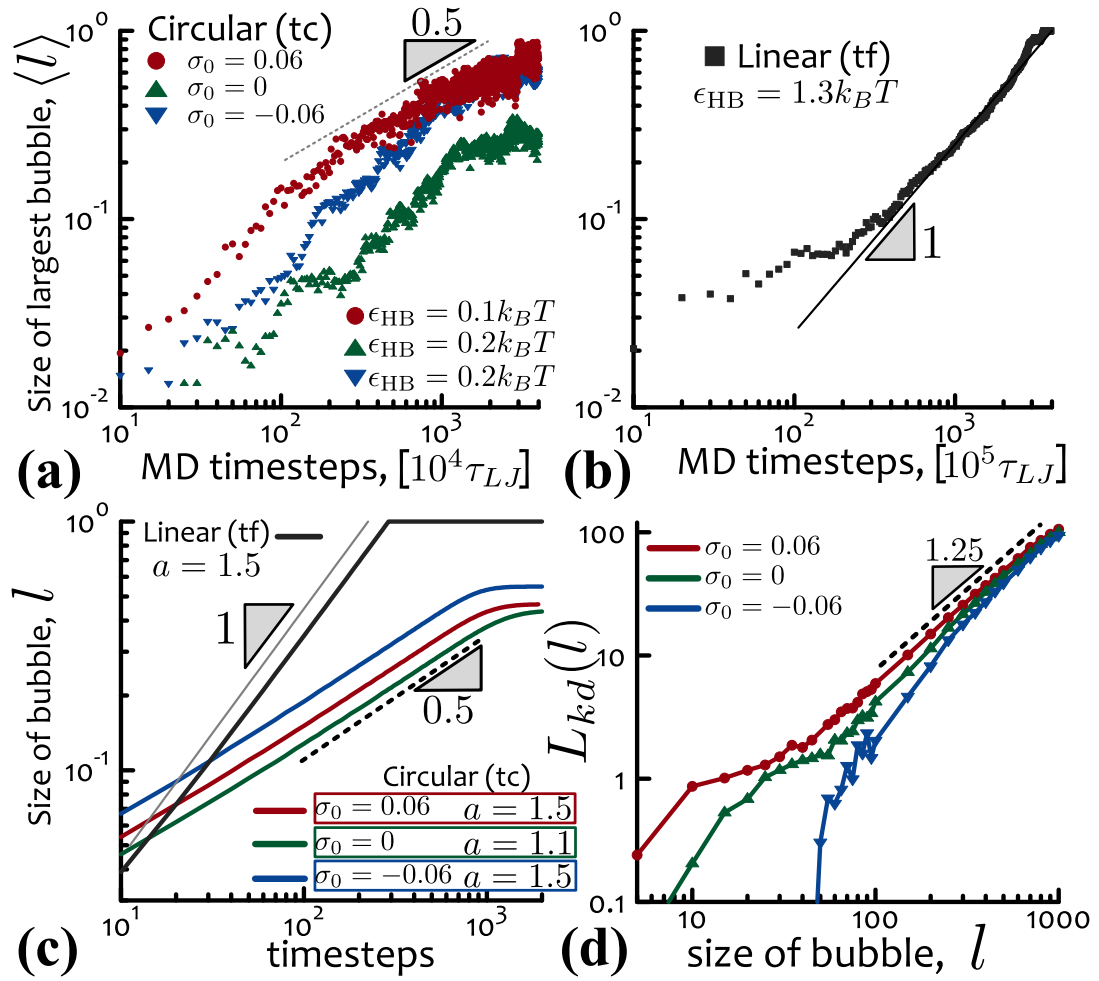


Figure 4.10 Dynamical Scaling. (a)-(b) show results from BD simulations. The size of the largest denatured bubble $\langle l \rangle$ (averaged over 5 replicas) is plotted against time from the moment of the quench. (a) shows tcDNA while (b) refers to tfDNA. (c) shows the size of a single growing bubble, l , within our mean field model, Eqs. (4.22). PDE and BD simulations show similar behaviours, which suggest a universal dynamical scaling with topology-dependent exponent ($\alpha = 1$ for $\chi = 0$ and $\alpha = 1/2$ for $\chi > 0$). (d) shows the linking number, L_{kd} , stored inside a denatured bubble of fixed size l computed from BD simulations (see text and Appendix B.4 for details).

We propose the following argument to explain the values of α . For tfDNA (e.g., nicked or linear), we can assume that the supercoiling field relaxes quickly, and gets expelled, without affecting the dynamics of the denaturation field. In this case, the system can lower its free energy $f \simeq (\epsilon_{HB} - T\Delta S)l$ for each unpaired (denatured) base-pair at temperatures above a critical $T_c = \epsilon_{HB}/\Delta S$. In other words, there is a constant increase in entropy per each denatured bp when $T > T_c$. This implies that [145] $df/dl \simeq \psi dl/dt \sim const$, with ψ an effective

constant friction; as a result we obtain $l(t) \sim t$.

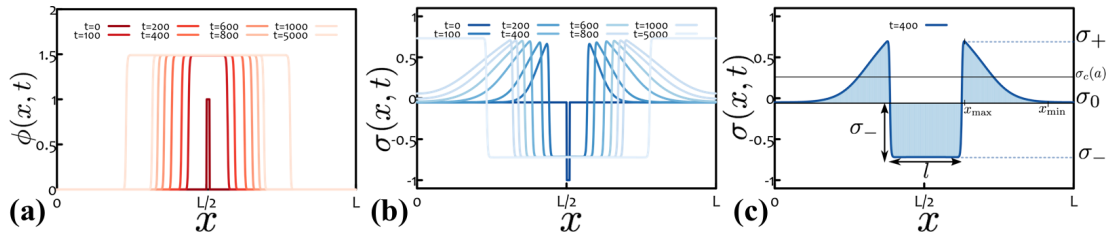


Figure 4.11 (a-b) Evolution of the fields $\phi(x, t)$ and $\sigma(x, t)$ as a function of position x starting from a small denatured bubble in the middle of the chain (see Appendix B.3 for full details on the calculations). Panel (c) highlights the area within the denatured bubble (square region with height $= |\sigma_- - \sigma_0|$ and base $l(t)$) which must equal the area outside it; in other words, the supercoiling front forming outside the growing bubble has an area that grows linearly with $l(t)$. Because the maximum of the front remains approximately constant at σ_+ , its base must then grow as $l(t)$. For this reason, one can approximate the slope of the “front” as $\partial_x \sigma(x, t)|_{\text{front}} \simeq (\sigma(x_{\text{max}}, t) - \sigma(x_{\text{min}}, t)) / (x_{\text{max}}(t) - x_{\text{min}}(t)) \simeq (\sigma_+ - \sigma_0) / l(t)$. Parameters: $a = 1.5$, $\chi = 2$ and $\sigma_0 = -0.06$.

The value of $\alpha = 1/2$ observed for tcDNA (e.g., circular non-nicked plasmids) can be understood by quantifying the slowing down of denaturation due to the accumulation of a “front” of supercoiling, raked up on either side of the growing bubble. By solving numerically the model C equations we can evolve the system and record the dynamics of the fields over time (see Appendix B.3). The typical profiles are reported in Figure 4.11(a-b), where the denaturation field ϕ and supercoiling field σ are shown for $a = 1.5$, $\chi = 2$ and $\sigma_0 = -0.06$. Let us call σ_- and σ_+ the supercoiling within the denatured bubble and in the dsDNA region just outside the bubble respectively. As the linking number is globally conserved, the amount of supercoiling which is expelled from the bubble after time t , i.e. $\sim |\sigma_- - \sigma_0| l(t)$ (see Fig. 4.11(c)), must be equal to supercoiling which builds up in the close proximity of the bubble. The corresponding “front” of positive supercoiling can be approximated by the area of a triangle with constant height $h = \sigma_+ - \sigma_0$ and growing base $b \sim l(t)$ (see Fig. 4.11(c) and Fig. 4.9(c)).

The supercoiling lost from the bubble is proportional to the increase in denatured fraction (or flux of ϕ) which can be estimated as $\sim |\sigma_- - \sigma_0| dl/dt$ (here, we are assuming that in the denatured region there is no supercoiling). On the other hand, the local supercoiling flux into the intact double-stranded

neighbour region ($J_\sigma = -\Gamma_\sigma \partial_x \sigma$), is proportional to the front of the triangle profile with slope $h/l = (\sigma_+ - \sigma_0)/l$. By equating the two fluxes of supercoiling ($dl/dt \sim 1/l$) we get the scaling law $l \sim t^{1/2}$.

4.4 Linking within denaturation bubbles

As a final result, we perform BD simulations to characterise the topology of a denaturation bubble by measuring the linking number that can be stored inside it (Fig. 4.10(d)). Naively one could expect $Lk = 0$ (and $\sigma = -1$) within a bubble [113], however our BD simulations show that a denatured region of *fixed* size l (imposed by selectively breaking *only* l consecutive bonds along an intact dsDNA molecule) has a non-zero linking number Lk_d (see Fig. 4.10(d)). In practice, Lk_d was obtained by removing the linking number of the chain *outside* the denatured region Lk_{out} from the initially set Lk , see Appendix B.4 for details.

We found that for small l , Lk_d displays a remarkable signature of global topology (through the value of σ_0 , see Fig. 4.10(d)); instead, the scaling behaviour at large l appears to follow $Lk_d \sim l^{1.25}$ irrespectively of σ_0 , until it reaches Lk_0 . The finding that a denaturation bubble displays a non-trivial and l -dependent linking number suggests that idealised ($Lk_d = 0$) bubbles may not always be reflecting realistic behaviour. Further, it may be of relevance for processes such as DNA replication, as it suggests that supercoiling or torsional stress may be able to diffuse past branching points such as replication forks [153].

4.5 Summary

In summary, we have studied the melting behaviour of topologically constrained DNA through a combination of large-scale BD simulations and mean field theory. A key result is that the phase diagram for tcDNA melting generally involves a phase coexistence region between a denatured and an intact phases, pre-empting a first-order denaturation transition as in tfDNA. This finding provides a theoretical framework to explain the long-standing experimental observation that the denaturation transition in circular, and not nicked, supercoiled plasmids is

seemingly less cooperative (smoother) than for linear, or nicked, DNA [42, 139].

We have further studied, for the first time, the coarsening dynamics of denaturation bubbles in tcDNA, and found a remarkable agreement between BD simulations and mean field theory, both reproducing similar topology-dependent scaling exponents that can be understood within our theoretical model. In the future, it would be of interest to investigate such dynamics experimentally as well as to include sequence heterogeneity in our theory.

5

Transcription-Driven 3D Supercoiling Dynamics

Abstract: I study the action of an RNA polymerase on the dsDNA model proposed in Chapter 3. The simulations capture the interplay between twist and writhe on this system and confirm the generation of the twin supercoiling domain. I find that the creation of plectonemes is associated with a competition between the relaxation of twist and writhe, where the latter is orders of magnitude slower than the former. The scenario in which the polymerase interacts with a nucleosome via the 3D conformations dictated by supercoiling is also investigated.

Irrespective of the organism, *in vivo* DNA is constantly being remodelled by proteins that drive the system away from equilibrium; these protein complexes, such as RNA polymerase II (polII), can apply forces and torques the order of 25 pN [154, 155] and 11 pN nm [156] respectively. The action of these enzymes results in the generation of supercoils: the classic example is the “twin supercoiling domain” [157, 158], where the assumption is that the rotation of an elongating polII protein is inhibited, leading to twisting of the DNA; the result is formation of positive supercoils ahead of the polymerase, with negative supercoils in its wake. It has been proposed that such dynamically generated supercoiling can play a regulatory role in gene expression either through the twist dependence of polII-DNA interactions [107, 159], or via supercoiling mediated generation of DNA loops [106]. While the twin supercoiling domain idea has been around for

many years, its consequences *in vivo* are still not well understood.

In this chapter I use Brownian dynamics simulations to study polII dynamics and the generation of a twin supercoiling domain on a DNA ring. Importantly, the simulations resolve the full three-dimensional (3-D) dynamics of the DNA, going beyond the 1-D description used in recent modeling efforts [107, 159, 160]. Contrary to typical text-book diagrams which show twist or writhe being generated close to the polymerase on both sides, the dynamics in the simulations is more complicated, and writhe (in the form of a plectoneme) can be produced far from the polymerase. This allows for an “action at a distance” phenomena, where transcription at one point on the DNA can affect its dynamics (and, for example, the binding of proteins) elsewhere.

First, I use simulations to characterise how the properties of the polymerase (i.e. the force it generates, or equivalently its speed) affect where and how the plectonemes form (considering a system where the DNA is subject to various topological constraints). Next, I demonstrate that supercoiling can affect the binding of proteins far from the polymerase, by simulating the unwrapping of a simple model nucleosome due to polymerase action. In order to better understand these observations, I then measure the diffusion of twist and relaxation of writhe. Finally I show that, depending on the elongation speed or the size of the DNA molecule, the twin supercoiling domain can still lead to the formation of plectonemes in circular DNA molecules even without any topological constraints.

5.1 3D Modelling of transcription

We simulate a circular DNA molecule, so the linking number is fixed and the White-Fuller-Calugernau (Eq. (2.29)) relationship holds. In all cases, this molecule overall is not supercoiled, i.e. its linking number is equal to the “relaxed” value $Lk = Tw_0 = N/p$ (see section 2.3).

To investigate the effect of polII elongation on the DNA topology, a simple model which mimics the mechanical action that a polymerase applies to the DNA is used. In short, the model polymerase is a rigid body consisting of a ring which encircles the double helix, and a “crossbar” which passes through the helix between the two strands and attaches to the ring on either side (see Fig. 5.1a(i)).

The section of DNA which passes through the polymerase experiences an active force directed perpendicular to the crossbar – this acts to drive relative motion of the polymerase and DNA. Once in relative motion the crossbar forces the opening up of the double strand and leads to positive twisting (double strand overwinding) in front of the rod and negative twist behind (double strand unwinding). Thus the mechanical actions of an elongating polymerase (which must unwind a section of the helix in order to “read” the nucleotides [4, 5]) is naturally reproduced. Full details of the polymerase model are given in Appendix C.1.

For simplicity here we focus on a geometry where the polymerase is fixed in place; this could represent an *in vitro* experimental setup where a polII molecule is attached to a bead held by optical tweezers, but is also likely to be relevant to the *in vivo* case, since it is possible that the large elongation complex experience a larger rotational drag than the DNA. We confirm that our results are also applicable for the case of a moving polymerase in Appendix C.2.

5.2 Twin supercoiling domains drive plectoneme formation at a distance

As discussed above, the mechanical consequence of transcription is that positive DNA twist is generated locally in front of the polymerase, with an equal negative twist left behind it. Since we typically simulate relatively short circular DNA molecules of 1000 bp, we expected that the generated twist would easily travel along the DNA around the circle, whereupon the positive and negative twist will annihilate (previous work has hypothesised that the twist will move diffusively [107]). In the crowded environment of a cell, and for a much larger DNA molecule one would expect that the twist would not be able to quickly diffuse away; to realize this in the simulations, a topological barrier is introduced – a section of DNA which could not rotate, and as such prevented the passage of supercoiling. We investigated the effect of twist generation in this context by examining different values of the polymerase force for a DNA molecule which was initially topologically relaxed ($Lk = Lk_0$).

When a sufficiently large force is applied by the polymerase, we observe the emergence of plectonemic structures on each side. Closer inspection reveals that

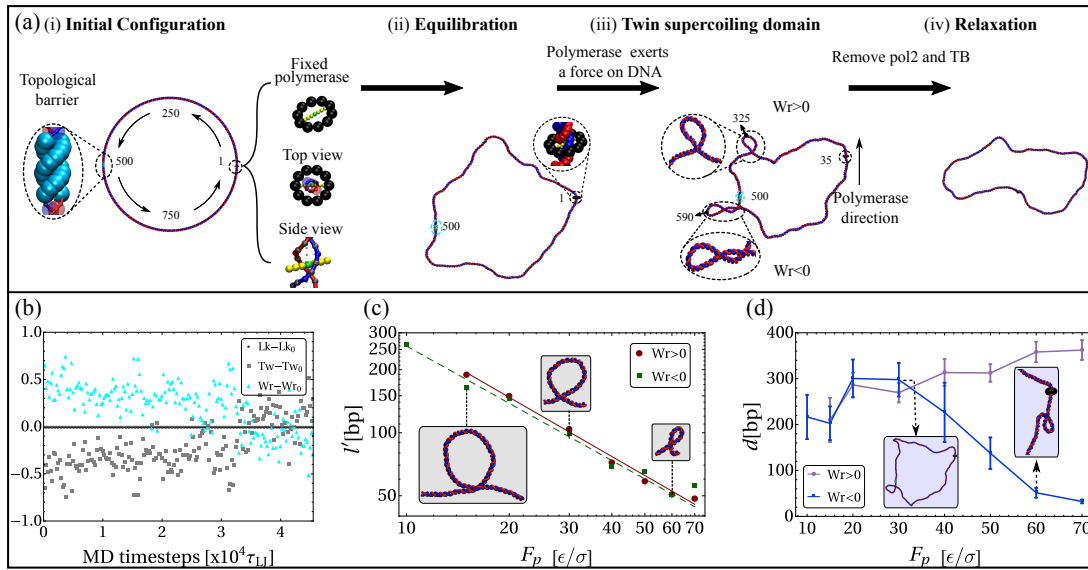


Figure 5.1 (a) Generation of a twin supercoiling domain is shown schematically. (i) A 1000 bp DNA molecule is initialized as a perfect ring with $Lk = Lk_0$. Base-pairs (bp) are labelled 1...1000 as indicated. A polymerase is attached at bp 1, and consists of a rod or “crossbar”, responsible for breaking the hydrogen bonds, and a ring (see Appendix C.1 for details). The position of the polymerase is fixed throughout the simulation (we present simulations for a moving polymerase in Appendix C.2). A topological barrier (TB) is positioned opposite the polymerase (bp 495-505), and these bp cannot rotate during the simulation. (ii) The simulation is run for $1 \times 10^5 \tau_{LJ}$ to allow the ring to relax to an equilibrium configuration. The polymerase is then “activated”, i.e. an active force \mathbf{F}_p is applied to all DNA beads which are within a distance 0.5 nm of the centre of the polymerase ring, in a direction perpendicular to the plain of the ring. This generates a twin supercoiling domain. (iii) After the polymerase has moved 34 bp, plectonemes have been generated with their tips at bases 325 and 590. Specifically a positively writhed plectoneme with one crossing has formed in front of the polymerase, and a negative plectoneme with two crossings has formed behind. (iv) We continue the simulation after removing the TB and polIII to study the relaxation of the writhe. (b) Temporal evolution of Tw , Wr and Lk with respect to their corresponding value in the relaxed state: Tw_0 , Wr_0 and Lk_0 . The time corresponds to the evolution between states ii and iii. (c) log-log plot of the length l' , of the plectoneme when it first forms when a force of magnitude F_p (in simulation units, ϵ/σ) is applied by the polymerase. Red points are for plectonemes with positive writhe, and green points for negative writhe. Lines are fits to the data: red (solid) shows $l' \sim F_p^{-0.94}$, and green (dashed) $l' \sim F_p^{-0.9}$. Snapshots are shown as insets for $F_p = 15, 30$ and $60 \epsilon/\sigma$. (d) Distance d along the molecule, from the polymerase to the first plectoneme crossing to form, as a function of F_p . Blue points are for $Wr < 0$ and violet points for $Wr > 0$; connecting lines are a guide for the eye. Snapshots are shown as insets for $F_p = 30 \epsilon/\sigma$ and $F_p = 60 \epsilon/\sigma$.

these plectonemes can appear at large distances from the polIII complex, and that each can store different amounts of writhe (Fig. 5.1a(iii)) – after the polymerase

has travelled the equivalent of 3.4 turns of the helix, a plectoneme with one right-handed crossing, or $Wr \simeq +1$, forms ahead of the polIII, and a plectoneme with two left-handed crossings, or $Wr \simeq -2$ forms behind). This is contrary to the usual twin supercoiled domain picture [157], which would suggest equal but opposite writhing on each side close to the polymerase.

Even though the writhe is unbalanced at the level of the plectonemes, the total linking number must still be conserved to satisfy the WFC theorem, i.e. $Lk(t) = Lk_0 = Tw(t) + Wr(t)$ [161]. In order to verify that this is the case we compute the global twist and writhe as a function of time (see Fig. 5.1(b) and Appendix C.3 for details on the computation). We find that these fluctuate around their relaxed values Tw_0 and Wr_0 , and that the total linking number $Lk = Lk_0$ as expected. These results indicate that not all of the writhe is necessarily stored in the plectonemes but can be delocalised over the whole polymer conformation (counting positive and negative crossings would give $Wr = -1$, whereas overall we measure $Wr \approx Wr_0 = 0$).

In order to understand the formation of the plectonemes more quantitatively, we developed a strategy to monitor both their position and length as they are generated. Briefly, this is done by examining at each time step the “contacts” between non-adjacent bp (which tend only to come into close proximity at the crossing points within plectonemes – the length of a plectoneme is defined as the bp distance between the two points of the outermost crossing). For details see Appendix C.4. First, we find that the length, l' , of the plectoneme which forms first depends on the magnitude of the force applied by the polymerase, F_p : the larger the force the smaller the plectoneme (Fig. 5.1(c)) and specifically $l' \sim F_p^{-\alpha}$ with $\alpha \simeq 0.9 - 1$. This scaling can be understood with a simple argument: injecting twist faster than it can diffuse away triggers the formation of a plectonemic structure. The rate of injection depends on the velocity of the polymerase with respect to the DNA, v_p , and a characteristic time scale for this is l/v_p , where l is a typical length scale; if the twist diffuses with an effective diffusion constant D_0 , then the time to diffuse across this length scale is l^2/D_0 . We expect twist to convert into writhe and plectonemes to form when these two times are equal, $l/v_p \approx l^2/D_0$, leading to a scaling $l \sim v_p^{-1}$. Assuming a linear relationship between force and velocity ($F_p \sim \gamma v_p$ with γ an effective frictional drag) then we recover $l \sim F_p^{-1}$. Here is worth mentioning that this linear relation

holds for $F_p < 30 \epsilon/\sigma$ (see Appendix C.5) and we still need to think about what is happening at larger forces.

Second, there are two regimes for how the relative distance from the polymerase to the base of the first plectoneme, d , depends on F_p (Fig. 5.1(d)). For large forces ($F_p \geq 30 \epsilon/\sigma$), this depends on the sign of the writhe, or equivalently on the relative position of the plectoneme with respect to the polII complex (in front or behind). For $Wr > 0$ (in front of the polII) the plectoneme appears at a distance d that *increases* with the force, while for $Wr < 0$ (behind the polII), d instead *decreases* with F_p . We attribute this observation to the tension experienced by the DNA in front of the polII as it is “reeled in” – this is large enough to lead to the DNA close to the polII being pulled taut, which suppresses writhing (Fig. 5.1(d), rightmost inset). At the same time there is an accumulation of relaxed DNA behind the polII which can readily form a plectoneme. For small forces ($F_p < 30 \epsilon/\sigma$), we observe a markedly different behaviour – in this regime the DNA has time to relax as it is pulled through the polII and does not become taut. We would expect plectoneme formation to be suppressed in the immediate vicinity of the topological barrier and polymerase due to the reduced freedom to move and rotate; since there is no other position dependence for plectoneme formation we expect this on average will occur close to the half way point between these topological constraints on both sides, at $L/4$.

A final intriguing observation is that in most of the simulations ($\simeq 80\%$) we observe that the first plectoneme to appear is the one behind the polymerase carrying negative writhe, i.e. where the DNA is underwound. This is in agreement with both the results in reference [82] (see also section 2.3.2) and the observation in our simulations (see Appendix C.6) that negative supercoiling is more difficult to twist (easier to bend) compared with the positive one. It further implies that our model should correctly capture the asymmetric torsional rigidity seen for real dsDNA [74] and discussed in section 2.3.1. To corroborate this, we set a simulation with a model polymerase moving along DNA while the hydrogen bond energy was decreased to $\epsilon_{\text{HB}} = 0.1k_B T$ (see Fig. 5.2). Under these circumstances, the plectoneme in front of the polymerase (with positive writhe) was the first to appear. In addition, small melted “bubbles” became visible in the region behind the polymerase. After enough stress was accumulated, the plectoneme carrying negative writhe emerged. Since in these simulations the dihedral and

the Kratky-Porod interactions between nucleotides which are part of a bubble were not erased, the flexibility of the ssDNA in the bubbles was larger than in real DNA (see section 3.5 for more details). We expect that if we erase these interactions, the bubbles will grow and the plectoneme with negative writhe might not be created.

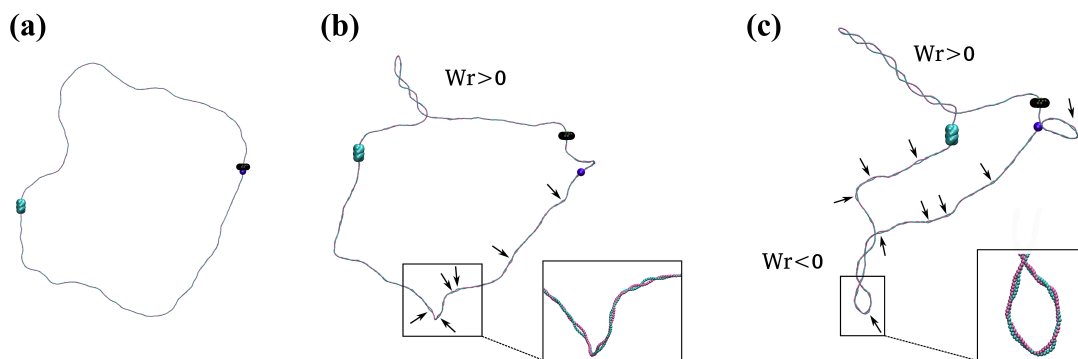


Figure 5.2 *Melting during transcription.* The snapshots from the simulations displayed in this panel do not show the DNA beads, they show the hydrogen bond sites (patches) as small beads in pink and cyan. The TB is represented by the large beads in cyan and the PolII complex is shown in black. In addition a purple bead is shown as references of the relative motion between DNA and the polymerase. (a) shows the configuration at $t = 0$, when ϵ_{HB} is about to be quenched. (b) After the quench in hydrogen bond energy the plectoneme with positive writhe is the first to appear in the simulations, and the section behind the polymerase starts melting. The arrows show the positions of the bubbles. (c) After more stress is accumulated in the system, particularly behind the polymerase, a plectoneme with negative writhe appears. Stably denatured bubbles localise at regions of high curvature (tips of plectonemes, in agreement with reference [143]).

5.3 Nucleosomes can be unwrapped by transcription at a distance

From now on we use the model as original parametrised (with $\epsilon_{\text{HB}} = 6.0k_B T$ to avoid melting) to show how this “action at a distance” effect might act on proteins; specifically we focus on the effect on nucleosomes, which is relevant *in vivo* for eukaryotic cells. To this end we construct a simple model histone octamer made up from a spherical bead of diameter 10 nm, with 20 “sticky patches”

tracing a left-handed helical path on its surface [162], and add a short range attraction between the patches and the DNA beads. This model can recapitulate the self-assembly of a realistic nucleosome (Fig. 5.3(a)). We started from a closed loop of DNA with $Lk = Lk_0$ (i.e. it is in a relaxed state), with an inactive (no force applied) polymerase on one side, and an octamer positioned on the other; when the simulation starts the DNA proceeds to wrap around the octamer due to the attraction to the patches. Our nucleosome is consistent with the “linking number paradox” [161], meaning that while the DNA completes 1.7 turns around the histone core, the wrapped section is also slightly twisted so the net change in linking number stored within the nucleosome is only approximately $\Delta Lk_{\text{nuc}} \approx -1$ (the plot in Fig. 5.3(a) shows that while through most of the DNA molecule $\theta(x) - \theta_0 \simeq 0$, the segment wrapped around the nucleosome is overtwisted $\theta(x) - \theta_0 > 0$). Overall for the entire loop we measure $Wr - Wr_0 = -0.657$ and $Tw - Tw_0 = 0.657$; no DNA crossings are formed during nucleosome assembly, indicating that to balance ΔLk_{nuc} , writhe must be spread through the part of the loop which is not within the nucleosome. We use this configuration (snapshot at the right of Fig. 5.3(a)) as an initial condition for simulations with an activated polymerase.

In order to isolate and independently investigate the effect of negative and positive supercoiling on the nucleosomal structure we place a topological barrier either directly in front of, or directly behind the nucleosome, and then activate the polymerase. In Fig. 5.3(b) we plot the local twist $\theta(x) - \theta_0$ at a time $t = 3000 \tau_{LJ}$ after the polymerase was activated, for the case where the topological barrier is placed so as to subject the nucleosome to positive supercoiling. Negative twist accumulates between the barrier and the polIII, while positive twist spreads towards the nucleosome (indicated as a blue segment). Importantly, at this time we observe that the DNA in the nucleosome starts to unwrap; eventually it completely unfolds, as shown in the right snapshot of Fig. 5.3(b). In marked contrast, when the barrier is placed such that the nucleosome is subject to negative supercoiling, i.e. isolated from the positive twist, there is no unravelling (Fig. 5.3(c)), and actually the nucleosome becomes more stable (see also Fig. C.6 in Appendix C.7).

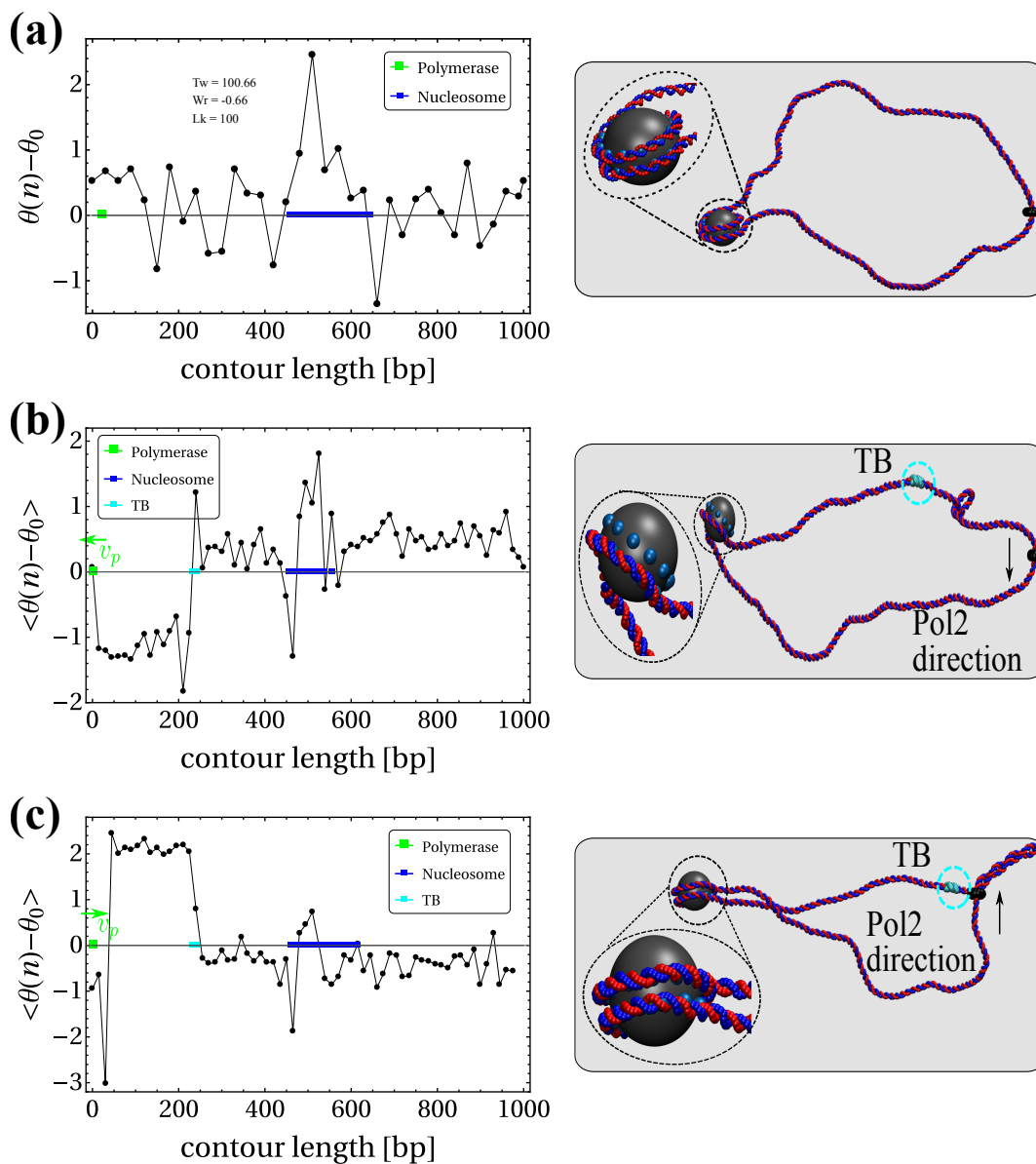


Figure 5.3 (a) Plot shows how the twist angle $\theta(n)$ between the n and $n+1$ bps varies around the loop – the region of DNA attached to the nucleosome is displayed in blue, there the DNA is slightly over-twisted. The snapshot shows the initial configuration: an equilibrated 1000 bp dsDNA ring with $Lk = Lk_0$, and with a fully wrapped nucleosome positioned on the opposite side to the (initially inactive) polymerase. The patches in the model nucleosome attract the DNA beads and follow a left-handed path. We measure $Wr - Wr_0 = -0.657$ and $Tw - Tw_0 = 0.657$ for the snapshot shown, confirming the conservation of Lk . (b) Configuration where a TB (shown in cyan) has been introduced to isolate the nucleosome from the negative twist produced by the polymerase. Plot shows $\theta(n)$ after the polymerase has moved 18 bp, averaged over 50 independent replica simulations. Snapshot shows that at this point the DNA has started to unwrap from the nucleosome. (c) Configuration where the nucleosome is instead isolated from the positive twist. Again plot shows $\theta(n)$ after the polymerase has moved 50 bp, averaged over 50 simulations. The nucleosome remains attached to the DNA during the whole simulation (snapshot). Further plots shown in Appendix C.7 reveal that in this case the nucleosome actually becomes more stable.

It has been suggested that *in vivo* the positive supercoils generated ahead of a transcribing polymerase are removed by topoisomerase enzymes as it goes along – our simulations suggest that it may not be necessary to remove all of the generated supercoiling, as this could naturally promote the unfolding of nucleosomes and the disassociation of histones in the down-stream region. That is to say, the twin supercoiling domain might facilitate the removal of obstacles to transcription, and thus increase its efficiency. While we have not attempted to realistically represent the DNA-histone interactions here, our results clearly predict that downstream nucleosomes should become less stable. Finally, it is intriguing to note that the action of the polIII can unwrap a nucleosome without any direct contact with it – it occurs as a result of the supercoiling travelling along the DNA.

5.4 Two modes of supercoiling relaxation

Above we observed that supercoiling generated by a polymerase spreads along the DNA to act at a distance; indeed previous 1-D modelling work has hypothesised that supercoiling moves diffusively [107, 160]. In this section we use our 3-D simulations to better understand how the independent modes of twist and writhe relax and distribute themselves through the dsDNA molecule.

First we consider twist relaxation. We initialise a 1000 bp, dsDNA loop in a non-relaxed state $Tw - Tw_0 = 3$, this is done by fixing the twist angle contained within a 30 bp long segment at $\theta(n) = 72^\circ$, away from the preferred value of $\theta_0 = 36^\circ$; we then run the simulation with this imposed constraint to generate a molecule where the rest of the chain is in a steady state conformation (we monitor the radius of gyration, R_g , and run until this has reached a steady value). We then release the constraint, and allow the twist locked in that short segment to freely relax until the molecule reaches an equilibrium conformation. The diffusion of the twist is monitored by measuring its local value along the molecule (see Appendix C.8) and averaging this over one thousand independent replica simulations (Fig. 5.4(a)).

The twist does indeed spread diffusively. Writing down and solving analytically a diffusion equation with an initial condition which corresponds to our

simulation we obtain an equation for the twist angle

$$\theta(x, t) = \theta_0 + \frac{\theta_0}{2} \left[\operatorname{erf} \left(\frac{x - x_i}{2\sqrt{Dt}} \right) - \operatorname{erf} \left(\frac{x - x_f}{2\sqrt{Dt}} \right) \right], \quad (5.1)$$

where $\theta_0 = 36^\circ$ is the preferred angle and x_i and x_f are the initial and final base-pair positions of the segment storing the deformation respectively. By fitting this profile to our simulation data we can extract a diffusion coefficient $D_{+\text{Tw}} = 478 \pm 9 \text{ bp}^2/\tau_{LJ}$. A similar but slightly smaller value or $D_{-\text{Tw}} = 312 \pm 12 \text{ bp}^2/\tau_{LJ}$, was obtained for an undertwisted chain, initialized with $\theta_n = 0^\circ$ within the same interval. These numerical estimates are robust for different lengths of the over/under-twisted segment.

Second, in order to study the relaxation of writhe, we consider initial configurations which display stable plectonemic structures (localised writhe) and a twist which is close to the relaxed value ($\theta(n) \simeq \theta_0$); i.e., we use configurations from the end of simulations shown in Fig. 5.1a(iii). We reset the time to $t = 0$ and remove both the polII complex and the topological barrier, and then allow the dsDNA molecule to relax to an equilibrium state. The evolution of supercoiling is monitored by computing the local writhe $\chi(n, t)$ [163, 164] (see Appendix C.9 for details). In Fig. 5.4(b), we show typical profiles for $\chi(n, t)$ at different times $t > 0$. The peaks and troughs in this function identify the locations of the positively and negatively writhed plectoneme tips; we can quantify the evolution in time of both the magnitude of local writhe and its localisation along the DNA [163]. This allows us to characterise the relaxation and diffusional dynamics of plectonemes.

We observe that the relaxation of plectonemes occurs in two steps: an initial fast relaxation followed by a slowing down at longer times (Fig. 5.4c). This can be explained with the following argument: high levels of writhe stored in the plectonemes carry conformational stress that is quickly released as soon as the chain is allowed to relax. Then, at later times, the remaining writhe becomes more delocalised (as confirmed by the broadening of the peaks in $\chi(n)$); this means that there is a lower conformational stress within these regions. This is confirmed by the results shown in Appendix C.10, where the total bending energy of the system displays the same behaviour. As a consequence, the long-time relaxation of the supercoiling occurs through a slower process. We find that a fit to the different regions (fast and slow relaxations) with an exponential decay agrees with the

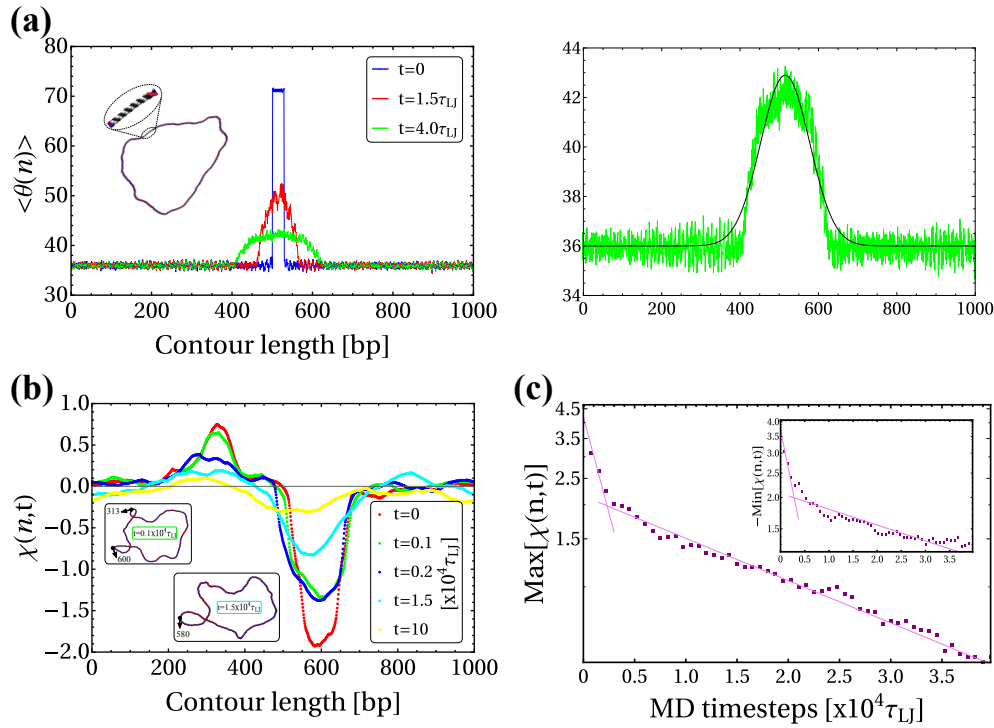


Figure 5.4 (a) Plot showing local twist angle $\theta(n)$ averaged over 1000 simulations. Different colours show the values for different times after the constraint was removed. In the inset, a snapshot of the initial configuration is depicted. The right plot shows the data at $t = 4 \times 10^7 \tau_{LJ}$ in green and the black line is the fit to Eq.(5.1). (b) A local writhe measure $\chi(n,t)$ (see Appendix C.9) is shown at different time points as a DNA loop (which initially had two plectonemes) relaxes after removing the TB and the polII from the system. Insets show snapshots of configurations at $t = 0.1 \times 10^4 \tau_{LJ}$ and $t = 1.5 \times 10^4 \tau_{LJ}$. (c) Log-linear plot of the maximum of $\chi(n,t)$ as a function of time. Purple squares are the data extracted from simulations. The solid lines represent the fit to this data with an exponential decay function $e^{-t/\tau}$ in the two different regions. Each region gives a decay constant (τ_1^+ and τ_2^+). In the inset the same information is depicted for the magnitude of the minimum of $\chi(n,t)$. More details are provided in the Appendix C.9, where we also show the same analysis for different initial configurations (see Fig. C.8). These configurations differ in the amplitude of the writhe and the number of plectonemes at $t = 0$.

observed data (Fig. 5.4(c)), and the relaxation time scale at long times is about an order of magnitude larger than the initial one. Remarkably, we find that the timescales associated with the relaxation of negatively writhed plectonemes are systematically longer (by about a factor of two) than those associated with positive writhe. We believe that this is again in agreement with the asymmetric torsional rigidity of dsDNA [74], i.e. negative writhe has less steric repulsion than

positive one thus entailing a lower conformational stress (see Appendix C.6), in turn reflected by a slower relaxation. We confirmed this observation through independent replicas that displayed different levels of local writhe stored within the plectonemes. The average timescales of positive writhe relaxation were $\tau_1^+ \simeq 2.9 \cdot 10^3 \tau_{LJ}$ and $\tau_2^+ \simeq 29 \cdot 10^3 \tau_{LJ}$ whereas for negative writhe we measured $\tau_1^- \simeq 4.4 \cdot 10^3 \tau_{LJ}$ and $\tau_2^- \simeq 44.5 \cdot 10^3 \tau_{LJ}$, see Appendix C.9 for details.

5.5 Transcription without topological constraints

Above we set up our simulations with a topological barrier, so as to observe the effect of the twin supercoil domains. One might expect that in absence of such a barrier, the supercoiling produced on each side of the polIII complex would travel around the DNA and cancel out, meaning no plectonemes would form. However, we show here that this is only true in the limit of low applied force, i.e. when the time it takes for the twist to diffuse around the DNA loop is smaller than the rate at which supercoils are injected by the polymerase. In other words, there is a competition between the rate of injection and the rate of dissipation that may favour or disfavour the creation of plectonemes.

To show the effect of this competition we perform a similar simulation to that described in Fig. 5.1, but without a topological barrier; we compute the total unsigned writhe $\zeta(t)$ (see Appendix C.3 for a definition) as a function of time after transcription is initiated. This quantifies the appearance of DNA crossings, irrespectively of their sign [163, 164]. $\zeta(t)$ is zero at the beginning of a simulation and reaches a steady state $\bar{\zeta}$ at large times. When plotted as a function of the force (Fig. 5.5), $\bar{\zeta}$ shows a transition between a relaxed ($\bar{\zeta} = 0$) and a writhed ($\bar{\zeta} > 0$) regime at forces $F_p \simeq 20 \epsilon/\sigma$. As expected there is also a dependence on the length of the DNA loop – for a larger loop the twist has to diffuse over a greater distance, so $\bar{\zeta}$ starts to show non-zero values at lower force.

Though we expect that *in vivo* the force generated by an elongating polymerase is relatively constant, these results demonstrate that the relevance of the twin supercoiling domain – its ability to generate plectonemes or affect protein binding – depends on the relative rate of supercoil injection and dissipation (the latter being determined by the length of DNA and any topological constraints it

is subject to).

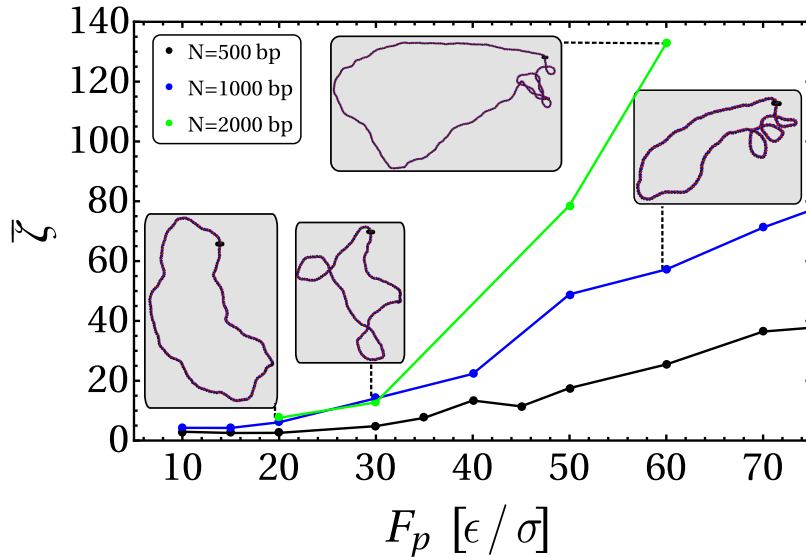


Figure 5.5 Plot showing the unsigned writhe $\bar{\zeta}$ once it has reached steady state for different values of the force applied by one polymerase. Values for $L = 500$ (black), 1000 (blue) and 2000 (green) bp DNA molecules are shown. Connecting lines are a guide for the eye. Insets show typical equilibrated configurations for the $F_p = 20, 30$ and $60 \epsilon/\sigma$ cases with $L = 1000$ bp, and the $F_p = 60 \epsilon/\sigma$ case with $L = 2000$ bp.

5.6 Summary

In this work we used the coarse-grained model introduced in Chapter 3 to study the generation of supercoiling by RNA polymerase. We confirmed that when the rotational motion of the polymerase is hindered, its action generates a “twin supercoiling domain”, where DNA in front of the polymerase becomes positively supercoiled, and DNA behind it becomes negatively supercoiled. In most simulations we studied a geometry where the polymerase was held in place, but have also confirmed that the twin supercoiling domain is still generated for the case of a mobile polymerase, due to it experiencing a larger rotational drag than the DNA. Importantly our simulations resolved the full three-dimensional DNA dynamics, going beyond recent works which have studied the dynamics of supercoiling in one-dimension [159, 162].

The main result of our work is that the action of the polymerase can operate at a distance. Contrary to typical text book pictures of the twin supercoiling domain, we found that the generated twist diffuses away from the polymerase and generates writhe (in the form of plectonemes) some distance away. This could be beneficial for the transcription process, because it gives some room for the polymerase to transcribe a segment of DNA before reaching the plectoneme structure created by its own movement. This plectoneme could interfere with the elongation process (in the absence of topoisomerases).

We also demonstrated that the generated supercoiling can have an effect on DNA-protein binding – focussing on the example of a nucleosome, we found that negative supercoiling reduces the stability of the complex which, in our simulations, led to DNA unwrapping from the histone octamer. Interestingly this also occurred “at a distance”, i.e. long before there was any physical contact between the polymerase and the nucleosome. This result is likely to be relevant *in vivo* for eukaryotes, since it implies that transcription itself can facilitate the destabilisation of downstream nucleosomes well ahead of the transcriptional machinery.

In order to better understand these results, we performed further simulations to measure the dynamical properties of supercoiling. We found that twist moves diffusively along the DNA and we measured its diffusion constant to be $D_{Tw} \approx 500 \text{ bp}^2/\tau_{LJ}$. We observe that writhe, in the form of plectonemes, relaxes much more slowly (consistent with it requiring large-scale rearrangements of the DNA), and does so with two characteristic time scales – initially writhe reduces following an exponential decay (with decay constant $\tau_{1Wr} \approx 10^3\tau_{LJ}$); later it still decays exponentially, but the rate slows down by about an order of magnitude ($\tau_{2Wr} \approx 10^4\tau_{LJ}$). This could be important after *termination* – a process in which the piece of RNA recently transcribed is released (for translation) and also the polymerase detaches from DNA; in this scenario most of the writhe produced by the transcription of a gene will quickly vanish, thereby leaving the same gene ready to be transcribed once more. Intriguingly, the relaxation timescales of negative writhe are systematically longer than the ones for positive writhe, indicating an asymmetric response, with negative writhe carrying less steric repulsion and thus entailing a lower conformational stress than the positive writhe.

6

DNA binding proteins

Abstract: In this chapter I show results from simulations of model proteins that selectively bind to the DNA major groove of a dsDNA ring molecule. As a consequence of this interaction the DNA structure deforms and this eventually leads to supercoiling. These observations suggest that supercoiling plays an important role in the regulation of DNA-protein binding interactions.

Proteins such as histones (seen in Chapter 5) bind to most parts of the chromosome, with only a slight preference for base sequence [4]. In eukaryotes, they serve mainly to package the long DNA into the small volume of the cell nucleus. By altering the structure of the chromosome, this compaction controls partially the access of enzymes to the promoter sites (start of the genes), affecting the transcription of genes [165, 166]. Therefore, this is one of the mechanisms to regulate gene expression. It is worth noting here that such a general mechanism does not explain how the expression of specific genes, related to important processes (growth of cells, cellular response to environmental stimuli or the housekeeping activities) is regulated. Thus, scientists have studied a different kind of protein that only binds to DNA with a specific nucleotide sequence [167]. The presence of these proteins alters the transcription, repair or unwinding of specific segments of DNA and therefore regulates the expression of particular genes.

Each of these specific-binding proteins identify small variations (generated by the base sequence) on the geometry of the DNA grooves [168, 169]. In addition, the DNA sequence provides a chemical signature that is recognised by chemically-complementary sections of the protein, this promotes the interaction between the protein and the bases [170]. Experiments on the structures of phage [171] and bacterial [172] repressor proteins, suggest that special motifs (like α -helices or β -sheets) that protrude from their surface, have the ideal structure to fit into the DNA grooves [173] and reach the bases. However, in some cases like the TATA binding protein, the DNA is significantly deformed to accommodate the protein [174]. In principle, this deformation could cause the supercoiling of the molecule, or alternatively, the supercoiling could influence the binding interaction between DNA and proteins.

Recently, experiments have explored the effects of DNA topology on DNA-protein binding interactions. In reference [175] for example, the topology-dependent DNA binding was measured for several enzymes (helicase, type IB topoisomerase, type IIA topoisomerase, non-specific mitochondrial DNA binding proteins and type II restriction endonuclease). Each of the studied proteins show an increased affinity for more highly supercoiled DNA despite their differences in structure, function and cellular compartment. However, this type of experiments do not resolve how is that the binding alters the topology of DNA and more importantly if there is a supercoiling-binding feedback mechanism in turn. Here I use our coarse-grained dsDNA model with explicit phosphates (see Appendix A.1) to get some insight into these subjects. In particular, I study the interactions of a dsDNA ring with model proteins that bind specifically to the major groove, with no sequence affinity (our DNA model represents an homo-polymer).

6.1 Protein model

Instead of modelling an entire protein, we focus in the particular sections that recognise the DNA major groove. For example [4] the “fingers” of the *Zif268* protein, the α -helix of the *bZip* activator protein or the β -sheets of the met repressor protein. In the following, by protein I refer to these specific motifs unless otherwise stated. In Fig. 6.1 the representation of the structure of protein *Zif268* obtained from x-ray diffraction is shown.

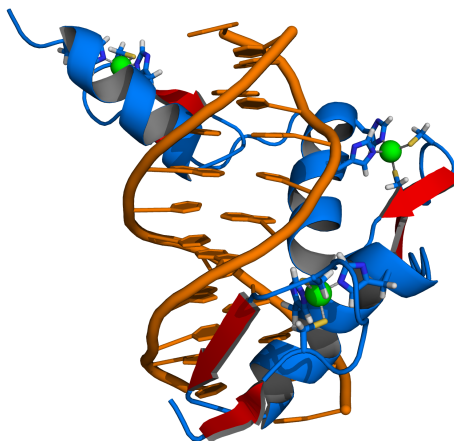


Figure 6.1 *Cartoon representation of a complex between DNA (shown in orange) and the Zif268 protein (shown in blue). The image is based on the x-ray structure of the protein. The 3 zinc finger motifs are shown in green and the coordinating residues of the middle zinc finger are highlighted. This image was obtained from reference [176].*

Proteins are modelled as a rigid body made of three patchy-beads (see Fig. 6.2(a)). The protein beads (shown in white) have an excluded volume of 2 nm which fit into the major but not the minor groove of the DNA, and the protein patches (shown in purple) have an excluded volume of 0.5 nm and can bind to the DNA base sites (cyan and pink). In Fig. 6.2(b) the relative position between consecutive patchy-beads in a protein is shown. This is obtained by rotating each patchy-bead complex from the previous one an angle of 72° about the centre of the purple bead, and rising it 0.68 nm along the direction perpendicular to the plane of rotation (the equivalent position between every three nucleotides in the B-DNA). Since the protein beads are spherical, the structure formed by the three complexes would involve a large degree of overlapping and therefore a large steric repulsion. To avoid this, we ensure that the components of the same protein (i.e. of the same rigid body) do not interact with each other.

The steric interaction between protein beads (with 2 nm size) and the DNA phosphates (with 0.2 nm size, see Appendix A.1) is set through the truncated and shifted Lennard-Jones potential of Eq. (3.4), with σ_s replaced by the average size of the two interacting particles; i.e. $\sigma_s = 1.1$ nm. We set the cut-off distance of the potential to $r_c = 2^{1/6}\sigma_s$ to model purely repulsive interactions. The proteins need to fit into the DNA major groove to interact with the bases, therefore a

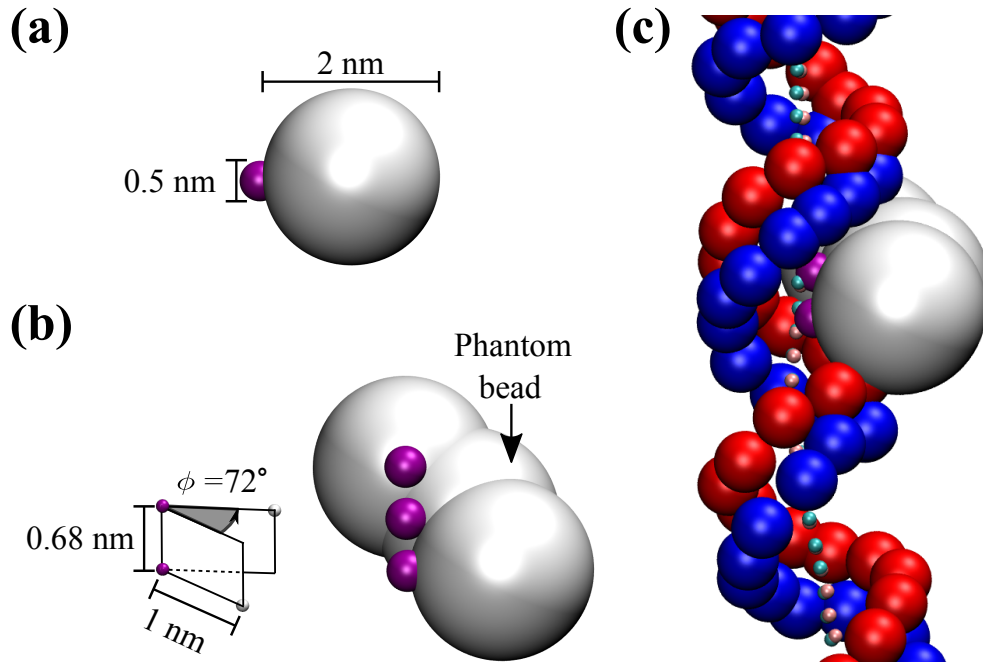


Figure 6.2 (a) The level of coarse-graining of one of the three components, that forms the functional part (of a protein) that recognises the major groove, is shown here. The white bead represents the excluded volume of the monomer. Its interaction with all the particles that do not belong to the same protein is purely repulsive (see main text). The small purple bead only interacts with the DNA base sites (patches) and phosphates. These purple particles are responsible for the binding interaction between proteins and DNA. (b) Left shows schematically the geometry between two consecutive protein beads. The white and purple spheres represent the centre of the protein beads and patches respectively. The right panel shows the right-handed helical conformation of the three patchy-beads in the same rigid body. To avoid large steric repulsions, the bead in the middle (phantom bead) do not interact sterically within the same protein motif. (c) One protein attached to the DNA major groove is shown. In this figure: the phosphates of the different strands are shown in red and blue respectively, the DNA patches are shown in cyan and pink, and the DNA beads are not shown to ease the visualisation.

hard-repulsive interaction between protein beads and DNA beads (with 1 nm size) was not desired and we chose instead a soft interaction, represented by:

$$U_{\text{soft}} = \epsilon_{\text{soft}} \left[1 + \cos \left(\frac{\pi r}{r_{cs}} \right) \right], \quad (6.1)$$

where $\epsilon_{\text{soft}} = 25k_B T$ is the strength of the potential, and the cutoff distance (r_{cs}) is set to the average size of the particles involved in the interaction (1.5 nm).

The protein patches (with 0.5 nm size) repel the DNA phosphates in the same way that protein beads repel the phosphates (through the potential Eq. (3.4), with $\sigma_s = 0.35$ nm). In order to model the interaction between protein patches and the DNA base sites (with 0 nm size, see Chapter 3) we use again the potential of Eq. (3.4), but this time $\sigma_s = 0.25$ nm and the cut-off distance is $r_c = 3\sigma_s$, so as to include attractive interactions. The strength of this potential represents the effective binding energy ϵ_{bind} between proteins and DNA. Due to the steric interactions, proteins only bind to the DNA major groove (see Fig. 6.2(c)).

The initial configuration of the system includes a dsDNA ring ($N = 1000$ bp long) that is not supercoiled ($\text{Lk} = \text{Lk}_0 = N/p$) and two thousand proteins ($M = 2000$) with no attractive interaction to DNA ($\epsilon_{\text{bind}} = 0$). All the particles lie inside a cubic box of side 140 nm long and periodic boundary conditions. We then let the system equilibrate to a relaxed state before introducing the binding of proteins to DNA.

6.2 Binding curve and the dynamics of the DNA-proteins system

After performing a sudden quench of ϵ_{bind} we record the time evolution of the fraction (Υ) of base-pairs in contact with proteins, until the steady state (characterized by a constant value of Υ) is reached. The plot of the equilibrium value $\langle \Upsilon \rangle$ as a function of the binding energy is identified with the binding curve for the protein-DNA system. This is depicted in Fig. 6.3(a), where we can appreciate that there is a critical binding energy $\epsilon_{\text{bind}}^* = 4k_B T$. Below this energy, proteins do not bind to DNA ($\langle \Upsilon \rangle = 0$) and for energies above it, $\langle \Upsilon \rangle$ jumps discontinuously to a non zero value: the expected behaviour of a first order transition (see section 4.2.1).

The state of each base-pair (bound or not to a protein) against time is captured by the kymographs in Fig. 6.3. For an energy $\epsilon_{\text{bind}} = 8k_B T$ (Fig. 6.3(b)) the proteins start binding at several locations along the DNA (covering a fraction of 9.6% of its contour length at $t = 25 \times 10^4 \tau_{LJ}$). When the energy is decreased to a value $\epsilon_{\text{bind}} = 6k_B T$ (Fig. 6.3(c)) the binding of proteins starts at fewer discrete sites (this time proteins cover a fraction of 5% of the DNA contour

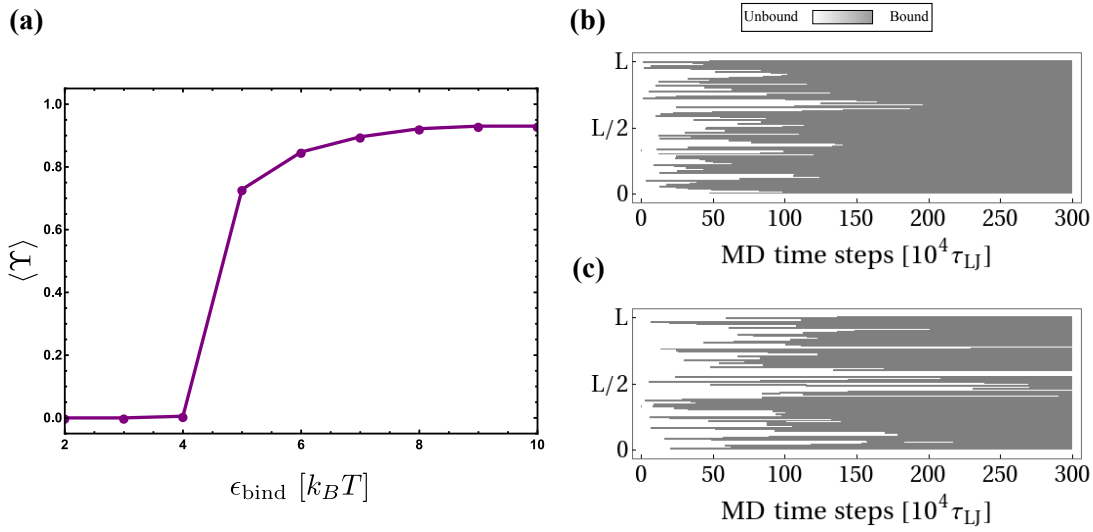


Figure 6.3 (a) Protein binding curve for a ring DNA molecule 1000 bp long, initialized in a non supercoiled state ($Lk = Lk_0$). The dots show the fraction of base-pairs in equilibrium $\langle \Upsilon \rangle$ bound to any of the proteins as a function of the binding energy ϵ_{bind} . The line connecting dots is only a guide to the eye and it should not be confused with the continuity of the plot (i.e. with a second order transition). The critical energy of the transition between the bound and unbound states is $\epsilon_{\text{bind}}^* = 4k_B T$. (b)-(c) represent the kymographs from BD simulations. At time $t=0$ the binding energy is quenched and the local state of every base-pair is recorded (gray for bound and white for unbound). Panel (b) shows the case for $\epsilon_{\text{bind}} = 8k_B T$ and panel (c) for $\epsilon_{\text{bind}} = 6k_B T$.

length at $t = 25 \times 10^4 \tau_{LJ}$) along the DNA molecule. In both cases, the binding of new proteins appears to spread near to these initial binding sites, suggesting a cooperative effect, in agreement with the first order transition mentioned above. This may be explained as it is energetically favourable for the proteins to bind near to sites that have already been deformed, rather than going through the trouble of deforming the DNA themselves before binding. This should be more evident as we get closer to the critical binding energy (ϵ_{bind}^*).

The trajectories from BD simulations showed that depending on the binding energy, the proteins either remain attached to the DNA in the same initial location where they were bound, or they detach from DNA. In the latter scenario, sometimes the protein reattaches at a different location and sometimes it is kicked out by a second protein that takes its place and binds to DNA. This is important because it has been shown that the rate at which proteins bind to, and unbind from, DNA can arise in several biologically relevant scenarios [177]. For instance, a complex of transcription factors and an RNA polymerase might

stabilize a promoter-enhancer loop; upon transcription termination, the complex could dissociate and the loop disappear [5, 178].

6.3 Protein binding induces supercoiling

The binding of proteins produces the decrease in the local twist angle near to the binding sites. In Fig. 6.4(a) the time evolution of the total writhe (see Appendix C.3 for details in the computation) of the DNA after the quench in the binding energy is shown for $\epsilon_{\text{bind}} = 4, \dots, 8k_B T$. When looked at together with the kymographs in Fig. 6.3, one notes that when enough proteins attach to DNA, the molecule writhes (positively) to compensate the change in twist. To better appreciate this, in Fig. 6.4(b) we show that the writhe increases with the fraction of base-pairs in contact with proteins, for the case $\epsilon_{\text{bind}} = 6k_B T$. The snapshots from the simulation (see Fig. 6.4(c)) show configurations with positive supercoiling after several proteins attach to DNA when $\epsilon_{\text{bind}} = 4, 6$ and $8 k_B T$.

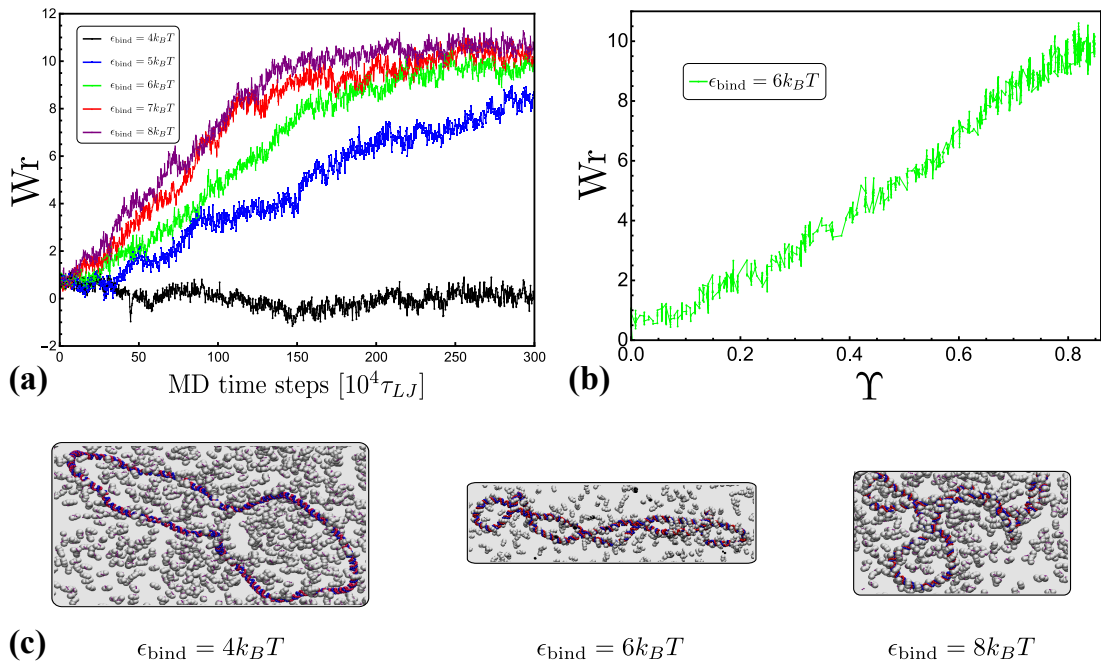


Figure 6.4 (a) Writhe evolution after the quench in binding energy. For $\epsilon_{\text{bind}}^* = 4k_B T$, $W_r \approx 0$ because proteins do not attach to DNA. For $\epsilon_{\text{bind}} > \epsilon_{\text{bind}}^*$ the proteins bind to DNA and W_r increases. (b) W_r as a function of Υ for $\epsilon_{\text{bind}} = 6 k_B T$. (c) Snapshots from simulations of molecules when $\epsilon_{\text{bind}} = 4, 6$ and $8 k_B T$.

From the previous results we infer (and expect to corroborate in the near future) that if we start from a molecule that is positively supercoiled, it would be more difficult for the proteins to bind. Thus, in general one would expect that the critical binding energy should depend on the level of supercoiling.

6.4 Summary

A novel coarse-grained model for proteins that bind to the DNA major groove has been introduced. The results in this chapter show that the phase transition between the bound-unbound states depends on the order parameter ϵ_{bind} and resembles a first order transition. When near to the critical energy (ϵ_{bind}^*), the binding of proteins looks cooperative. The local twist angle of DNA is modified by the binding of proteins and this yields to the writhing of the molecule. It would be interesting to study whether the level of supercoiling produces a shift in the critical binding energy or how the system (formed by the DNA and proteins) behaves during melting when decreasing the hydrogen bond energy ϵ_{HB} . In addition, the dependence on several parameters of the model, such as: the size of the protein beads, the dihedral angle between consecutive protein patchy-beads, the total number of proteins, among others, needs to be characterised in the future.

7

Conclusions

Recent improvements in experimental techniques, have allowed scientists to study in more detail the behaviour of DNA under different circumstances, thus providing relevant information about its elastic and topological properties. This progress requires a similar increase in the theoretical understanding of the DNA, in order to adequately interpret each experimental result and also to get some insight into the biological relevant scenario. With this in mind, I focused in a particular subject of interest that is the study of DNA supercoiling.

The results presented in this thesis certainly show that the structure of DNA is richer, more intricate and not as uniform as was first introduced in Chapter 2. Something as simple as constraint the movement of the ends of a linear DNA molecule, or close them into a circular shape (both common scenarios *in vivo*) has huge consequences in the geometry and dynamics of the polymer. These systems obey topological rules that are important in several biological processes. In particular, supercoiling because its richness, has provided a steady source of interesting problems and new challenges for physicists to think about. Some of them were here investigated using the coarse-grained computational model introduced in Chapter 3, which despite its simplicity, can provide us with some intuition into the still puzzling behaviour of DNA.

In Chapter 4 I have shown that the simulations reproduce experimental melting curves of DNA molecules that are either, topological free or topological constraint. In the context of a mean field theory proposed in this Chapter, I explained how the melting of DNA is regulated by the level of supercoiling and

in particular how the first order transition of tfDNA is hindered for tcDNA. A noteworthy behaviour from the simulations is that the denatured regions can store linking number. This becomes clearer when we think in the case of a completely melted dsDNA ring whose backbone has not been broken. This molecule has to preserve its linking number and therefore this has to be somehow distributed in the melted DNA.

The supercoiling dynamic during transcription has been explored in Chapter 5. One major finding is that in the presence of a topological barrier, the twin-supercoiling domain creates plectonemes of opposite sign at both sides of the polymerase, that: do not carry the same units of writhe and are not located at the same distance from the polymerase complex. This is opposite to the usual view of this process. I showed that the behaviour of the model can be understood in terms of the relaxation of twist and writhe. Furthermore, the local disruption of DNA in the vicinity of the polymerase can produce the unbinding of a model nucleosome located ahead of the polymerase. This is an effective action at a distance that might play a major role in the replication process.

In Chapter 6 I study a system composed by DNA and DNA-binding proteins. The preliminary results seem to suggest that there is a feedback mechanism between supercoiling and binding, but further analysis needs to be carried out in this direction.

When asking about the relevance of these theoretical results in the real case, it should be taken into account that some simplifications and approximations have been made, which allowed us to solve and explore particular aspects of the systems mentioned above. The results and techniques discussed in this thesis, should be considered as a small contribution towards the construction of a broader picture in the understanding of the complex mechanisms of DNA. Yet, I have tried to explore situations that could be relevant and hopefully tested in experiments.

Finally, the computational approach employed in this thesis, has proved to be relatively simple and fairly efficient to study and describe some biological processes. That is why I consider that would be interesting to use it in the near future to explore the interplay of for example: transcription in the presence of major-groove binding proteins, several nucleosomes or even during melting. This could be important in the understanding of gene regulation.

A

DNA model: grooves and closure procedure

A.1 Major and minor grooves

As mentioned in Section 3.6, the presence of grooves can be incorporated into our model by adding a third spherical monomer per nucleotide, representing the phosphate group. In Fig. A.1 a top view of a base-pair is sketched. For one of the nucleotides, the excluded volume of the bead is shown in blue (with diameter of 1 nm), the patch is marked with pink (with no excluded volume) and its corresponding phosphate is shown in green at a distance of 1.02 nm from the center of the helix axis and with an excluded volume of 0.2 nm. Similarly, for the complementary nucleotide the excluded volume of the bead is shown in red, the patch in cyan and the phosphate in yellow. All the interactions mentioned in Section 3.2 remain unchanged.

Including the phosphate in the model will add an additional degree of freedom per nucleotide, which is regulated by a harmonic angle interaction between two consecutive patches and a phosphate in the same strand, similar to the one imposing the planarity between consecutive nucleotides (see Eq. (3.9)) and with the same value of the involved parameters (see Table 3.1). The smallest angle between the two phosphates in a base-pair and the helix axis (shown in light grey in Fig. A.1) is $\phi = 130^\circ$ and results in the minor groove when following the helical path of the dsDNA, as can be seen in Fig. A.2(a). The conjugate angle

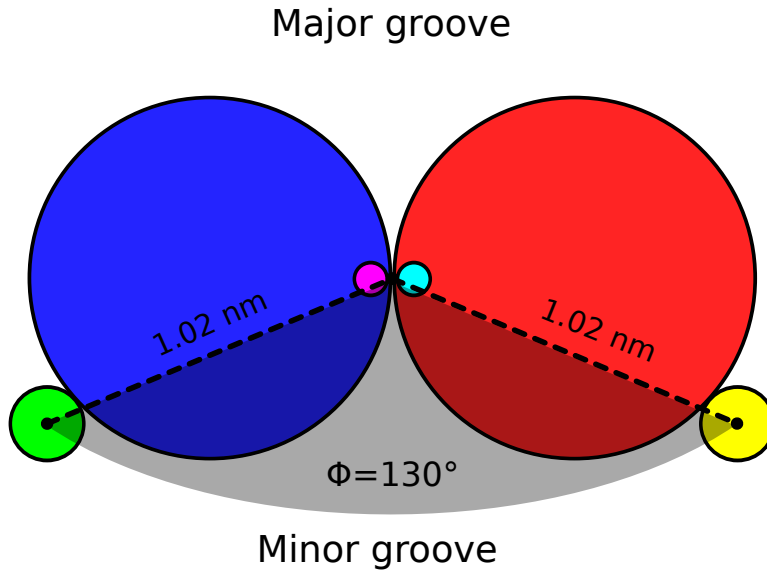


Figure A.1 Sketch of a base-pair with the phosphate-group included explicitly. The bead-patch-phosphate complex (Blue-Pink-Green for the first strand and Red-Cyan-Yellow for the second strand) acts as a rigid body representing one nucleotide.

is 230° and it gives rise to the major groove. If the pitch of the chain is 10 bp and therefore the helical angle between two consecutive base pairs is 36° , then the minor groove is made by $130/36 = 3.62$ bp with a total length of around $3.62 \times 0.34 = 1.2$ nm. Correspondingly, the length of the major groove is 2.2 nm.

One way of determining the presence of grooves in our model is by comparing the average distance between one fixed phosphate chosen randomly from one of the strands and the ten consecutive phosphates in the complementary strand (including the one in the same base-pair). Since the grooves have a different size, the resulting plots differ from one another depending on the position of the fixed phosphate, whether it is on the first or the second, strand. In Fig. A.2(b) we show these plots where the top graphic displays a global minimum that is related to the presence of the minor groove. In this plot, the blue dots represent the distance (averaged over time and base-pairs) between phosphates, measured from the simulation of a chain made by 300 bp. The spline interpolation of the blue dots is shown in black and the red point represents the inferred position of the the minor groove, located at 3.62 bp with a distance of 1.22 nm. The bottom graphic in Fig. A.2(b) is related to the major groove, in this case the interpolation gives a total width of 2.18 nm. The length of the grooves can be modified by changing the angle (ϕ) between the phosphates, but as the pitch remains constant

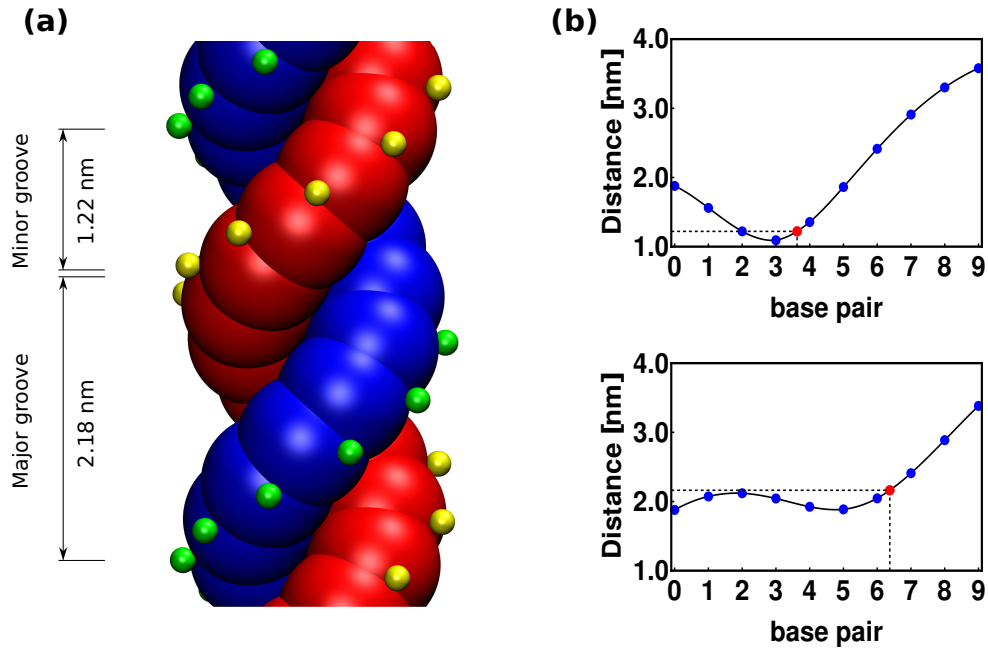


Figure A.2 (a) Representation of the double-stranded DNA model with phosphates. Interactive and ghost beads are shown in red for one of the strands with the corresponding phosphates in yellow. The complementary strand and its phosphates are shown in blue and green respectively. The hydrogen bond sites (patches) are not depicted in this picture. (b) The top graphic shows (blue dots) the average distance between one phosphate chosen randomly from the red strand and the ten consecutive phosphates in the blue strand. The width of the minor groove can then be extracted from the interpolation curve (in black) at a distance of 3.62 bp, giving a value of 1.22 nm. This is when the green phosphate is on top of the previously chosen phosphate in the red strand, as depicted in (a). In a similar way the bottom graphic shows the average distance between one phosphate chosen randomly from the blue strand and the ten consecutive phosphates in the complementary strand. This time the major groove is extrapolated from the black curve at a distance of 6.38 bp, giving a width of 2.18 nm.

the sum of the total widths of the grooves will always be 3.4 nm.

A.2 Closure for linear DNA

In this section we review the procedure to compute the linking number of an open segment of dsDNA. For clarity we report a schematic in Fig. A.3. Given two curves C_R and C_B mapping the interval $I = [0 : 1] \rightarrow \mathbb{R}^3$, it is possible to formally compute their linking number only if closed, *i.e.* $C_R(0) = C_R(1)$ and $C_B(0) = C_B(1)$. For a linear open segment of dsDNA, a pair of closed

strands can be defined by considering the vectors tangent to the terminal pair of beads of the two single strands forming the dsDNA segment and extending the curves away from the pair of strands. After reaching a certain distance by following, for instance, \mathbf{t}_{1R} and \mathbf{t}_{2R} , one can close the contour by defining a vector \mathbf{f}_R that joins the two new terminal beads (see Fig. A.3). By following this procedure one can finally construct a pair of closed oriented curves γ_R and γ_B , for instance “stitching” C_R , $\mathbf{t}_{1R}, \mathbf{f}_R, -\mathbf{t}_{2R}$, and similarly for the blue curve. Their linking number can be computed through the numerical evaluation of the double integral

$$Lk(C_R, C_B) = \frac{1}{4\pi} \int_{\gamma_R} \int_{\gamma_B} \frac{|\mathbf{r}_R - \mathbf{r}_B|}{|\mathbf{r}_R - \mathbf{r}_B|^3} \cdot (d\mathbf{r}_R \times d\mathbf{r}_B), \quad (\text{A.1})$$

where \mathbf{r}_R and \mathbf{r}_B are the vectors defining the position of the segments along the curves γ_R and γ_B , respectively. If the centreline running through the pair of curves has no self-intersections (null writhe) then the linking number is equal to the twist.

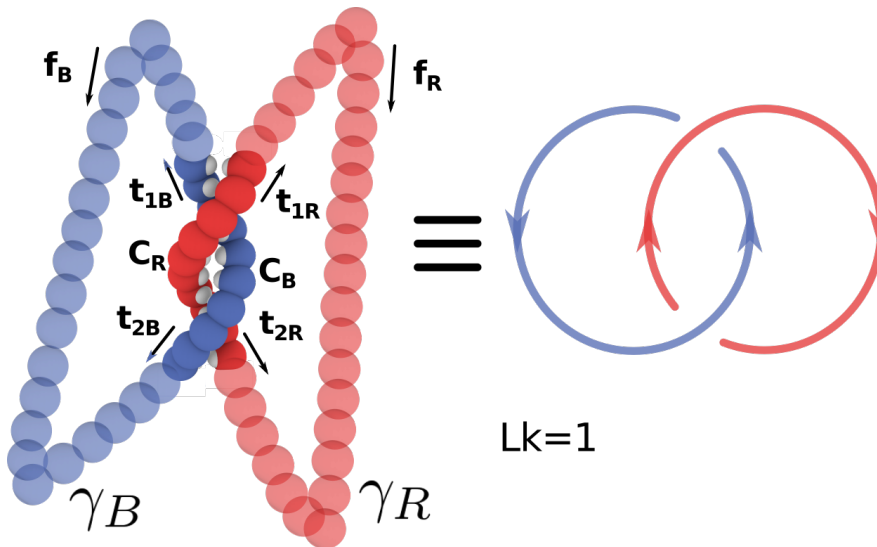


Figure A.3 The “closure” procedure can be performed on a pair of linear open curves to construct a closed pair whose linking number can be formally defined through the Gauss’ integral (see Eq. (A.1)). In this case the curves are linked once. See text for further details.

It is also worth mentioning that tightly wound curves, such as those obtained from dsDNA configurations, can lead to imprecise numerical evaluation of the integrals in Eq. (A.1). In fact, the computation of Lk can become unreliable

when $|\mathbf{r}_R - \mathbf{r}_B| \simeq d\mathbf{r}_R \simeq d\mathbf{r}_B$. The numerical evaluation can be arbitrarily improved by replacing the DNA backbones by contours more finely interspersed with points, *i.e.* enhancing the resolution of the integral by decreasing the infinitesimal element $d\mathbf{r}$. Clearly, this can slow down the computation of Lk . We found a good compromise between precision and speed by adding three intermediate points every pair of beads for which we consistently measured the correct linking number during topology-preserving simulations (for instance by considering circular dsDNA).

B

Details on the calculations for the melting of tcDNA

B.1 Supercoiling field

One of the central points of Chapter 4 is that we define a *local* supercoiling field $\sigma(x, t)$ as

$$\sigma(x, t) \equiv \frac{Lk(x, t) - Lk_0}{Lk_0}, \quad (\text{B.1})$$

where $Lk(x, t)$ is the effective linking number between the two single-strands at location x and time t along the DNA and Lk_0 is the linking number in the relaxed state (i.e. 1 per pitch). In order to have a meaningful measure of the local linking number, it is useful to consider a stretch of the DNA in the range of one pitch length. In light of this, $Lk(x, t)$ can actually be computed by considering, for example, a sliding window of two pitch lengths (~ 20 bp) centred in x . The four strands selected within the window can be joined to form two closed curves (C_R and C_B), so that the Gauss integral

$$Lk(C_R, C_B) = \frac{1}{4\pi} \int_{C_R} \int_{C_B} \frac{\mathbf{r}_R - \mathbf{r}_B}{|\mathbf{r}_R - \mathbf{r}_B|^3} \cdot (d\mathbf{r}_R \times d\mathbf{r}_B), \quad (\text{B.2})$$

reflects the linking number between the two strands of the double-helix (see also below for more details on the closure procedure). If the window is too short (say, much shorter than a pitch length), the Gauss integral would always return zero;

on the other limit, too large windows (say, the size of the DNA molecule) cannot capture the local variations in supercoiling density. The best choice falls in the region of 2-3 pitch lengths.

This operational definition allows us to compute the local supercoiling field directly from our BD simulations. In addition, it allows us to unambiguously employ $\sigma(x, t)$ in our mean-field theory to capture the average behaviour (i.e. without fluctuations) of the local linking number $Lk(x, t)$ with respect to the one expected in the relaxed state Lk_0 . Fig. B.1 shows this approach for a negative supercoiled ring ($\sigma = -0.06$) and three different values of the bond energy. In Fig.B.1(a) we compute the local linking number $Lk(x, t)$ at time t , and its average over configurations for a chain with $\epsilon_{HB} = 0.9 k_B T$. This chain is closed everywhere except in the vicinity of base-pair $x = 800$, where the presence of the only denatured bubble accounts for the drop in the value of the linking number. In Fig.B.1(b) we show $Lk(x, t)$ for a fully denatured chain ($\epsilon_{HB} = 0.1 k_B T$, black) and a chain near to the transition ($\epsilon_{HB} = 0.18 k_B T$, red). In both, $Lk(x, t)$ fluctuates around the relaxed value $Lk_0 = 2$, with the second case showing a peak near to $x = 75$, consistent with the presence of a small closed region of the chain.

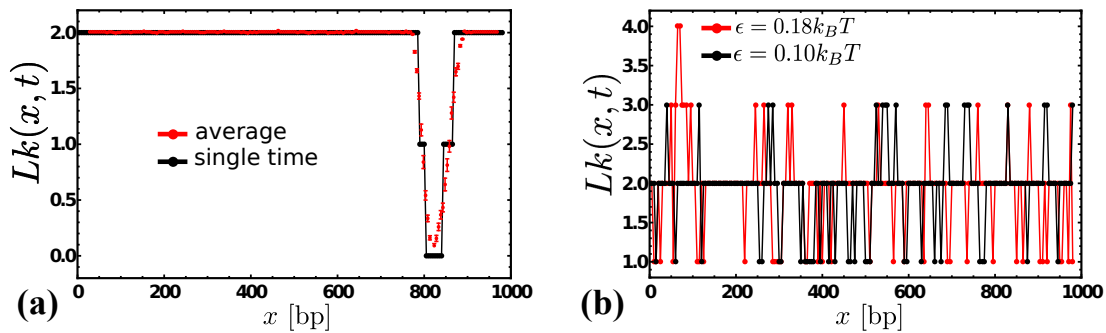


Figure B.1 Local supercoiling. (a) Shows the local linking number at timestep t , $Lk(x, t)$ (black points and line), computed as described in text (via a sliding window of 20bp) for a configuration in which the large part of the chain is intact (dsDNA) and a small part (to the far right) displays a denatured bubble. The plot also shows the time-average, $Lk(x)$ (in red), computed over a time-window in which the open denatured bubble displays little displacement. (b) Shows the local linking number for a configuration near the melting transition. The chain at $\epsilon_{HB} = 0.18 k_B T$ (red) shows a small region in which the linking number is higher (to the left), whereas the same chain at $\epsilon_{HB} = 0.1 k_B T$ is fully denatured. The value of the local $Lk(x, t)$ is seen to fluctuate around the relaxed value $Lk_0 = 2$.

B.2 The three parts of the free energy

Here we analyse the behaviour of each of the terms in the free energy (Eq.(4.3)) as independent quantities, in order to gain some insight into the system.

Free energy of the non-conserved parameter

According to the experimental melting curves shown in Fig.4.2, the tfDNA exhibits a first order transition. Depending on the temperature, the molecule can experience a transition between completely closed $\phi = 0$ and completely open $\phi > 0$. This means, that the free energy must have two local minima (corresponding to closed and open), and any of them can convert into the global minimum depending on the temperature of the system (see Fig.B.2). The simplest form of a free energy that satisfies these conditions is:

$$\beta f_\phi = \left(\frac{b^2}{4c} + 1 - a(T) \right) \phi^2 + b\phi^3 + c\phi^4. \quad (\text{B.3})$$

The extremes of the previous equation are obtained through the derivative of the free energy, $\frac{df_\phi}{d\phi} = 0$. This gives

$$\phi_1 = 0 \quad (\text{B.4})$$

$$\phi_2 = \frac{-3b + \sqrt{b^2 - 32c(1-a)}}{8c} \quad (\text{B.5})$$

$$\phi_3 = \frac{-3b - \sqrt{b^2 - 32c(1-a)}}{8c}. \quad (\text{B.6})$$

In the most general case, these solutions correspond to the presence of two minima (ϕ_1 and ϕ_2) and one local maximum (ϕ_3). When the system is in the helix state (closed), the global minimum of the free energy is located at $\phi_1 = 0$, and $f_\phi(\phi_1) = 0$ according to Eq.(B.3). At the critical temperature, the two minima produce the same value of the free energy ($f_\phi(\phi_1) = f_\phi(\phi_2)$), this imposes the value of $a(T_c) = a_c = 1$. It also ensures that the value of $\phi_3 = -b/4c$ and $\phi_2 = -b/2c$.

If the coefficient of the term in the free energy with the largest exponent is positive ($c > 0$), then f_ϕ will be positive at large values of ϕ , implying that the free energy is bounded from below. This ensures the description of a system which is globally stable. Without loss of generality we can set $\phi_2 = 1$ at the critical temperature, which will determine the value of $c = -b/2$, with $b < 0$.

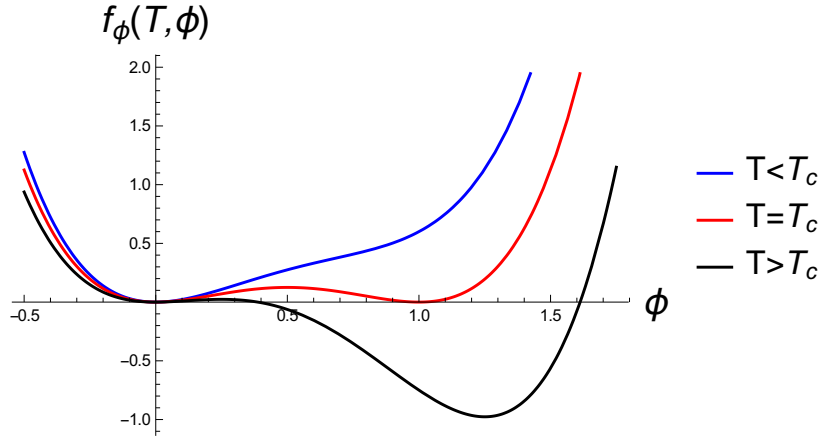


Figure B.2 *Free energy of the non-conserved parameter.* Profiles of the free energy of the non-conserved parameter (see Eq.(B.3)) for different values of the temperature when $b = -4$ and $c = -b/2 = 2$.

Free energy of the conserved parameter

When the chain is intact ($\phi = 0$), small torsional strains should be symmetric with respect to σ [118]. This means that the supercoiling field exhibits an inversion symmetry. Therefore, only even powers of σ are allowed in the expansion of the free energy:

$$\beta f_\sigma = a_\sigma \sigma^2 + b_\sigma \sigma^4, \quad (\text{B.7})$$

which satisfies $f_\sigma(\sigma) = f_\sigma(-\sigma)$. The extremes of the previous equation are:

$$\sigma_1 = 0 \quad (\text{B.8})$$

$$\sigma_2 = -\sqrt{-a_\sigma/2b_\sigma} \quad (\text{B.9})$$

$$\sigma_3 = \sqrt{-a_\sigma/2b_\sigma}, \quad (\text{B.10})$$

The requirement that the free energy is bounded from below implies that $b_\sigma > 0$. Therefore the three solutions shown in the previous equations would exist only if $a_\sigma < 0$. In the case that a_σ is also positive the system will have only one global minimum located at $\sigma = 0$. The two sceneries mentioned above are illustrated in Fig. B.3.

At the critical temperature, we expect the chain to denature locally and therefore $\phi = 1$. In this denatured region, the twist tends to zero and if the section is small enough, also the local writhe. Thus, we can make the approximation of $Lk = 0$ or what is the same $\sigma = (0 - Lk_0)/Lk_0 = -1$ in the bubbles. Since Eqs.(B.9 and B.10) imply $\sigma^2 = -a_\sigma/2b_\sigma$, then $b_\sigma = -a_\sigma/2$.

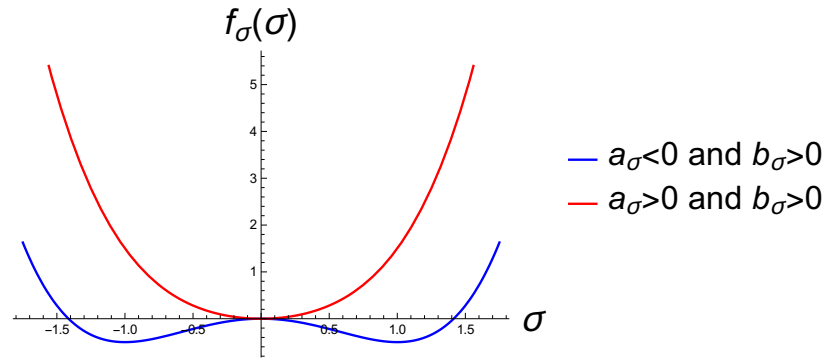


Figure B.3 *Free energy of the conserved parameter.* Profiles of the free energy of the conserved parameter (see Eq.(B.7)) for $a_{\text{sigma}} = -1.0$, $b_{\text{sigma}} = 0.5$ (blue) and $a_{\text{sigma}} = 1.0$, $b_{\text{sigma}} = 0.5$. When the coefficients in the expansion have different sign, the free energy shows two symmetric minima. When both coefficients have the same sign, there is only one minimum at $\sigma = 0$.

Coupling

The form of the coupling term between the two order parameters is

$$\chi\sigma\phi^2, \quad (\text{B.11})$$

which is thought to produce a linear dependence of $a(T)$ on the supercoiling at the critical temperature, i.e., that the denaturation temperature will depend on the level of supercoiling, as shown in Fig. 4.2. The total free energy results from

adding all the terms seen in this section.

B.3 Equations of motion: Model C

The equations describing the dynamics of the non-conserved field ϕ and the conserved one σ can be derived from the Landau free energy as [145]

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} &= -\Gamma_\phi \frac{\delta\mathcal{H}(\phi,\sigma)}{\delta\phi} \\ \frac{\partial\sigma(x,t)}{\partial t} &= \Gamma_\sigma \nabla^2 \frac{\delta\mathcal{H}(\phi,\sigma)}{\delta\sigma}\end{aligned}\quad (\text{B.12})$$

where Γ_ϕ and Γ_σ are relaxation constant, and \mathcal{H} denotes the total free energy,

$$\mathcal{H}(\phi,\sigma) = \int \left(f(\phi,\sigma) + \gamma_\phi (\nabla\phi)^2 + \gamma_\sigma (\nabla\sigma)^2 \right) dx. \quad (\text{B.13})$$

The equations of motion are therefore:

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} &= \\ &= -\Gamma_\phi \left[2 \left(1 - a + \frac{b^2}{2c} \right) \phi + 3b\phi^2 + 4c\phi^3 + 2\chi\phi\sigma - \gamma_\phi \nabla^2 \phi \right] \\ \frac{\partial\sigma(x,t)}{\partial t} &= \Gamma_\sigma \nabla^2 \left[2a_\sigma\sigma + 4b_\sigma\sigma^3 + \chi\phi^2 - \gamma_\sigma \nabla^2 \sigma \right].\end{aligned}\quad (\text{B.14})$$

Equations (B.14) are solved numerically on a 1D lattice of size $L = 1000$ via Euler method with integration time $dt = 10^{-3} t$ (t is the time corresponding to 1 simulation unit) and lattice spacing $dl = 1$. The parameters are set as before: $b = -4$, $c = -b/2$, $a_\sigma = 1$ and $b_\sigma = 1/2$. The surface tensions γ_σ and γ_ϕ are both set to 10 in units of squared lattice spacing. The mobility Γ_σ is set so that the supercoiling field σ relaxes more quickly than ϕ (as we expect to be the case for DNA during melting), i.e. $\Gamma_\sigma = 10$ (units $dl^2 t^{-1}$) while $\Gamma_\phi = 1$ (units t^{-1}).

The system is initialised with a small denatured bubble of size $l_b = 10 \ll L$ so that

$$\phi(x,0) = \begin{cases} 1 & \text{for } x \in \text{bubble} \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.15})$$

while the supercoiling field is set to

$$\sigma(x, 0) = \begin{cases} -1 & \text{for } x \in \text{bubble} \\ \frac{L\sigma_0 + l_b}{L - l_b} & \text{otherwise.} \end{cases} \quad (\text{B.16})$$

Finally, since the molecule is circular-closed, we set periodic boundary conditions along L for the values of the fields and their derivatives, i.e. $\phi(-L/2, t) = \phi(L/2, t)$, $\partial_x \phi(-L/2, t) = \partial_x \phi(L/2, t)$ and $\sigma(-L/2, t) = \sigma(L/2, t)$, $\partial_x \sigma(-L/2, t) = \partial_x \sigma(L/2, t)$ for any time t .

Importantly, we can model topologically unconstrained molecules by setting $\chi = 0$, for which the denaturation field decouples from σ . This scenario can be viewed as an approximation for nicked molecules where the supercoiling relaxes infinitely fast (through strand rotations at the nick).

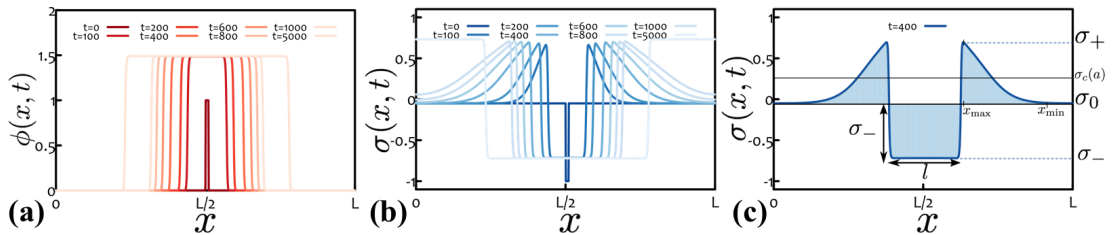


Figure B.4 (a-b) Evolution of the fields $\phi(x, t)$ and $\sigma(x, t)$ as a function of position x starting from a small denatured bubble in the middle of the chain. Panel (c) highlights the area within the denatured bubble ($= l(t)\sigma_-$) which must equal the area outside it; in other words, the supercoiling front forming outside the growing bubble has an area that grows linearly with $l(t)$. Because the maximum of the front remains approximately constant at σ_+ , its base must then grow as $l(t)$. For this reason, one can approximate the slope of the “front” as $\partial_x \sigma(x, t)|_{\text{front}} \simeq (\sigma(x_{\text{max}}, t) - \sigma(x_{\text{min}}, t)) / (x_{\text{max}}(t) - x_{\text{min}}(t)) \simeq (\sigma_+ - \sigma_0) / l(t)$. Parameters: $a = 1.5$, $\chi = 2$ and $\sigma_0 = -0.06$.

Once the simulations are initialised from the above-mentioned conditions, we can evolve the system and record the dynamics of the fields over time. The typical profiles are reported in Figure B.4(a-b), where the denaturation field ϕ and supercoiling field σ are shown for $a = 1.5$, $\chi = 2$ and $\sigma_0 = -0.06$. For these values of σ_0 and a , the uniform solution is (linearly) unstable and for this reason the initially small denatured bubble grows in time and eventually arrests at the point where two stable phases coexist. At the same time, the supercoiling field

is “emptied” from inside the bubble, where it takes the stable value of σ_- , and “poured” into the $L - l(t)$ portion of the chain, where it eventually attains the value σ_+ .

As discussed in the main text, the size of the denatured bubble $l(t)$ has been found to scale in time as

$$l(t) \sim t^{1/2} \quad \text{for topologically constrained DNA} \quad (\text{B.17})$$

$$l(t) \sim t^1 \quad \text{for topologically unconstrained DNA.} \quad (\text{B.18})$$

The exponent for unconstrained DNA, such as nicked or linear, can be simply explained by assuming that for temperatures above a critical $T_c = \epsilon/\Delta S$ the system can lower its (free) energy for each un-paired (denatured) base-pair as

$$f \simeq (\epsilon - T\Delta S)l. \quad (\text{B.19})$$

This directly implies that [145]

$$\frac{dl}{dt} \simeq \frac{1}{\gamma} \frac{df}{dl} \sim \text{const} \quad (\text{B.20})$$

and therefore $l(t) \sim t$. Note that a previous theory described in Ref. [179] found a different behaviour: this is because a lot of the analysis in [179] builds on the assumption that there is a decoupling between denaturation and untwisting, whereas the denaturation and supercoiling dynamics are always coupled in the simulation results we report here.

On the other hand, the value of α observed for topologically constrained, such as circular closed, molecules can be understood by quantifying the slowing down of the denaturation field due to the accumulation of a “front” of supercoiling on either side of the growing region. First, we note that, due to the global conservation of σ , the amount of supercoiling lost by denaturation, $|\sigma_- - \sigma_0|l(t)$, needs to be incorporated into the intact dsDNA region. By looking at our simulation, we can safely assume that this supercoiling is mostly absorbed by the “front” of positive supercoiling ¹ immediately outside of the bubble ². This

¹More precisely, $\sigma > \sigma_0$ in this region.

²In principle, some of this supercoiling may be absorbed by DNA regions far from the bubble. Provided that the fraction absorbed in the flanking region does not scale with l , our argument holds in this case as well.

front has an approximately triangular shape with constant height $h = \sigma_+ - \sigma_0$. Conservation of supercoiling therefore requires that the base increases with $l(t)$, $b \sim l(t)$ (as in this way the area of the flanking triangular fronts equals the supercoiling expelled from the denatured bubble, which is $\sim l(t)$, Fig. B.4). Now, by continuity the local supercoiling flux exiting the bubble at any point, which is $\sim |\sigma_- - \sigma_0| dl/dt$, needs to be equal to the supercoiling flux into the intact flanking region. Given the shape of the flanking profile just discussed (a triangular front with slope $\frac{\sigma_+ - \sigma_0}{b} \sim \frac{\sigma_+ - \sigma_0}{l}$), we can estimate the latter quantity as

$$J_\sigma = -\Gamma_\sigma \partial_x \sigma \simeq \Gamma_\sigma \frac{\sigma_+ - \sigma_0}{l}. \quad (\text{B.21})$$

Therefore, we obtain that

$$\frac{dl}{dt} \sim \Gamma_\sigma \frac{\sigma_+ - \sigma_0}{|\sigma_- - \sigma_0| l} \quad (\text{B.22})$$

which leads to $l(t) \sim t^{1/2}$.

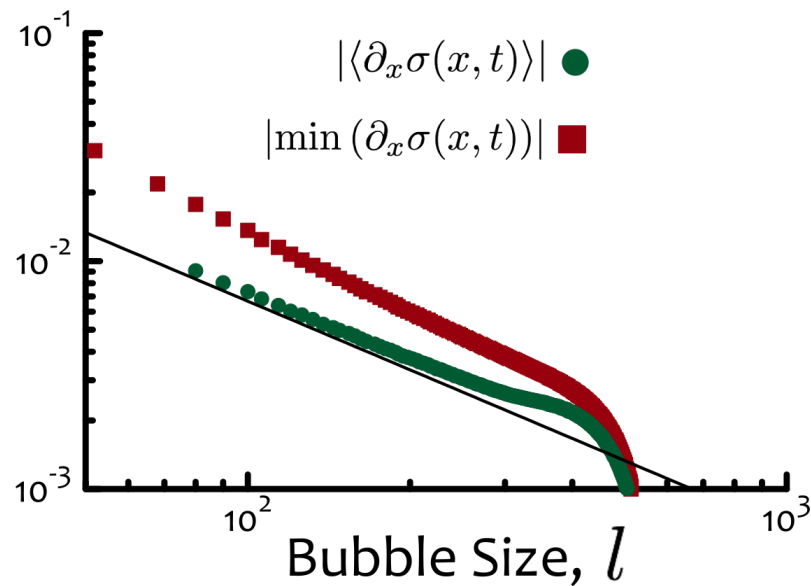


Figure B.5 Slope of the front of the σ field forming on either side of a growing bubble. Parameters as in Fig. B.4.

Equation (B.22) may also be interpreted as an effective force balance corresponding to a case of a polymer with free energy favouring denaturation

and friction dependent on (here, linearly proportional to) $l(t)$. Balances with non-constant frictions were used in recent work to rationalise scaling laws in the rotational dynamics of torsionally free molecules [180, 181]. In our case, the increase of the friction with l is due to the build-up of positive supercoiling in the regions flanking the bubbles, hence critically depends on the fact that the linking number is a topological invariant (i.e., that supercoiling is globally conserved).

Because the approximation $\partial_x \sigma(x, t)|_{\text{front}} \simeq (\sigma_+ - \sigma_0)/l(t)$ might look crude, we numerically computed the minimum and the average slope attained by the σ field on the supercoiling front. These two quantities are plotted in Fig. B.5 against the size of the bubble l . From the figure one can readily notice that the approximation is in fact not far from the real behaviour.

B.4 Computing linking number in denatured bubbles

For a topologically constrained dsDNA the linking number Lk has to be preserved at any time. In particular, even when a circular closed molecule of DNA is fully denatured, all the initial linking number has still to be preserved. One may imagine this situation as an extreme case where the only denaturation bubble takes up the whole DNA molecule. In light of this, one may ask whether a generic bubble of size l can store some non-zero linking number, Lk_d . This case is generally discarded in simple models [113, 115, 136], which consider no supercoiling within denatured regions. On the contrary, arguments from elasticity theory [118] suggest and support the idea that some linking may still be present within denatured bubbles, since a complete expulsion of this would then contribute to increase the torsional (twist) or bending (writhe) energies of the intact (non-denatured) segments through the well-known $Lk = Wr + Tw$ relation [63].

In order to test this argument we performed simulation in which we kept the strength of the hydrogen bond $\epsilon_{HB} = 6.0k_B T$ constant and high enough so denaturation could not take place. Then, for rings with different level of supercoiling ($\sigma = 0.06, 0, -0.06$, or $Lk = 106, 100, 94$, respectively), we created a bubble of size l by manually deleting the corresponding consecutive hydrogen-

bonds along the chain.

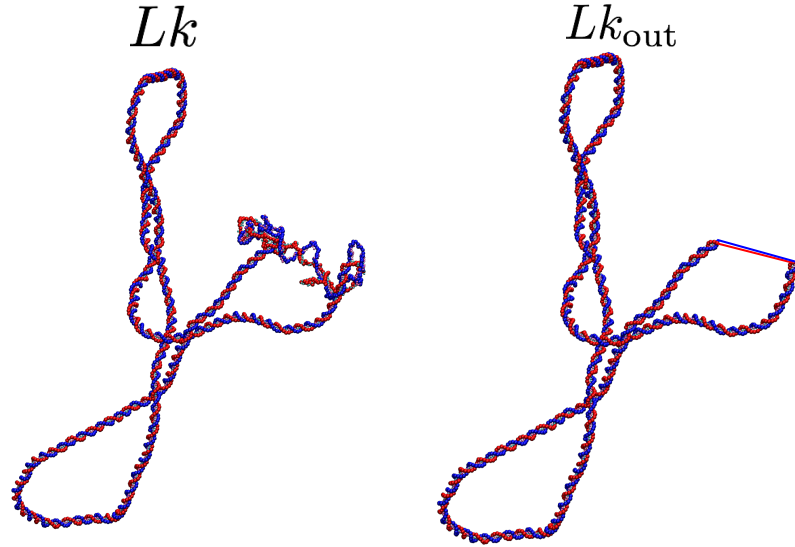


Figure B.6 *Computing linking number in denatured molecules* Sketch of the process through which we compute Lk_d . We take a denatured molecule with fixed bubble size l (here $l = 200$ bp) and replace the denatured strands with straight lines joining the first non-denatured beads on either side of the bubble. Then, we compute the linking number outside the bubble (Lk_{out}) using Eq. (B.23) and derive the linking stored within the bubble as $Lk_d = Lk - Lk_{\text{out}}$.

After equilibration we computed the linking number by numerically calculating

$$Lk(C_R, C_B) = \frac{1}{4\pi} \int_{C_R} \int_{C_B} \frac{\mathbf{r}_R - \mathbf{r}_B}{|\mathbf{r}_R - \mathbf{r}_B|^3} \cdot (d\mathbf{r}_R \times d\mathbf{r}_B), \quad (\text{B.23})$$

between curves C_R and C_B . In Eq. (B.23), \mathbf{r}_R and \mathbf{r}_B are the vectors defining the position of the beads along the curves C_R and C_B , respectively.

In order to find the linking number *inside* the denatured region, Lk_d , we first replaced the bubble by straight lines joining the first non-denatured base-pairs (see Fig. B.6), so that Eq. (B.23) returned the linking number *outside* the denatured region, Lk_{out} . Finally, we simply use the invariance of the total linking number to find $Lk_d = Lk - Lk_{\text{out}}$ (see Fig. B.7).

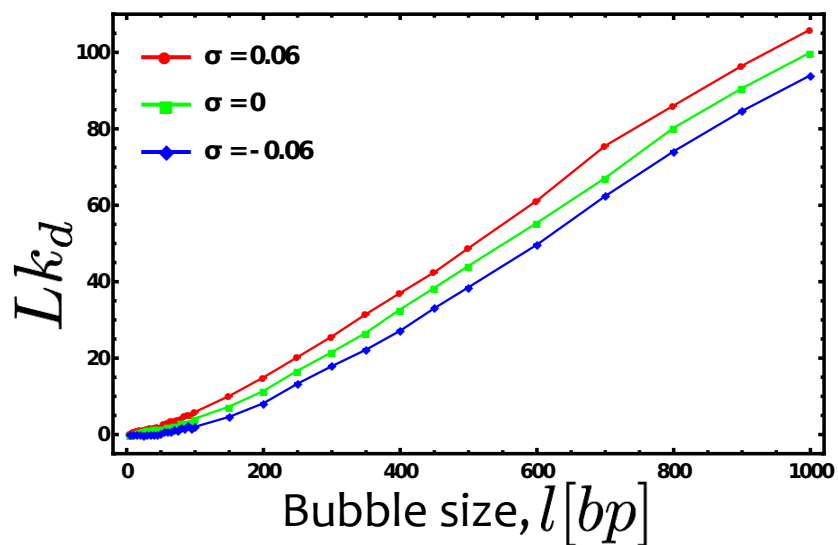


Figure B.7 Lk_d grows with the size of the denatured bubble. The figure shows that the linking number stored within the denatured region, Lk_d , grows as function of its length l . The curves (red, green and blue) correspond to circular closed molecules with different levels of supercoiling: $\sigma = 0.06, 0, -0.06$, or $Lk = 106, 100, 94$, respectively.

C

Details on the transcription of model polymerase

C.1 Polymerase model

The polymerase is modelled as a rigid body with two parts (see Fig. C.1(a)): a ring (shown in black) and a segment across its diameter (shown in yellow) which is referred to as the "crossbar" in the main text. The crossbar is placed in between two consecutive base-pairs in the initial configuration (Fig. C.1(b)) with its axis perpendicular to the DNA center line. It is responsible for breaking the hydrogen bonds and it is made by seven small beads. While the six beads colour in yellow have an excluded volume of 0.34 nm to repel and break the HB of the dsDNA; the bead in the centre of the finger (colour in green) can also apply a force (\mathbf{F}_p) on the nearest base-pairs (in a range of 0.5 nm) and directed perpendicular to the crossbar. Since in our current model we have fixed the polymerase in the 3D space, the force applied by the polymerase imposes the movement of DNA through it; along the clockwise direction when $\mathbf{F}_p > 0$ and counter-clockwise direction in the opposite case. Finally, the ring is made by ten beads (1 nm in diameter each) surrounding the crossbar and preventing the polymerase complex to be expelled from the DNA molecule (Fig. C.1(c)) during the whole simulation.

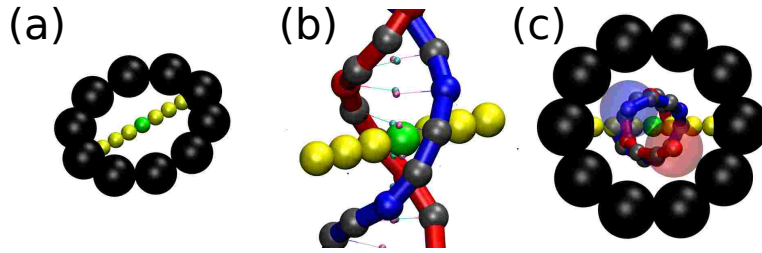


Figure C.1 (a) Representation of a polymerase in the model. (b) The polymerase is initialized with the crossbar in between two base-pairs, to ease the visualization we do not show the black beads. (c) The ring surrounding both, the crossbar and the DNA, is designed to ensure that the polymerase remains in the middle of the two strands during the whole simulation.

C.2 Moving polymerase

In the main text we describe simulations where the position and orientation of the polII complex is fixed. We can also perform simulations with a mobile polymerase. In this case, the generation and dynamics of supercoiling will also depend on the relative rotational motion between the DNA and polymerase [159]. If, for example, the polymerase crossbar rotates 360° (with the same handedness of the DNA) as it covers a distance of 10 bp along the DNA axis, no supercoiling will be generated, even in the presence of a topological barrier. Therefore, twin supercoiling domain formation requires the rotation rate of the polymerase to be small, meaning that the rotational drag it experiences must be sufficiently large. *In vivo* this is likely to be the case: a transcribing polymerase is associated with a large number of co-factors, there is a nascent RNA connected to the polymerase, and the whole complex may be connected to distant DNA regions or other cellular components.

In our simulations each rigid body (each DNA bead-patch pair or the polII model) experiences both a translational drag γ (see Eq. (3.2) above) and a rotational drag γ_r . This drag can be related to an inertial time scale $\tau_{\text{in}} = m/\gamma$, the time over which the velocity decorrelates (directional information is lost). Here m is the total mass of the rigid body. Similarly for the rotational drag $\tau_{\text{in rot}} = I/\gamma_r$, where I is the moment of inertia. For a sphere Stokes' law relates the drag to the viscosity of the solvent ν : $\gamma = 3\pi\nu\sigma$ and $\gamma_r = \pi\nu\sigma^3$ where σ is the

diameter of the sphere. The moment of inertia for a solid sphere is $I = m\sigma^2/10$. So if we take the simple approximation that the polymerase rotates like a solid sphere, then in simulations where the polymerase can move we should set

$$\tau_{\text{in rot}} = \frac{3}{10}\tau_{\text{in}}. \quad (\text{C.1})$$

In this case, the rotational drag is small enough such that the polymerase rotates as it moves and no supercoiling is generated. If we instead increase the rotational drag (decrease the rotational damp) to account for the possible effects listed above, then the polymerase does indeed generate a twin supercoiling domain (Fig. C.2 - here we set $\tau_{\text{in}} = 1$ and $\tau_{\text{in rot}} = 0.1$ (smaller than $0.3\tau_{\text{in}}$) both given in simulation units).

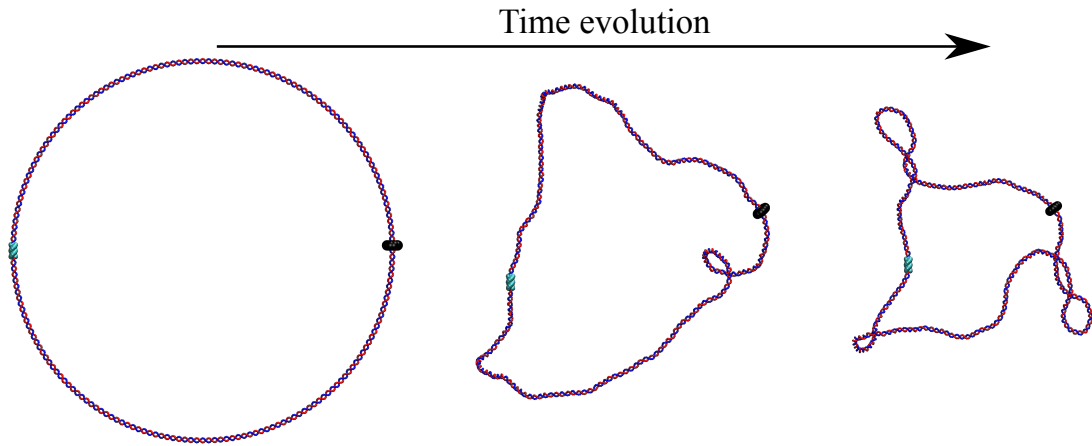


Figure C.2 Temporal evolution of the system during transcription, when the model polymerase is allowed to move, and has a rotational drag ten times larger than the translational drag.

C.3 Algorithms to compute the topological quantities

Local and total twist of DNA

To obtain the local twist angle $\theta(n)$ between basepairs n and $n + 1$ we followed the procedure described in Ref. [182]. We consider the ring DNA molecule as

a discrete chain consisting of a set of N vertices, with their locations given by $\mathbf{r}(n)$ for $n = 1, \dots, N$. In our model, $\mathbf{r}(n)$ corresponds to the mid-point of the line connecting the centres of the two beads which form the n th bp. Then we define a local reference frame at each vertex. This is specified by the unitary tangent vector $\hat{\mathbf{t}}(n) = (\mathbf{r}(n+1) - \mathbf{r}(n))/|\mathbf{r}(n+1) - \mathbf{r}(n)|$, one unitary normal vector $\hat{\mathbf{f}}(n)$ obtained from the projection of the vector connecting two beads in a base-pair onto the plane perpendicular to $\hat{\mathbf{t}}(n)$, and a third vector $\hat{\mathbf{m}}(n) = \hat{\mathbf{t}}(n) \times \hat{\mathbf{f}}(n)$. Note that since the system forms a ring, the last tangent vector $\hat{\mathbf{t}}(N)$, is the one joining the last base-pair with the first one. We define the binormal vector $\hat{\mathbf{b}}(n) = \hat{\mathbf{t}}(n-1) \times \hat{\mathbf{t}}(n)/|\hat{\mathbf{t}}(n-1) \times \hat{\mathbf{t}}(n)|$ and the turning angle $\alpha(n) = \arccos(\hat{\mathbf{t}}(n-1) \cdot \hat{\mathbf{t}}(n))$ between the incoming and outgoing tangent vectors. For the case $n = 1$, the vector $\hat{\mathbf{t}}(n-1)$ refers to the last tangent vector. With this information we can apply to the vectors $\hat{\mathbf{m}}(n-1)$, a rotation around $\hat{\mathbf{b}}(n)$ an angle $\alpha(n)$ to find the material shifted vector $\hat{\mathbf{m}}^*(n)$. Finally, the local twist angle can be computed with the use of the following relation.

$$\sin \theta(n) = \hat{\mathbf{t}}(n) \cdot [\hat{\mathbf{m}}^*(n) \times \hat{\mathbf{m}}(n)]. \quad (\text{C.2})$$

The total twist Tw is found by adding the value of the local twist along all the base-pairs

$$\text{Tw} = \frac{1}{2\pi} \sum_n \theta(n). \quad (\text{C.3})$$

Writhe, unsigned-writhe and linking number of DNA

The writhe of a closed curve C is defined through the Gauss double integral:

$$\text{Wr} = \frac{1}{4\pi} \int_C \int_C \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|} d\mathbf{r}_2 \times d\mathbf{r}_1, \quad (\text{C.4})$$

where \mathbf{r}_1 and \mathbf{r}_2 are any two points along the curve. In the discrete case, the curve C is represented by the set of vectors joining consecutive base-pairs (\mathbf{t}_n as defined in the previous section) and the double integral becomes a double sum. The numerical evaluation of this integral can be computed following the methods presented in Ref. [164], where they find that Eq.(C.4) along a curve with

N segments can be expressed as:

$$\text{Wr} = 2 \sum_{i=2}^N \sum_{j<i} \frac{\Omega_{ij}}{4\pi}, \quad (\text{C.5})$$

where $\Omega_{ij}/4\pi$ is the Gaussian integral along the segment i, j . These are denoted by $\mathbf{t}_{i,i+1}$ and $\mathbf{t}_{j,j+1}$, which represent two different vectors linking any two consecutive vertices ($i-i+1$ and $j-j+1$) along the DNA axis. In this case, the absolute value of the Gauss integral times 4π represents all the perspectives in which the two previous vectors seem to cross in the solid angle Ω^* . This quantity can be found through the definition of the unitary vectors:

$$\hat{\mathbf{n}}_i = \frac{\mathbf{t}_{i,j} \times \mathbf{t}_{i,j+1}}{|\mathbf{t}_{i,j} \times \mathbf{t}_{i,j+1}|}, \quad (\text{C.6})$$

$$\hat{\mathbf{n}}_{i+1} = \frac{\mathbf{t}_{i,j+1} \times \mathbf{t}_{i+1,j+1}}{|\mathbf{t}_{i,j+1} \times \mathbf{t}_{i+1,j+1}|}, \quad (\text{C.7})$$

$$\hat{\mathbf{n}}_j = \frac{\mathbf{t}_{i+1,j+1} \times \mathbf{t}_{i+1,j}}{|\mathbf{t}_{i+1,j+1} \times \mathbf{t}_{i+1,j}|}, \quad (\text{C.8})$$

$$\hat{\mathbf{n}}_{j+1} = \frac{\mathbf{t}_{i+1,j} \times \mathbf{t}_{i,j}}{|\mathbf{t}_{i+1,j} \times \mathbf{t}_{i,j}|}. \quad (\text{C.9})$$

These vectors satisfy

$$\begin{aligned} \Omega_{ij}^* = & \arcsin(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_{i+1}) + \arcsin(\hat{\mathbf{n}}_{i+1} \cdot \hat{\mathbf{n}}_j) + \arcsin(\hat{\mathbf{n}}_j \cdot \hat{\mathbf{n}}_{j+1}) + \\ & \arcsin(\hat{\mathbf{n}}_{j+1} \cdot \hat{\mathbf{n}}_i). \end{aligned} \quad (\text{C.10})$$

The contribution to the Gauss integral depends on the crossing orientation between two segments and therefore

$$\frac{\Omega_{ij}}{4\pi} = \frac{\Omega_{ij}^*}{4\pi} \text{sign}[(\mathbf{t}_{j,j+1} \times \mathbf{t}_{i,i+1}) \cdot \mathbf{t}_{i,j}], \quad (\text{C.11})$$

The total writhe is obtained by replacing this result into Eq.(C.5). It is worth mentioning here that the unsigned writhe ($\bar{\zeta}$) which we refer to in the main text, is computed by neglecting the sign function in the previous equation before adding

all the contributions of the segments i and j . Finally, once the values of both, writhe and twist have been found, the linking number is computed by adding these quantities, $Lk = Tw + Wr$.

C.4 Length of the first plectoneme and distance to the polymerase

In order to calculate the length (l') of a plectoneme, and its distance (d) from the polymerase, we examine the contacts made between non-adjacent bp – where a contact is defined as when the separation between two bp is less than a threshold distance $r_{threshold}$. A contact map is a visual representation of all the contacts within the dsDNA molecule, where a matrix is obtained by filling the mn element with 1 if two bp (m and n) are in contact, and 0 otherwise. Self-contacts are not allowed and therefore the diagonal of the matrix is full of zeros. We set $r_{threshold} = 3\sigma = 3 \text{ nm}$; since this value is larger than the rise of the B-DNA (0.34 nm), we can expect that close to the diagonal of the matrix the elements will be one.

As in appendix C.3, the middle point between two complementary beads represents the position \mathbf{r}_n of the n th base-pair in our model. In Fig. C.3(a) we show schematically this position for a small section of the ringDNA molecule, lying between base-pair i and base-pair j . If this piece of DNA forms a plectoneme like the one shown in Fig. C.3(b), that would mean that the contour length (l') of the plectoneme is $|i - j|$ (in base-pairs). Furthermore, the ends of the plectoneme correspond to base-pairs i and j respectively, and given $r_{threshold} = 3\sigma$, we can infer that they are close enough to define a contact. The same would be true for base-pairs $i + 1$ and $j - 1$, $i + 2$ and $j - 2$, and so on and so forth, until we reach the tip of the plectoneme located approximately at the base-pair with index $(i + j)/2$. In the contact map, this situation is displayed in the form of a line (full of ones) perpendicular to the diagonal of the matrix.

In figures C.3(c)-(d), we show the contact map after the index of the base-pairs is shifted so the polymerase always is located in the middle of the molecule (at position $N/2 = 500$). c) represents a typical case of the first plectoneme appearing in our simulations with negative writhe (behind the polymerase) and

d) represents a typical case with positive writhe (in front of the polymerase). Therefore, if we generate a contact map every δt timesteps, we can look for the first one which contains information of a plectoneme according to the criteria mentioned above, and then extract the length of this plectoneme ($l' = |i - j|$). For the same timestep, we can also compute the distance between the polymerase and the ends of the plectoneme: $|N/2 - i|$ and $|N/2 - j|$. The minimum of these two quantities correspond to the distance from the polymerase to the beginning of the first plectoneme.

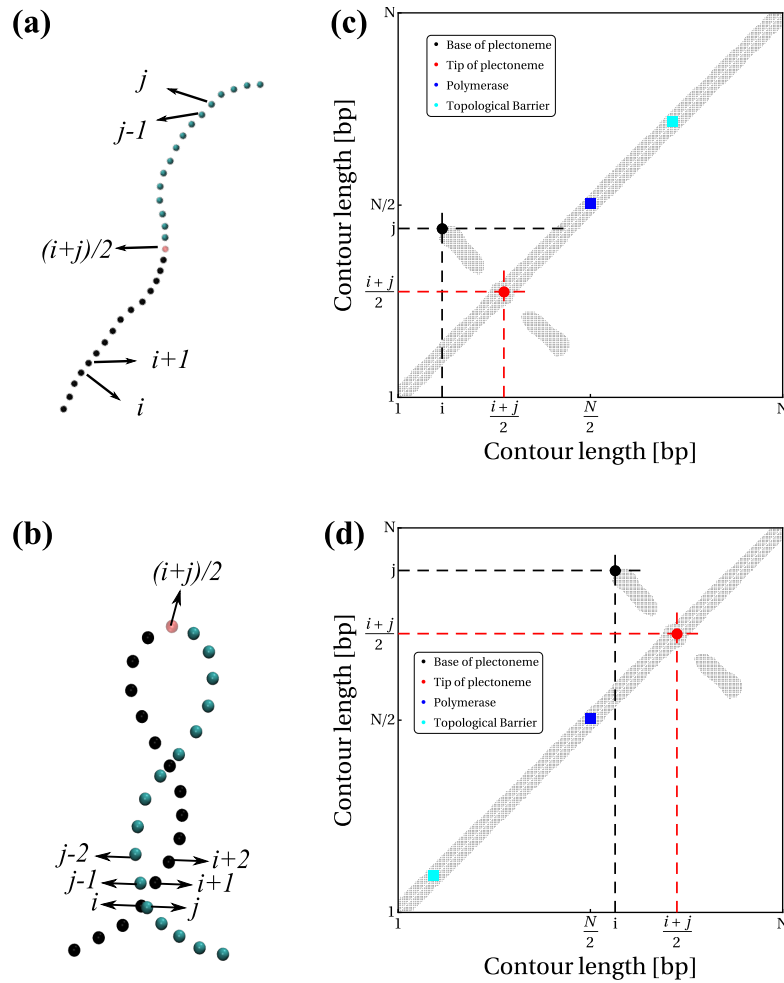


Figure C.3 (a) Segment of the ring DNA molecule containing from base-pair i to j . (b) After the polymerase has applied a force on the DNA, the segment forms a plectoneme of length $l' = |i - j|$, the tip of the plectoneme is located at base-pair $(i + j)/2$. (c) Sketch of a contact map for the first plectoneme with negative writhe. (d) Sketch of a contact map for the first plectoneme with positive writhe. In both, (c) and (d), the base-pairs positions were shifted so the polymerase is located at $N/2$

C.5 Polymerase-DNA relative velocity

In Fig. C.4(a) we show the position of the polymerase (to the closest base-pair) x_p as a function of time for simulations with the setup as in Fig. 5.1a(ii), for different forces F_p . The slope of these curves represents the velocity of the polymerase (v_p), in units of bp/ τ_{LJ} . Figure C.4(b) shows this velocity (averaged over 10 simulation) as a function of F_p (plotted in units of ϵ/σ). In the low force regime ($F_p < 30 \epsilon/\sigma$), v_p is linear in the applied force. By extrapolating this result and taking into account that $\tau_{LJ} = 2.228\text{ns}$, we found that a force $F_p \simeq 1 - 5 \epsilon/\sigma$ would lead to a rate of transcription similar to the one found in experiments (10-50 bp/s). However, this would require to perform impractical long simulations, so the polymerase breaks enough base-pairs before producing the twin supercoiling domain.

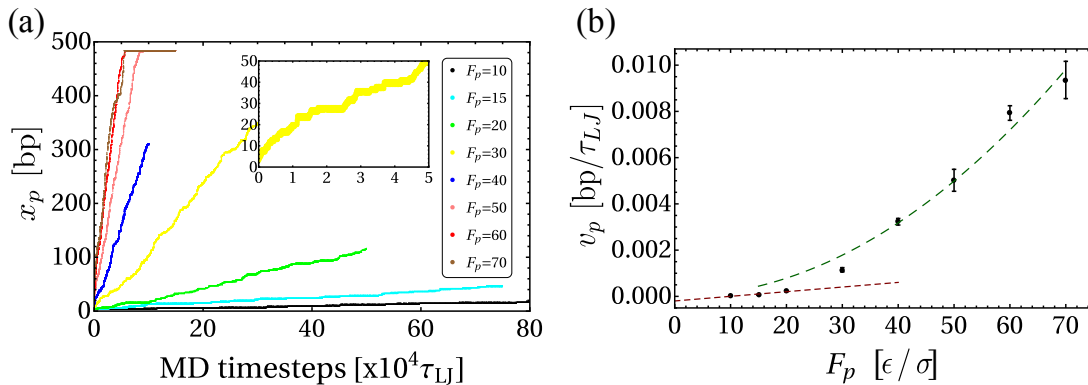


Figure C.4 (a) Temporal evolution of the position (x_p , in base-pairs) of the polymerase along the dsDNA molecule. Different colors represent the results obtained for different forces F_p . In the inset of this figure we show the results at early times for $F_p = 30 \epsilon/\sigma$, since the slope of this curve fluctuates, we fit a line over large time-intervals to obtain the velocity. (b) For a fixed force, we repeat the procedure to compute the slope of the curves in the previous plot, for 10 different replicas. The average of these results represents the velocity (v_p) of the polymerase as a function of the applied force.

Intriguingly, in the high force regime ($F_p > 30 \epsilon/\sigma$) the velocity is quadratic with the force. For the case of a single bead subjected to a force we would expect it to accelerate until that force is cancelled by the frictional drag, achieving a velocity $v_p = F_p/\gamma$. In our simulations, each DNA bead within a distance of 0.5 nm from the centre of the polIII ring experiences a force F_p ; however, each bead

is also subjected to forces and torques due to the DNA interactions, and forces due to the excluded volume interaction with the polymerase crossbar and ring. For sufficiently large force the DNA is pulled so quickly into the polymerase it does not have time to rearrange its configuration, and in front of the polIII it is under significant tension and gets pulled taut. Also, at different forces the DNA beads and the polymerase beads may overlap to different extents, meaning there is an effective friction between the two. Together this means that the total drag force on the DNA may not be linear in the velocity, so the polIII velocity scales with the force in an unpredictable way.

C.6 Asymmetry in the coarse-grained model due to steric repulsion

As discussed in section 2.3.1, the chirality of the DNA causes that the molecule responds different to overtwisting and undertwisting. In particular, when no extreme structural changes of DNA are allowed (like denaturation or the Pauling-DNA form), in reference [82] the authors find that if the bending-coupling is considered, then twisting a ringDNA molecule will change its persistence length, making underwounds molecules harder to twist and easier to bend than overwound molecules (see section 2.3.2).

To corroborate this, we ran simulations of 1000 bp long topological-constrained dsDNA molecules (closed into a loop) with different levels of supercoiling: $\sigma = 0.06$ and -0.06 , respectively. Then, the total dihedral energy of the system is computed as a function of time (shown in Fig C.5(a)). We can see that in general, the dihedral energy for the negative supercoiling case (blue) is larger than the positive (red) one. Since this energy is related to the torque response of the DNA in the simulations, this means that is more difficult to keep undertwisting a negative supercoiled ring than overtwist a positive supercoiled one (when initially both have the same magnitude of supercoiling). In Fig C.5(b) we show the temporal evolution of the steric energy in the system. In this case the configuration with the lowest steric energy is the one with $\sigma = -0.06$. This is in agreement with the fact that undertwist molecules are more open (if the deficit of turns is enough, the geometry of the two strands is closer to a

ladder structure) and therefore there are less steric interactions. Therefore, we would expect that underwound molecules carry a lower conformational stress than overwound molecules.

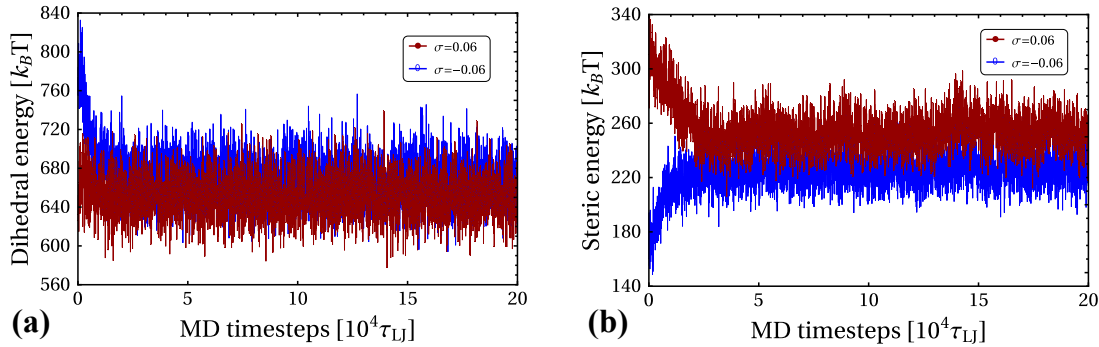


Figure C.5 Temporal evolution of the total dihedral energy (a) and the steric energy (b) computed from simulations for two ring molecules with $\sigma = 0.06$ and $\sigma = -0.06$.

C.7 Dependence of nucleosome wrapping on supercoiling

As described in the main text, we model the nucleosome as a big bead (10nm in diameter) with 20 patches on it, following a left-handed helical path for 1.7 turns. The patches have a short range attractive interaction on the DNA beads. In addition, the separation between these patches is set to cover a distance of 150 bp, as in real nucleosomes. We observed in our simulations that a dsDNA molecule 1000 bp long, initialized with $Lk = Lk_0$ (relaxed state), can be completely attached to the nucleosome without writhing at any other location (apart from the binding sites of the nucleosome). This is depicted in Fig. C.6(a). The DNA base-pair indices are shown in black, and the index of the nucleosome patches in red. After the nucleosome is completely attached to the DNA, we add a topological barrier in between base-pair 230 and base-pair 240 (shown in cyan in Fig. C.6(a))

Here the kymographs capture the state of each nucleosome patch (attached or not to the DNA) against time. In Fig. C.6(b), the kymograph is displayed for the case in which the TB isolates the nucleosome from the effect of the

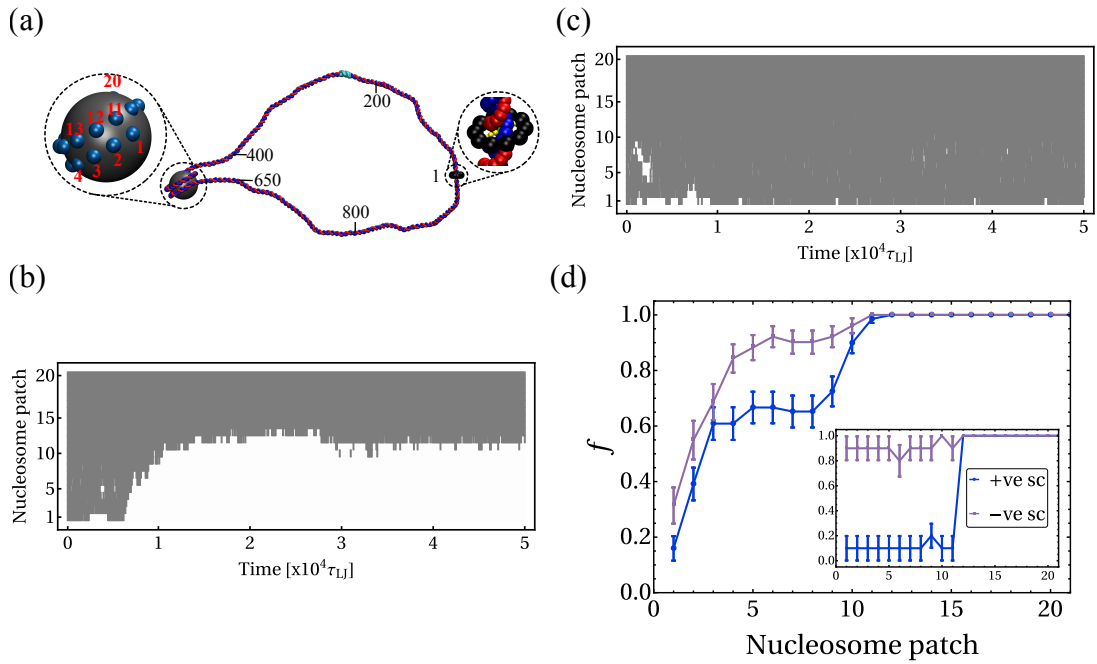


Figure C.6 (a) Initial configuration of the system with the DNA wrapped around the nucleosome. (b)-(c) Kymographs obtained from simulations when the +ve and -ve supercoiling reaches the nucleosome, respectively. At time $t = 0$ the nucleosome is completely attached to DNA, and the polIII is activated as the state (gray for nucleosome patches attached to DNA and white otherwise) is recorded. (d) Fraction (f) of configurations, at a fixed time $t = 1 \times 10^4 \tau_{LJ}$, for which each of the nucleosome patches was in contact with the DNA. In the inset we show the same variables at a late time $t = 5 \times 10^4 \tau_{LJ}$. Lines connecting dots are shown to ease visualization.

negative supercoiling. This means that at $t = 0$, the polymerase is activated so it moves on the clockwise direction along the DNA in Fig. C.6(a). This produces the unwrapping of the nucleosome, starting from the patch closer to the positive supercoiling "wave" (i.e. that DNA detaches from nucleosome, starting from patch 1 labelled in red). The kymograph for the same system when the polymerase moves in the opposite direction is shown in Fig. C.6(c). This time the TB prevents the +ve supercoiling from reaching the nucleosome and the -ve supercoiling do not produce the unwrapping of the nucleosome.

In agreement with these results and the ones shown in the main text (Fig. 5.3), at a fixed time-step we compute the fraction (f) of times (over configurations) that each of the nucleosome patches was in contact with DNA. This is depicted at an early time ($t = 1 \times 10^4 \tau_{LJ}$) in Fig. C.6(d) and for a late time ($t = 5 \times 10^4 \tau_{LJ}$) in the inset of the same figure. Both, positive (+ve) and negative (-ve)

supercoiling, produce a perturbation of the nucleosome patches at early times. This perturbation causes the unbinding of the nucleosome in the case of the +ve supercoiling reaching the nucleosome, but not in the case of -ve supercoiling (as seen from the data at late times in Fig C.6(d)). Finally, the nucleosome patches closer to the TB do not show this perturbation, probing the effectiveness of the TB.

C.8 Twist diffusion coefficient

If the contour length of the dsDNA ring is expressed through the continuum variable $x \in [0, N]$, with N the number of base-pairs along the molecule. Then, the diffusion equation of the local twist $\theta(x, t)$ at time t and position x is given by:

$$\frac{\partial \theta(x, t)}{\partial t} = D \frac{\partial^2 \theta(x, t)}{\partial x^2}, \quad (\text{C.12})$$

where D is the twist diffusion coefficient. Assuming that in the initial configuration a certain region along the chain, located at $x_i < x < x_f$, has a local twist θ , different from the one of B-DNA in its natural form ($\theta_0 = 36^\circ$). Then, $\Delta\theta = \theta - \theta_0$ represents the excess (or deficit) of twist in this region of the chain, and the initial condition can be expressed as:

$$\theta(x, 0) = \theta_0 + \Delta\theta H(x - x_i) H(x_f - x). \quad (\text{C.13})$$

where H is the so called Heaviside step function. The solution of Eq.(C.12) with the initial condition Eq.(C.13) is

$$\theta(x, t) = \begin{cases} \frac{\theta_0}{2} \left[2 + \operatorname{erf}\left(\frac{x-x_i}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{x-x_f}{2\sqrt{Dt}}\right) \right], & \text{for } \Delta\theta > 0 \\ \frac{\theta_0}{2} \left[2 + \operatorname{erf}\left(\frac{x-x_f}{2\sqrt{Dt}}\right) - \operatorname{erf}\left(\frac{x-x_i}{2\sqrt{Dt}}\right) \right], & \text{for } \Delta\theta < 0, \end{cases} \quad (\text{C.14})$$

where erf represents the error function. Finally, the data obtained at any timestep t , can be fitted with Eq.(C.14) to obtain the Twist diffusion coefficient.

C.9 Writhe field

In order to study the relaxation of plectonemes in our model, we ran simulations starting from a configuration of a DNA molecule which has developed the twin-supercoiling domain by the action of a polymerase (see Fig. 5.1a(iii), for example). Then we remove both, the constriction on the topological barrier and the polII complex, and we analyse the evolution of this new system to its relaxed state (Fig. 5.1a(iv)). Under these conditions, the two opposite sign plectonemes will eventually disappear. One possible way of studying this event is by defining an effective local writhe field $\chi(n, t)$ at position n along the chain and time t , that can be tracked directly from our BD simulations.

The idea is that at fixed time t , we define a sliding window centre at base-pair n . The length of this windows should be larger or equal than the length of the biggest plectoneme in our initial configuration. It is then possible to create an artificially closed curve that follows first the open path generated by the DNA axis contained inside the window, and then is extended by adding extra beads along the tangent direction at both ends of this path. After reaching a certain distance, the curve is closed by defining a vector joining the two new terminal beads. Then we compute the writhe contained in this artificial curve by solving the Gaussian integral (Eq.(C.4)). Once this is done, we move the centre of the window one base-pair and repeat the procedure until we have covered the whole contour length of the chain. With this we obtain χ as a function of n for a fixed time. Therefore, the temporal evolution of the field χ is computed by doing the same calculation at different time-steps. We repeated the procedure mentioned above for four different cases (case 1 is shown in Fig. 5.4(b) and cases 2-4 correspond to (a)-(c) in Fig. C.7), whose main differences are the number of plectonemes in the initial configuration and the magnitude of local writhe stored in these plectonemes.

The writhe field of these systems is displayed in Fig. C.7(a)-(c). We show how the amplitude of the local writhe in the plectonemes decreases as the system evolves towards the relaxed state. In the initial configuration, the maximum of $\chi(n, 0)$ corresponds to the writhe stored in the plectoneme located in front of the polymerase. Similarly, the minimum corresponds to the writhe stored in the plectoneme behind the polII complex. The position of the two extremes in this

plot coincide with the indices of the base-pairs at the tip of the plectonemes.

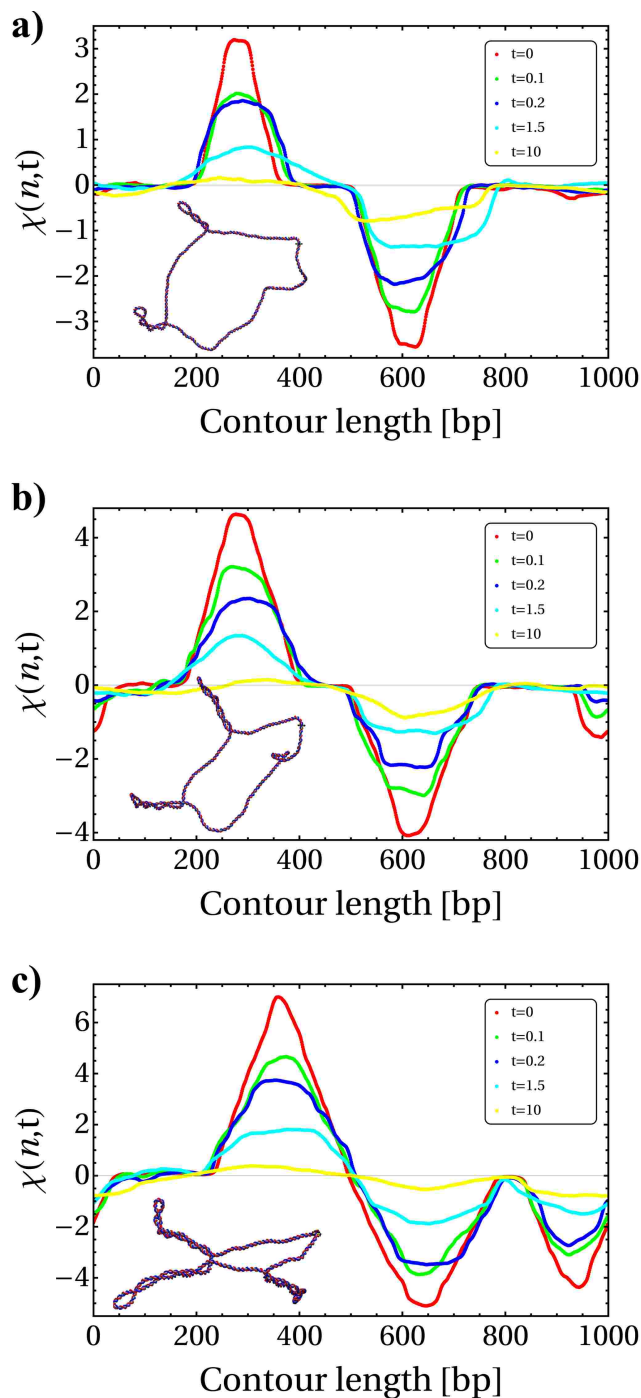


Figure C.7 Relaxation of the writhe field, $\chi(n,t)$, extracted from simulations. The different plots correspond to different initial configurations (shown in the insets of each plot). We will refer to these configurations as: case 1 (shown in Fig. 5.4(b)), cases 2-4 correspond to a-c in this panel.

Extremes of $\chi(n, t)$ during relaxation

In Fig. C.8 we show the temporal evolution of the maximum of $\chi(n, t)$ for the cases mentioned in the previous section. We observe that the relaxation of the local writhe in all of them involves two timescales: a rapid relaxation at the beginning with the magnitude of the writhe decreasing as e^{-t/τ_1^+} , and then the process slows down, decreasing as e^{-t/τ_2^+} (with τ_1^+ and τ_2^+ , two positives satisfying $\tau_2^+ > \tau_1^+$).

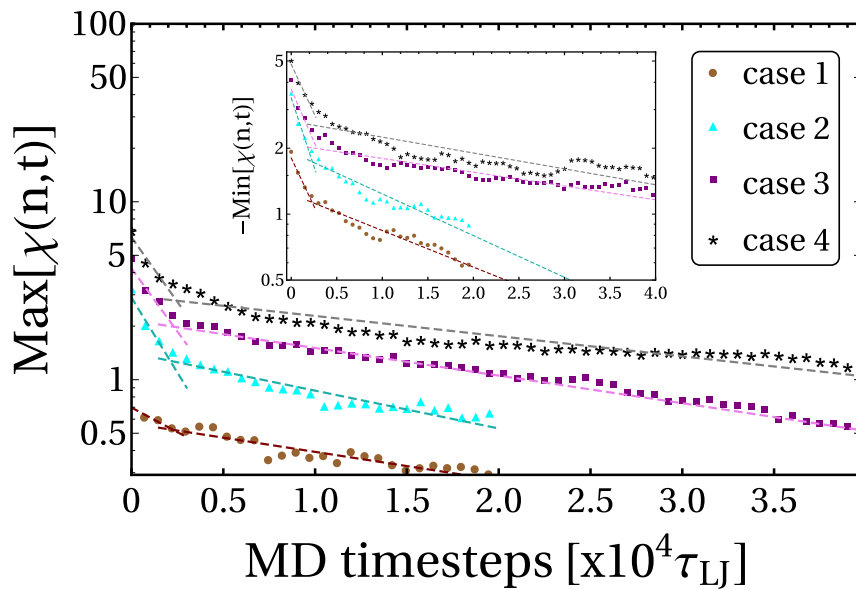


Figure C.8 Log-linear plot of the maximum of $\chi(n, t)$ as a function of time. The different colours represent the evolution from the different initial configurations mentioned in the previous section. These configurations differ in the amplitude of the writhe and the number of plectonemes at $t = 0$. In the inset the same information is depicted for the magnitude of the minimum of $\chi(n, t)$. Points represent data from the simulations, and the lines the fit to exponential decay. The decay constants are summarized in Table C.1.

The evolution of the magnitude of the minimum of $\chi(n, t)$ is shown on the inset of Fig. C.8. It exhibits the same tendency as before, with a relaxation split by two timescales τ_1^- and τ_2^- . The value of all the decay constants is summarized in Table C.1, where $\tau_1^+ < \tau_1^-$ (except in case 1, where the plectoneme with positive writhe is so short that the first time-scale is negligible) and $\tau_2^+ < \tau_2^-$. This implies that the relaxation of the positive writhe is faster than the negative

one, this might explain why we observed that the formation of the first negative plectoneme occurs earlier than the first positive plectoneme in our simulations.

Case	Maximum $[\chi(n, t)]$		-Minimum $[\chi(n, t)]$	
	τ_1^+ [τ_{LJ}]	τ_2^+ [τ_{LJ}]	τ_1^- [τ_{LJ}]	τ_2^- [τ_{LJ}]
1	— — —	27213±623	4915±207	26036±416
2	2580±120	20443±373	3469±79	22811±405
3	3066±119	27976±284	4531±206	68989±959
4	2991±131	38252±333	4857±161	59976±488
average	2879±71	28471±212	4443±86	44453±306

Table C.1 Decay constants of the maximum (τ_1^+ and τ_2^+) and the minimum (τ_1^- and τ_2^-) of the local writhe $\chi(n, t)$. Their average is shown in the last row, ignoring the result of τ_1^+ for the case 1.

C.10 Bending energy during relaxation

In the previous section we showed that the relaxation of both, positive and negative writhe, occurs in two time-scales. One explanation to this behaviour is that the tension accumulated in the relative short plectonemes produces the fast relaxation at the beginning. When a large part of this tension has been released by decreasing the local writhe in the plectoneme, the relaxation slows down accounting for the second timescale. In order to test this hypothesis, here we compute the evolution of the bending energy (E_{bend}) of the DNA molecule during the same relaxation stage (iii to iv in Fig. 5.1(a)).

From section 2.2.2 we know that a curve $\mathbf{r}(s)$, parametrized by arc-length (s) according to the mapping: $s \in [0, L] \rightarrow \mathbf{r}(s)$, with L the total length of the curve; exhibits a bending energy:

$$\frac{E_{\text{bend}}}{k_B T} = \frac{l_p}{2} \int_0^L \kappa^2(s) ds, \quad (\text{C.15})$$

where l_p is the persistence length, $\kappa(s) = \frac{d\mathbf{t}}{ds} = \frac{d^2\mathbf{r}}{ds^2}$ is the curvature at position s along the curve and \mathbf{t} its tangent vector. For the discrete case of a DNA molecule from our simulations, the bending energy is obtained from the curve at fixed

time which follows the path of the tangent vectors $\mathbf{t}(n)$. We only have to use the discrete version of Eq.(C.15):

$$\frac{E_{\text{bend}}}{k_B T} = \frac{l_p}{2L} \sum_{i=0}^N |\mathbf{t}_{(n+1)} - \mathbf{t}_{(n)}|^2, \quad (\text{C.16})$$

with $l_p = 50$ nm and $L = 0.34N$ nm for DNA. At different times the shape of the molecule will change and the temporal evolution of the bending energy can be tracked. After average over five different configurations we obtain the curves shown in Fig. C.9, each one represents the evolution from different initial configurations (case 1-4 in the previous section). Clearly, the configurations with larger plectonemes at $t = 0$, start with the highest bending energy. We see that the value of E_{bend} decays quickly at the beginning of the relaxation, in a similar way as τ_1^+ (or τ_1^-), and then stabilizes. This last part would correspond to the slow relaxation time τ_2^+ (or τ_2^-)

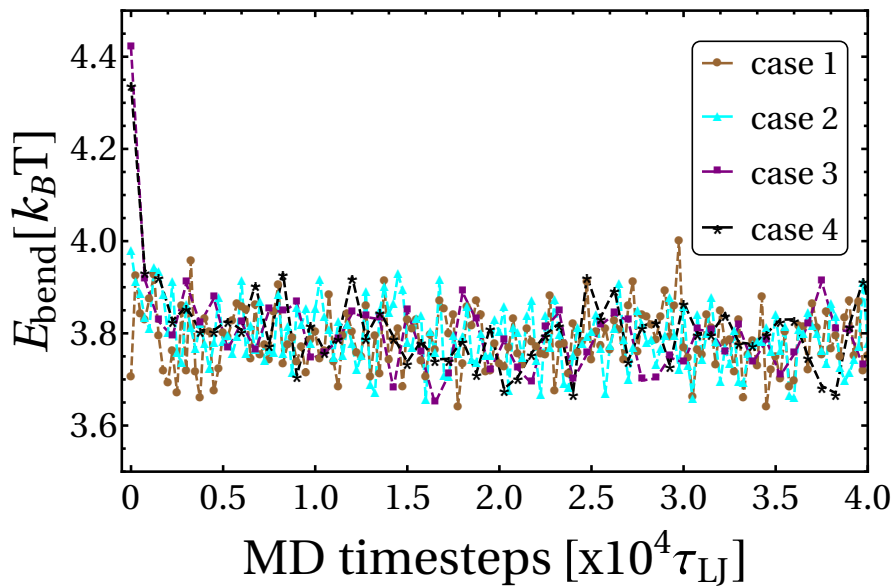


Figure C.9 Bending energy plot as a function of time. The different colours represent the evolution from the different initial configurations mentioned in the previous section. These configurations differ in the amplitude of the writhe and the number of plectonemes at $t = 0$. Lines connecting dots are shown to ease visualization.

Bibliography

- [1] Y A G Fosado, D. Michieletto, J. Allan, C. A. Brackley, O. Henrich, and D. Marenduzzo. A single nucleotide resolution model for large-scale simulations of double stranded DNA. *Soft Matter*, 12:9458–9470, 2016.
- [2] Y A G Fosado, D Michieletto, and D Marenduzzo. Dynamical Scaling and Phase Coexistence in Topologically Constrained DNA Melting. *Phys. Rev. Lett.*, 119:118002, Sep 2017.
- [3] Oliver Henrich, Yair Augusto Gutiérrez Fosado, Tine Curk, and Thomas E. Ouldridge. Coarse-grained simulation of DNA using LAMMPS. *The European Physical Journal E*, 41(5):57, May 2018.
- [4] C R Calladine, H R Drew, B F Luisi, and A A Travers. *Understanding DNA: The molecule and how it works*. Elsevier Academic Press, third edition.
- [5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, fifth edition.
- [6] J D Watson, T A Baker, S P Bell, A Gann, M Levine, R Losick, and S C Harrison. *Molecular Biology of the Gene*. Pearson Education, seventh edition.
- [7] Michail Monastyrsky. *Topology in Molecular Biology*. Biological and Medical Physics, Biomedical Engineering. Springer Berlin Heidelberg, 2007.
- [8] Joseph E. Deweese, Michael A. Osheroff, and Neil Osheroff. DNA topology and topoisomerases: Teaching a “Knotty” Subject. *Biochemistry and Molecular Biology Education*, 37(1):2–10, 2009.
- [9] A. Griswold. Genome packaging in prokaryotes: the circular chromosome of *E. coli*. *Nature Education*, 1(57), 1 2008.
- [10] Andrew D. Bates and Anthony. Maxwell. *DNA topology*. Oxford University Press, 1993.

-
- [11] T. Christian Boles, James H. White, and Nicholas R. Cozzarelli. Structure of plectonemically supercoiled dna. *Journal of Molecular Biology*, 213(4):931 – 951, 1990.
- [12] Nick Gilbert and James Allan. Supercoiling in DNA and chromatin. *Current Opinion in Genetics & Development*, 25(1):15–21, 2014.
- [13] A Annunziato. DNA packaging: Nucleosomes and chromatin. *Nature Education*, 1(26), 01 2008.
- [14] Haojie Yu and Peter Dröge. Replication-induced supercoiling: a neglected DNA transaction regulator? *Trends in Biochemical Sciences*, 39(5):219–220, 2014.
- [15] Eldra Pearl Solomon. *Biology*. Cengage Learning, tenth edition, 2015.
- [16] J. D Watson and F. H. C Crick. Genetical Implications of the Structure of Deoxyribonucleic Acid. *JAMA*, 269(15):1967–1969, 1993.
- [17] Jie Ma and Michelle Wang. DNA supercoiling during transcription. *Biophysical Reviews*, 8(Supplement 1):75–87, 2016.
- [18] *Molecular cell biology*. Palgrave Macmillan, Basingstoke, sixth edition, 2007.
- [19] M.T.J. van Loenhout, M.V. de Grunt, and C. Dekker. Dynamics of DNA supercoils. *Science*, 338(6103), 2012.
- [20] He Meng, Johan Bosman, Thijn Van der heijden, and John Van Noort. Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains. *Biophysical Journal*, 106(5):1174–1181, 2014.
- [21] Antonella Vettone, Giuseppe Perugino, Mosè Rossi, Anna Valenti, and Maria Ciaramella. Genome stability: recent insights in the topoisomerase reverse gyrase and thermophilic DNA alkyltransferase. *Extremophiles*, 18(5):895–904, 2014.
- [22] Pavel Lulchev and Dagmar Klostermeier. Reverse gyrase—recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic acids research*, 42(13), 2014.
- [23] Giacomo Cavalli and Tom Misteli. Functional implications of genome topology. *Nat. Struct. Mol. Biol.*, 20(3):290–9, March 2013.
- [24] C A Brackley, Stephen Taylor, Argyris Papantonis, Peter R Cook, and Davide Marenduzzo. Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA*, 110(38):E3605–11, September 2013.

- [25] Peter R Cook and Davide Marenduzzo. Entropic organization of interphase chromosomes. *J. Cell. Biol.*, 186(6):825–34, September 2009.
- [26] Peter R Cook. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.*, 32(november):347–352, 2002.
- [27] Amir Bar, Alkan Kabakçoğlu, and David Mukamel. Denaturation of circular DNA: Supercoil mechanism. *Phys. Rev. E*, 84:041935, Oct 2011.
- [28] R Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, UK, 1976.
- [29] Madeleine Price Ball. https://en.wikipedia.org/wiki/File:DNA_chemical_structure.svg.
- [30] Zephyris. https://en.wikipedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png.
- [31] Rosalind E. Franklin and Raymond G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171, 1953.
- [32] R Franklind and Gosling. Churchill College Archives, Cambridge; hitherto unpublished, described qualitatively in *Acta Crystallographica*. 6:673, 1953.
- [33] Vladimir N. Potaman and Richard R. Sinden. *DNA: Alternative Conformations and Biology*. Madame Curie Bioscience Database [Internet]., Austin (TX): Landes Bioscience, 2000-2013.
- [34] Aaron Klug. The Discovery of the DNA Double Helix. *Journal of Molecular Biology*, 335(1):3 – 26, 2004.
- [35] Alexander Rich and Shuguang Zhang. Z-DNA: the long road to biological function. *Nature Reviews Genetics*, 4:566–572, Jul 2003.
- [36] Donna R. Whelan, Thomas J. Hiscox, Julian I. Rood, Keith R. Bambery, Don McNaughton, and Bayden R. Wood. Detection of an en masse and reversible B- to A-DNA conformational transition in prokaryotes in response to desiccation. *J R Soc Interface*, 11(97), Aug 2014.
- [37] Jan Lipfert, Gary M. Skinner, Johannes M. Keegstra, Toivo Hensgens, Tessa Jager, David Dulin, Mariana Köber, Zhongbo Yu, Serge P. Donkers, Fang-Chieh Chou, Rhiju Das, and Nynke H. Dekker. Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *Proceedings of the National Academy of Sciences*, 111(43), 2014.
- [38] Xiang-Jun Lu and Wilma K Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research*, 31(17), 2003.

- [39] J T O Kirk. Determination of the base composition of deoxyribonucleic acid by measurement of the adenine-guanine ratio. *Biochem. J.*, 105(2):673–7, 1967.
- [40] Dieter W Gruenwedel and Chi-hsia Hsu. Salt Effects on the Denaturation of Dna. *Biopolymers*, 7:557–570, 1969.
- [41] Alexander Vologodskii. *Biophysics of DNA*. Cambridge University Press, 2015. Cambridge Books Online.
- [42] Jerome Vinograd, Jacob Lebowitz, and Robert Watson. Early and late helix-coil transitions in closed circular DNA the number of superhelical turns in polyoma DNA. *J. Mol. Biol.*, 33(1):173–197, apr 1968.
- [43] Carlos Bustamante, Steven R. B. Smith, Jan Liphardt, and Donald Smith. Single-molecule studies of DNA mechanics. *Current opinion in structural biology*, 10 3:279–85, 2000.
- [44] Doi M. and Edwards S.F. *The theory of polymer dynamics*.
- [45] Thomas T. Perkins, Douglas E. Smith, Ronald G. Larson, and Steven Chu. Stretching of a Single Tethered Polymer in a Uniform Flow. *Science*, 268(5207):83–87, 1995.
- [46] S.B. Smith, Laura Finzi, and Carlos Bustamante. Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads. 258:1122–6, 12 1992.
- [47] JD Moroz and Philip Nelson. Entropic elasticity of twist-storing polymers. *Macromolecules*, 9297(41):6333–6347, 1998.
- [48] David R. Latulippe and Andrew L. Zydney. Radius of gyration of plasmid DNA isoforms from static light scattering. *Biotechnology and Bioengineering*, 107(1):134–142, 2010.
- [49] Gordon C.K. Roberts. *Encyclopedia of biophysics*. 1 edition, 2013.
- [50] John F. Marko and Erick D. Siggia. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.
- [51] Siggia ED Smith S Bustamante C, Marko JF. Entropic elasticity of λ -phage DNA. *Science*, 265:1599–600, 1994.
- [52] Odijk. Stiff chains and filaments under tension. *Macromolecules*, 28:7016–7018, 1995.
- [53] T. Odijk. Physics of tightly curved semiflexible polymer chains. *Macromolecules*, 26,(25), (25), 1993.

- [54] Theo Odijk. Theory of lyotropic polymer liquid crystals. *Macromolecules*, 19(9):2313–2329, 1986.
- [55] Robert Landick Jeff Gelles Michelle D. Wang, Hong Yin and Steven M. Block. Stretching DNA with optical tweezers. *Biophysical journal*, 72:1335–46, 1997.
- [56] Ioulia Rouzina Jay R Wenner, Mark C Williams and Victor A Bloomfield. Salt dependence of the elasticity and overstretching transition of single DNA molecules. *Biophysical journal*, 82:3160–69, 2002.
- [57] Philip Charles Nelson. *Biological physics : energy, information, life*. W.H. Freeman, New York ; Basingstoke, first edition, 2008.
- [58] Stefanos K. Nomidis, Franziska Kriegel, Willem Vanderlinden, Jan Lipfert, and Enrico Carlon. Twist-Bend Coupling and the Torsional Response of Double-Stranded DNA. *Phys. Rev. Lett.*, 118:217801, May 2017.
- [59] Jan Lipfert, Jacob W. J. Kerssemakers, Tessa Jager, and Nynke H. Dekker. Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nature Methods*, 7:977 – 980, Oct 2010.
- [60] Jan Lipfert, Matthew Wiggin, Jacob W. J. Kerssemakers, Francesco Pedaci, and Nynke H. Dekker. Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids. *Nature Communications*, 2, Aug 2011.
- [61] C. A. Brackley, A. N. Morozov, and D. Marenduzzo. Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS. *J. Chem. Phys.*, 140(13):135103, 2014.
- [62] Xander J. A. Janssen, Jan Lipfert, Tessa Jager, Renier Daudey, Jaap Beekman, and Nynke H. Dekker. Electromagnetic Torque Tweezers: A Versatile Approach for Measurement of Single-Molecule Twist and Torque. *Nano Letters*, 12(7):3634–3639, 2012. PMID: 22642488.
- [63] F. B. Fuller. Decomposition of the linking number of a closed ribbon: A problem from molecular biology. *Proc. Natl. Acad. Sci. USA*, 75(8):3557–3561, August 1978.
- [64] Catherine Naughton, Nicolaos Avlonitis, Samuel Corless, James G Prendergast, Ioulia K Mati, Paul P Eijk, Scott L Cockroft, Mark Bradley, Bauke Ylstra, and Nick Gilbert. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.*, 20(3):387–395, 2013.
- [65] James C. Wang. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Reviews Molecular Cell Biology*, 3(6), 2002.

- [66] J J Champoux. DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry*, 70, 2001.
- [67] K J Mariani, J E Ikeda, S Schlagman, and J Hurwitz. Role of DNA gyrase in phiX replicative-form replication in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 74(5), 1977.
- [68] Paolo Bettotti, Valeria Visone, Lorenzo Lunelli, Giuseppe Perugini, Maria Ciaramella, and Anna Valenti. Structure and Properties of DNA Molecules Over The Full Range of Biologically Relevant Supercoiling States. *Scientific reports*, 8(1), 2018.
- [69] Purificación López-García. DNA Supercoiling and Temperature Adaptation: A Clue to Early Diversification of Life? *Journal of Molecular Evolution*, 49(4):439–452, 1999.
- [70] Jan Lipfert, Mina Lee, Orkide Ordu, Jacob W. J Kerssemakers, and Nynke H Dekker. Magnetic Tweezers for the Measurement of Twist and Torque. *Journal of Visualized Experiments*, (87), 2014.
- [71] Christian Matek, TE Ouldridge, J P K Doye, and A A Louis. *Scientific Reports*, 5, 2015.
- [72] Benedict E. K. Snodin, Ferdinando Randisi, Majid Mosayebi, Petr Šulc, John S. Schreck, Flavio Romano, Thomas E. Ouldridge, Roman Tsukanov, Eyal Nir, Ard A. Louis, and Jonathan P. K. Doye. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *The Journal of Chemical Physics*, 142(23):234901, 2015.
- [73] Stefanos K. Nomidis, Enrico Skoruppa, Enrico Carlon, and John F Marko. Twist-bend coupling and the statistical mechanics of DNA: perturbation theory and beyond. *arXiv:1809.07050*, Sep 2018.
- [74] Zev Bryant, Michael D. Stone, Jeff Gore, Steven B. Smith, Nicholas R. Cozzarelli, and Carlos Bustamante. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424(6946), 2003.
- [75] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), 1998.
- [76] J. F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko. Structural Transitions of a Twisted and Stretched DNA Molecule. *Physical Review Letters*, 83(5):1066–1069, 1999.

- [77] Jonathan M. Fogg, Daniel J. Catanese, Graham L. Randall, Michelle C. Swick, and Lynn Zechiedrich. Differences Between Positively and Negatively Supercoiled DNA that Topoisomerases May Distinguish. In *Mathematics of DNA Structure, Function and Interactions*, pages 73–121, New York, NY, 2009. Springer New York.
- [78] T.R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette. Behavior of Supercoiled DNA. *Biophysical Journal*, 74(4):2016–2028, 1998.
- [79] Franziska Kriegel, Niklas Ermann, Ruaridh Forbes, David Dulin, Nynke H Dekker, and Jan Lipfert. Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers. 45.
- [80] Sarah A Harris, Charles A Laughton, and Tanniemola B Liverpool. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic acids research*, 36(1), 2008.
- [81] Jonathan M Fogg, Natalia Kolmakova, Ian Rees, Sergei Magonov, Helen Hansma, John J Perona, and E Lynn Zechiedrich. Exploring writhe in supercoiled minicircle dna. *Journal of Physics: Condensed Matter*, 18(14):S145–S159, 2006.
- [82] J. F. Marko and E. D. Siggia. Bending and twisting elasticity of DNA. *Macromolecules*, 27(4):981–988, 1994.
- [83] Maksymilian Chruszcz, Dominika Borek, Marcin Domagalski, Zbyszek Otwinowski, and Wladek Minor. X-Ray Diffraction Experiment: The Last Experiment in the Structure Elucidation Process. *Advances in Protein Chemistry and Structural Biology*, 77(C):23–40, 2009.
- [84] H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. The Protein Data Bank. *Nucleic acids research*, 28(1), 2000.
- [85] M Humberto Reyes-Valdés, Amalio Santacruz-Varela, Octavio Martinez, June Simpson, Corina Hayano-Kanashiro, and Celso Cort es Romero. Analysis and optimization of bulk DNA sampling with binary scoring for germplasm characterization. *PLOS ONE*, 8(11).
- [86] Ignacio Tinoco and Ruben L Gonzalez. Biological mechanisms, one molecule at a time. *Genes & Development*, 25(30):1205–1231, 2011.
- [87] Terence R. Strick, Jean-François Allemand, David Bensimon, Richard Lavery, and Vincent Croquette. Phase coexistence in a single DNA molecule. *Physica A*, 263(1-4):392–404, 1999.

- [88] C Bustamante, Z Bryant, and SB Smith. Ten years of tension: single-molecule DNA mechanics. *Nature*, 421:423–7, 2003.
- [89] Zev Bryant, Michael D Stone, Jeff Gore, Steven B Smith, Nicholas R Cozzarelli, and Carlos Bustamante. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424(6946):338–341, 2003.
- [90] F Ritort. Single-molecule experiments in biological physics: methods and applications. *J. Phys.: Condens. Matter*, 18(32):R531–R583, 2006.
- [91] Juan A. Torreno-Pina, Bruno M. Castro, Carlo Manzo, Sonja I. Buschow, Alessandra Cambi, and Maria F. Garcia-Parajo. Enhanced receptor–clathrin interactions induced by N-glycan–mediated membrane micropatterning. *Proceedings of the National Academy of Sciences*, 111(30):11037–11042, 2014.
- [92] Furio Ercolessi. *A Molecular Dynamics Primer*. 1997.
- [93] Thomas E Cheatham III. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Current Opinion in Structural Biology*, 14(3):360 – 367, 2004.
- [94] Eva Fadrná, Nad’ a Špačková, Joanna Sarzynska, Jaroslav Koča, Modesto Orozco, III Thomas E. Cheatham, Tadeusz Kulinski, and Jiří Šponer. Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *Journal of Chemical Theory and Computation*, 5(9):2514–2530, 2009. PMID: 26616629.
- [95] Modesto Orozco, Agnes Noy, and Alberto Pérez. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Current Opinion in Structural Biology*, 18(2):185 – 193, 2008.
- [96] Ilda D’Annessa, Andrea Coletta, Thana Sutthibutpong, Jonathan Mitchell, Giovanni Chillemi, Sarah Harris, and Alessandro Desideri. Simulations of DNA topoisomerase 1B bound to supercoiled DNA reveal changes in the flexibility pattern of the enzyme and a secondary protein-DNA binding site. *Nucleic Acids Res.*, 42(14):9304–12, August 2014.
- [97] Davide Michieletto, Davide Marenduzzo, and Ajazul H Wani. Chromosome-wide simulations uncover folding pathway and 3D organization of interphase chromosomes. *arXiv:1604.03041*, pages 1–20, 2016.
- [98] A Rosa and Ralf Everaers. Structure and dynamics of interphase chromosomes. *PLOS Comp. Biol.*, 4(8):1, January 2008.
- [99] Davide Michieletto, Davide Marenduzzo, and Enzo Orlandini. Topological Patterns in two-dimensional gel electrophoresis of dna knots. *Proc. Natl. Acad. Sci. USA*, 2015.

- [100] Daniel M Hinckley, Gordon S Freeman, Jonathan K Whitmer, and Juan J de Pablo. An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, 139(14):144903, October 2013.
- [101] S. K. Nomidis, W. Vanderlinden, J. Lipfert, and E. Carlon. The effect of twist-bend coupling on the torsional properties of double-stranded DNA. *arXiv:1603.00835*, pages 1–12, 2016.
- [102] Thomas E Ouldridge, Ard A Louis, and Jonathan P K Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, 134(8):085101, February 2011.
- [103] Juan J de Pablo. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annu. Rev. Phys. Chem.*, 62:555–74, January 2011.
- [104] Steve Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117(1):1 – 19, 1995.
- [105] SB Smith, Y Cui, and C Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 5:795–799, 1996.
- [106] Yue Ding, Carlo Manzo, Geraldine Fulcrand, Fenfei Leng, David Dunlap, and Laura Finzi. DNA supercoiling: A regulatory signal for the λ repressor. *Proc. Natl. Acad. Sci. USA*, 111(43):15402–15407, October 2014.
- [107] C. A. Brackley, J. Johnson, A. Bentivoglio, S. Corless, N. Gilbert, G. Gonnella, and D. Marenduzzo. Stochastic Model of Supercoiling-Dependent Transcription. *Phys. Rev. Lett.*, 117:018101, Jun 2016.
- [108] E Orlandini, M C Tesi, and S G Whittington. Polymer entanglement in melts. *J. Phys. A: Math. Gen.*, 33:181–186, 2000.
- [109] Davide Michieletto, Davide Marenduzzo, Enzo Orlandini, Gareth P. Alexander, and Matthew S. Turner. Threading dynamics of ring polymers in a gel. *ACS Macro Lett.*, 3(3):255–259, 2014.
- [110] C Bouchiat, Michelle D Wang, J F. Allemand, T Strick, Steven M Block, and V Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.*, 76(1 Pt 1):409–413, 1999.
- [111] Laszlo Oroszi, Peter Galajda, Huba Kirei, Sindor Bottka, and Pol Ormos. Direct measurement of torque in an optical trap and its application to double-strand DNA. *Phys. Rev. Lett.*, 97(5):1–4, 2006.

- [112] Peter Botchan, J C Wang, and H. Echols. Effect of circularity and superhelicity on transcription from bacteriophage lambda DNA. *Proc. Natl. Acad. Sci. USA*, 70(11):3077–3081, 1973.
- [113] A. Kabakçioğlu, Enzo Orlandini, and David Mukamel. Supercoil formation in DNA denaturation. *Phys. Rev. E*, 80(1):1–4, 2009.
- [114] Giuseppe Lia, David Bensimon, Vincent Croquette, Jean-Francois Allemand, David Dunlap, Dale E A Lewis, Sankar Adhya, and Laura Finzi. Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc. Natl. Acad. Sci. USA*, 100(20):11373–7, 2003.
- [115] Jae-Hyung Jeon, Jozef Adamcik, Giovanni Dietler, and Ralf Metzler. Supercoiling Induces Denaturation Bubbles in Circular DNA. *Phys. Rev. Lett.*, 105(20):208101, nov 2010.
- [116] Christophe Lavelle. Pack, unpack, bend, twist, pull, push: The physical side of gene expression. *Curr. Opin. Genet. Dev.*, 25(1):74–84, 2014.
- [117] John J. Kozak and Craig J. Benham. Denaturation: an example of a catastrophe. *Proc. Natl. Acad. Sci. USA*, 71(5):1977–1981, 1974.
- [118] C J Benham. Torsional stress and local denaturation in supercoiled DNA. *Proc. Natl. Acad. Sci. USA*, 76(8):3870–3874, 1979.
- [119] G Wesley Hatfield and Craig J Benham. DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.*, 36:175–203, 2002.
- [120] Enrico Carlon, Enzo Orlandini, and Attilio Stella. Roles of Stiffness and Excluded Volume in DNA Denaturation. *Phys. Rev. Lett.*, 88(19):198101, apr 2002.
- [121] Steven P. Mielke, Niels Gronbech-Jensen, V. V. Krishnan, William H. Fink, and Craig J. Benham. Brownian dynamics simulations of sequence-dependent duplex denaturation in dynamically superhelical DNA. *J. Chem. Phys.*, 123(12):124911, 2005.
- [122] Francois Sicard, Nicolas Destainville, and Manoel Manghi. DNA denaturation bubbles: Free-energy landscape and nucleation/closure rates. *J. Chem. Phys.*, 142(3), 2015.
- [123] Huiquan Wang and Craig J. Benham. Superhelical destabilization in regulatory regions of stress response genes. *PLoS Comp. Biol.*, 4(1):0062–0076, 2008.
- [124] JL Wood. pH-controlled hydrogen-bonding. *Biochem J.*, 143:775–777, 1974.

- [125] W Beers, A Cerami, and E Reich. An experimental model for internal denaturation of linear DNA molecules. *Proc. Natl. Acad. Sci. USA*, 58:1624–1631, 1967.
- [126] Carl Schildkraut and Shneior Lifson. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, 3(2):195–208, 1965.
- [127] VV Rybenkov, NR Cozzarelli, and AV Vologodskii. Probability of DNA knotting and the effective diameter of the DNA double helix. *Proceedings of the National Academy of Sciences of the United States of America*, 90(11):5307–5311, 1993.
- [128] Ngo Minh Toan and Cristian Micheletti. Inferring the effective thickness of polyelectrolytes from stretching measurements at various ionic strengths: applications to DNA and RNA. *Journal of Physics: Condensed Matter*, 18(14):S269, 2006.
- [129] Jacob Lebowitz. Through the looking glass: the discovery of supercoiled dna. *Trends in Biochemical Sciences*, 15(5):202–207, 1990.
- [130] L.V. Crawford. The physical characteristics of polyoma virus: Iv. the size of the dna. *Virology*, 22(1):149–152, 1964.
- [131] R Dulbecco and M Vogt. Evidence for a ring structure of polyoma virus dna. *Proceedings of the National Academy of Sciences of the United States of America*, 50, 1963.
- [132] R Weil and J Vinograd. The cyclic helix and cyclic coil forms of polyoma viral dna. *Proceedings of the National Academy of Sciences of the United States of America*, 50, 1963.
- [133] Walter Fiers and Robert L. Sinsheimer. The structure of the DNA of bacteriophage ϕ X174: III. Ultracentrifugal evidence for a ring structure. *Journal of Molecular Biology*, 5(4):424–434, 1962.
- [134] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis. The Twisted Circular Form of Polyoma Viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 53(5):1104–1111, 1965.
- [135] RM Wartell and AS Benight. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.*, 126(2):67–107, 1985.
- [136] D Poland and H A Scheraga. Occurrence of a phase transition in nucleic acid models. *J. Chem. Phys.*, 45(5):1464–1469, 1966.
- [137] Yariv Kafri, David Mukamel, and Luca Peliti. Why is the DNA denaturation transition first order? *Phys. Rev. Lett.*, 85(23):4988–4991, 2000.

- [138] Enrico Carlon, Enzo Orlandini, and Attilio L. Stella. Roles of stiffness and excluded volume in dna denaturation. *Phys. Rev. Lett.*, 88:198101, Apr 2002.
- [139] A V Gagua, B N Belintsev, and L Lyubchenko Yu. Effect of base-pair stability on the melting of superhelical DNA. *Nature*, 294(5842):662–663, 1981.
- [140] Joseph Rudnick and Robijn Bruinsma. Effects of torsional strain on thermal denaturation of DNA. *Phys. Rev. E*, 65(3):2–4, 2002.
- [141] Srikanta Sen and Rabi Majumdar. Statistical mechanical theory of melting transition in supercoiled DNA. *Biopolymers*, 27(9):1479–1489, 1988.
- [142] WR Bauer and CJ Benham. The free energy, enthalpy and entropy of native and of partially denatured closed circular DNA. *J. Mol. Biol.*, 234:1184—1196, 1993.
- [143] Christian Matek, Thomas E. Ouldridge, Jonathan P. K. Doye, and Ard a. Louis. Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA. *Sci. Rep.*, page 5, 2014.
- [144] Gregg Jaeger. The Ehrenfest Classification of Phase Transitions: Introduction and Evolution. *Archive for History of Exact Sciences*, 53(1):51–81, 1998.
- [145] P.M. Chaikin and T.C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, 2000.
- [146] Stephen J. Blundell, S.J. Blundell, Katherine M. Blundell, and K.M. Blundell. *Concepts in Thermal Physics*, volume 9780199562091. Oxford University Press, 2 edition, 2009.
- [147] Debashish Chowdhury. *Principles of equilibrium statistical mechanics*. Wiley-VCH, Weinheim ; New York, first edition, 2000.
- [148] V Víglaský, M Antalík, J Adamčík, and D Podhradský. Early melting of supercoiled DNA topoisomers observed by TGGE. *Nucleic acids research*, 28(11):E51, 2000.
- [149] William F. Smith. *Foundations of materials science and engineering*. McGraw-Hill international editions;Engineering mechanics series. McGraw-Hill, Singapore, second edition, 1993.
- [150] Akihiko Matsuyama, R M L Evans, and M E Cates. Non-uniformities in polymer / liquid crystal mixtures I . Interfacial tension. *Eur. Phys. J. E*, 87:79–87, 2002.

- [151] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49(3):435–479, 1977.
- [152] Julien Kockelkoren and Hugues Chaté. Late stages of coarsening in model C. *Physica D: Nonlinear Phenomena*, 168(C):80–92, 2002.
- [153] L Postow, N J Crisona, B J Peter, C D Hardy, and N R Cozzarelli. Topological challenges to DNA replication: conformations at the fork. *Proc. Natl. Acad. Sci. USA*, 98(15):8219–26, 2001.
- [154] Michelle D. Wang, Mark J. Schnitzer, Hong Yin, Robert Landick, Jeff Gelles, and Steven M. Block. Force and Velocity Measured for Single Molecules of RNA Polymerase. *Science*, 282(5390):902–907, 1998.
- [155] Elio A. Abbondanzieri, William J. Greenleaf, Joshua W. Shaevitz, Robert Landick, and Steven M. Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438:460–465, 2005.
- [156] Qian Wang and B. Montgomery Pettitt. Modeling DNA thermodynamics under torsional stress. *Biophys. J.*, 106(5):1182–1193, 2014.
- [157] L F Liu and J C Wang. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A*, 84(20):7024–7027, 1987.
- [158] Yeou-Ping Tsao, Hai-Young Wu, and Leroy F. Liu. Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell*, 56(1):111 – 118, 1989.
- [159] Stuart A. Sevier and Herbert Levine. Mechanical Properties of Transcription. *Phys. Rev. Lett.*, 118(26):1–5, 2017.
- [160] A. Bentivoglio, M. Ancona, C. A. Brackley, G. Gonnella, and D. Marenduzzo. Non-equilibrium phase transition in a model for supercoiling-dependent DNA transcription. *Soft Matter*, 14:3632–3639, 2018.
- [161] AD Bates and A Maxwell. *DNA topology*. Oxford University Press, 2005.
- [162] Chris A Brackley, Jill M Brown, Dominic Waithe, Christian Babbs, James Davies, Jim R Hughes, Veronica J Buckle, and Davide Marenduzzo. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biology*, 17(1), 2016.
- [163] D. Michieletto. On the tree-like structure of rings in dense solutions. *Soft Matter*, 12:9485–9500, 2016.
- [164] Konstantin Klenin and Jörg Langowski. Computation of writhe in modeling of supercoiled DNA. *Biopolymers*, 54(5):307–317, 2000.

- [165] Cizhong Jiang and B. Franklin Pugh. Nucleosome positioning and gene regulation: advances through genomics .(Report). *Nature Reviews Genetics*, 10(3), 2009.
- [166] John Svaren and Wolfram Hörz. Regulation of gene expression by nucleosomes. *Current Opinion in Genetics & Development*, 6(2):164–170, 1996.
- [167] Stephen C. Harrison. A structural taxonomy of DNA-binding domains. *Nature*, 353(6346), 1991.
- [168] Colin W. Garvie and Cynthia Wolberger. Recognition of Specific DNA Sequences. *Molecular Cell*, 8(5):937–946, 2001.
- [169] David Jakubec, Roman A Laskowski, and Jiri Vondrasek. Sequence-Specific Recognition of DNA by Proteins: Binding Motifs Discovered Using a Novel Statistical/Computational Analysis. *PLoS ONE*, 11(7), 2016.
- [170] N M Luscombe, R A Laskowski, and J M Thornton. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic acids research*, 29(13), 2001.
- [171] W. F. Anderson, D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature*, 290(5809), 1981.
- [172] Carl O. Pabo and Mitchell Lewis. The operator-binding domain of λ repressor: structure and DNA recognition. *Nature*, 298(5873), 1982.
- [173] G M Church, J L Sussman, and S H Kim. Secondary structural complementarity between DNA and proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 74(4), 1977.
- [174] Anne Lebrun, Zippora Shakked, and Richard Lavery. Local DNA stretching mimics the distortion caused by the TATA box–binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), 1997.
- [175] Tamara R Litwin, Maria Solá, Ian J Holt, and Keir C Neuman. A robust assay to measure DNA topology-dependent protein binding affinity. *Nucleic acids research*, 43(7), 2015.
- [176] Thomas Splettstoesser. https://commons.wikimedia.org/wiki/File:Zinc_finger_DNA_complex.png.
- [177] Chris A. Brackley, Benno Liebchen, Davide Michieletto, Francois Mouvet, Peter R. Cook, and Davide Marenduzzo. Ephemeral Protein Binding to DNA Shapes Stable Nuclear Bodies and Chromatin Domains. *Biophysical Journal*, 112(6):1085–1093, 2017.

-
- [178] Peter R Cook and T Misteli. *Principles of nuclear structure and function*. Wiley New York, 2001.
- [179] D. M. Crothers and H. C. Spatz. Theory of friction-limited DNA unwinding. *Biopolymers*, 10(10):1949–1972, 1971.
- [180] J. C. Walter, M. Baiesi, G. T. Barkema, and E. Carlon. Unwinding Relaxation Dynamics of Polymers. *Phys. Rev. Lett.*, 110:068301, 2013.
- [181] J. C. Walter, M. Baiesi, E. Carlon, and H. Schiessel. Unwinding Dynamics of a Helically Wrapped Polymer. *Macromolecules*, 47:4840–4846, 2014.
- [182] Nicolas Clauvelin and Irwin Olson, Wilma K. and Tobias. Characterization of the Geometry and Topology of DNA Pictured As a Discrete Collection of Atoms. *J. Chem. Theory Comput.*, 8(3):1092–1107, 2015.