Imperial College London

Chemistry Department

Ph.D. dissertation

Chiral theory of DNA supercoiling

Author:
Ruggero Cortini

Supervisor:
Prof. Alexei Kornyshev

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy of The Imperial College of Science, Medicine and Technology.

November 2012

Abstract

DNA supercoiling is a fundamental biological process occurring in all cells. We developed a theory of braiding (supercoiling) of a pair of DNA molecules that takes into account the contribution of the bending and the electrostatic energy. The electrostatic interaction was calculated within the framework of the Kornyshev–Leikin theory of DNA interactions, which takes into account realistic helical patterns of charge. Because of the chirality of the charge patterns, we predict that left-handed braiding of a pair of DNA molecules is more favourable than right-handed braiding. Applying our model to the case of closed loop DNA supercoiling and to single molecule DNA micromanipulations, we predict novel effects that have not yet been experimentally observed. We show that supercoiling may occur in topologically relaxed plasmids, as a consequence of attractive chiral forces. We speculate about the potential biological role of the predicted effects in the case of topoisomerase action, and the occurrence of positively supercoiled DNA in hyperthermophilic bacteria and archea. Our findings also suggest alternative an explanation of well-known experiments that proved that divalent ions overwind DNA. We also give an explanation for pairing of homologous DNA molecules in monovalent salt, and explain the occurrence of tight supercoiling observed in cryo-electron and atomic force microscopy. The analysis of existing experimental data shows that in most cases the chiral effects that we predict remain elusive. The theory therefore awaits final experimental verification.

Declaration of originality

I, the author, hereby declare that all the work presented in this Thesis is my original work. All the work done by other individuals, where present, was cited appropriately in the text.

The material reported here has been partially published in our two original research articles (Cortini et al., 2011, 2012).

Introduction

Deoxyribonucleic acid (DNA) is the carrier of genetic information in all living matter. The discovery of the structure of DNA by Watson and Crick (Watson and Crick, 1953), based on the X-ray images taken by Franklin and Gosling (1953), is perhaps one of the most successful scientific achievements of the past century. It was soon realized that the double helical structure is intimately related to its function. It is probably not an overstatement to say that the example of DNA is the most famous example of the relationship between structure and function in a biological molecule. Transcription and replication could not occur without the remarkable property of single strand DNA complementarity. In these cases, DNA is unwound and the two self-complementary strands are read or duplicated. There are other less well understood aspects of this relationship that may play a fundamental role in biological systems.

Decades of research into the biological and physical behaviour of DNA revealed fascinating phenomena that are not as famous and iconic as are transcription and replication. In 1965, Jerôme Vinograd discovered that in the polyoma virus, DNA was found in a closed circular form (Vinograd et al., 1965). In 1971, Lerman (1971) reported the existence of DNA condensation, when adding multivalent cations in the solution. In 1961, the DNA liquid-crystalline phase was discovered (Robinson, 1961). In 1984, the work of Rau and Parsegian (Rau et al., 1984) paved the way towards the quantitative understanding of the forces that DNA molecules experience when coming into proximity. These and other studies revealed that under certain conditions, DNA molecules attract each others. DNA is highly charged, as each base pair carries approximately two electron charges. This naturally leads to the question: how can two like-charged objects attract each others? This question aroused great interest, and spawned several theoretical works which attempted to

answer it.

In 1997, Alexei Kornyshev and Sergey Leikin proposed a theory of DNA-DNA interactions which explained the presence of attractive forces, based on the helical structure of DNA (Kornyshev and Leikin, 1997). In their theory, the helical structure of the charge distributions on the surface of DNA molecules was taken into account. Negatively charged phosphate groups could align with positively charged counterions on an opposing molecule, resulting in an attractive force. Such mechanism, which was named "electrostatic zipper", is at the core of the work proposed here.

Shortly after their work on parallel DNA molecules, Kornyshev and Leikin found that the appearence of attractive forces is not the only consequence of taking into account the helical structure of the DNA charge distributions. If the helices are tilted with respect to each other, it is easy to imagine that the tilt in one direction is not equivalent to the tilt in another direction: the helical nature of the charge distribution results in an intermolecular torque (Kornyshev and Leikin, 2000). This intuitive phenomenon was confirmed in the theory of skewed DNA molecules, where it was found that the electrostatic energy of interaction depends on the direction of the tilt. The preferred tilt direction is directly related to the handedness (chirality) of the DNA molecule: right-handed double helices prefer right-handed crossovers. In such case, positive and negative charges on opposing molecules align themselves in an optimum configuration.

When two rigid molecules skew, their molecular tips move apart, thereby losing their interaction energy. In conditions in which the molecules attract each others, this is not favourable. It is clear that there comes the necessity to benefit from the attractive force, and at the same time take advantage of the torque. These two tendencies are satisfied at the same time if the two DNA molecules wrap around each others, forming an intertwined structure: a braid. DNA molecules are in fact not rigid, and they may invest elastic cost of bending to gain advantage from the optimum alignment of charges that results from tilt. In order to evaluate this effect, it became necessary to develop a theory of electrostatic interactions of two braided DNA molecules. The first part of my PhD project was entirely devoted to this topic.

The goal of the project was then to use the theory of braiding to describe

three target cases, which we will illustrate below: (a) the formation of braids of two free-ended DNA molecules; (b) the case of closed loop DNA and (c) single-molecule DNA micromanipulations. The available experimental data on DNA is immense, but relatively few studies explored the conditions under which the chiral electrostatic forces are expected to be maximum. Comparison of the theoretical predictions to the available experimental data will give insight into the validity of the theory. Also, the theoretical results will suggest new experiments to be performed.

Supercoiling (braiding) of a closed circular DNA molecule is of fundamental biological importance. In prokaryotes, DNA is found in small circular species called plasmids, where it adopts an intertwined, braided tertiary structure (plectoneme). Supercoiling is so important that a special class of enzymes exist to control its level, and if this falls out of the physiological range, cell death rapidly occurs. Understanding DNA supercoiling is then of crucial importance.

The theory of DNA braiding suggests that if attractive forces are present in a closed circular DNA molecule, they may drive a transition towards an intertwined state, even when the molecule is in an open circular state with no torsional stresses. Such possibility was never before explored theoretically, and was not systematically studied in experimental works on DNA. Our second goal was then to apply our model of DNA braiding to the case of closed loop DNA supercoiling. The effects due to chiral attractive forces are expected to be evident in that case.

Single-molecule DNA micromanipulations is an ideal method to test our theoretical predictions. We address two different kinds of experiments of this kind. In the first type, a single DNA molecule is attached to a glass surface at one end, and to a rotating bead at the other end. Supercoiling arises there when rotating the bead. The other kind of experiment involves the formation of a single DNA braid by attaching two different DNA molecules to four beads. In both experimental methods, measuring the force acting on the beads allows to probe the structure and the interactions of the molecules.

The case of four-bead DNA supercoiling also involved the collaboration with the group of Prof. Gert van der Heijden (University College London, Mathematics Department) and Prof. Gijs Wuite (Vrije Universiteit Amsterdam, Physics Department). The group of Prof. van der Heijden developed a sophisticated model

of the elasticity of braids with local interactions. Such theory will be applied to the case of four-bead experiments, and such studies are ongoing. In Amsterdam, the four-bead experiments are being performed in Prof. Wuite's laboratory.

This thesis is divided into five chapters. Chapter 1 describes the features of DNA-DNA interactions and introduces the Kornyshev–Leikin theory. The idea of extending the theory to the case of DNA braids will become apparent at the end of the Chapter. In Chapter 2 we will introduce our theory of DNA braids. Its results will be illustrated and compared them with experimental data. Due to the high complexity of the calculation that was performed, the technical details of it have been put in appendix A and B. The theory of DNA braiding will be then applied to the case of closed loop DNA supercoiling in chapter 3. Here we will extensively discuss the rich experimental data available, and compare them to our theoretical predictions. In the final two chapters we will discuss the case of single-molecule DNA supercoiling (Chapter 4) and the case of four-bead DNA experiments (Chapter 5).

Contents

Abstract **3**

Declaration of originality **5**

Introduction **6**

1 Kornyshev–Leikin theory of DNA interactions **15**

1.1 DNA-DNA interactions 17

1.1.1 Multimolecular aggregates 17

1.1.2 Liquid crystalline phases 18

1.1.3 DNA condensation 19

1.1.4 Homologous pairing 20

1.1.5 Other studies 21

1.2 General theoretical framework 23

1.2.1 Poisson–Boltzmann equation 23

1.2.2 Debye-Hückel-Bjerrum theory 25

1.2.3 Other theoretical approaches 26

1.3 The general problem 27

1.3.1 Formulation of the problem 27

1.3.2 Fixed charge patterns in KL theory 29

1.3.3 Fourier transform formulation 31

1.4 Parallel DNA molecules 32

1.4.1 Model outline 33

1.4.2 Results 34

1.4.3 Discussion 36

1.5 Non ideal DNA . . . 36
1.5.1 DNA non-ideality: homologous and non-homologous interactions . . . 37
1.5.2 Torsionally flexible DNA . . . 38
1.6 Skewed molecules . . . 41
1.6.1 Asymmetric crossings: the chiral torque . . . 41
1.6.2 The chiral torque and the DNA cholesteric phase . . . 42
1.7 Summary and outlook . . . 42

2 Free DNA Braids **44**
2.1 Model . . . 45
2.1.1 Geometry assumptions . . . 45
2.1.2 Energy function . . . 47
2.1.3 Calculation strategy . . . 48
2.1.4 Sequence-related and thermal fluctuations . . . 49
2.1.5 Energy and free energy functions . . . 50
2.1.6 Model limitations . . . 52
2.2 Results . . . 53
2.2.1 Analysis of the model . . . 54
2.2.2 Free energy landscape . . . 55
2.2.3 Electrolyte effects . . . 59
2.2.4 The effect of the DNA pitch . . . 62
2.3 Discussion . . . 62
2.3.1 The Prentiss experiments . . . 63
2.3.2 Evidence for asymmetry in DNA crossovers . . . 64

3 Closed circular DNA **66**
3.1 Introduction . . . 67
3.1.1 Mathematical description of supercoiling . . . 67
3.1.2 Supercoiling in vivo . . . 69
3.1.3 Control of supercoiling: topoisomerases . . . 69
3.1.4 Hyperthermophiles . . . 71
3.1.5 Structure of naturally occurring supercoils . . . 71

3.1.6 Thermodynamics of supercoil formation . . . 73
3.1.7 The tighly wound supercoil enigma . . . 74
3.2 Model . . . 74
3.2.1 Geometry and topological constraint . . . 75
3.2.2 Energy function . . . 77
3.2.3 End loop energy . . . 78
3.2.4 Nicked molecules . . . 81
3.2.5 Model limitations . . . 82
3.3 Results . . . 83
3.3.1 Model analysis . . . 83
3.3.2 Nicked molecules energy . . . 85
3.3.3 Energy landscape . . . 86
3.3.4 State diagram . . . 89
3.4 Discussion . . . 89
3.4.1 Enthalpy of DNA supercoiling . . . 92
3.4.2 The effect of divalent metal ions on the DNA secondary structure . . . 92
3.4.3 Tight supercoiling and writhing of relaxed circular DNA . . . 94
3.4.4 Biological evidence for asymmetric DNA crossovers . . . 97
3.5 Summary and outlook . . . 97

4 Single molecule supercoiling 99
4.1 Introduction . . . 99
4.2 Model . . . 102
4.2.1 General features of the system . . . 102
4.2.2 Worm-like chain behaviour of the straight portions . . . 103
4.2.3 Force calculation . . . 104
4.2.4 Plectoneme geometry and energy function . . . 105
4.2.5 Solenoid geometry and energy function . . . 108
4.2.6 Model limitations . . . 110
4.3 Results . . . 112
4.3.1 Force versus extension experiments . . . 113
4.4 Force versus rotation experiments . . . 116

4.5 Discussion . . . 116
4.5.1 Single molecule experiments in the presence of condensing agents . . . 119
4.5.2 Experiments to be performed . . . 121

5 Four-bead DNA braiding 123
5.1 Model . . . 124
5.1.1 Geometry and energy function . . . 125
5.1.2 Model limitations . . . 126
5.2 Results . . . 127
5.2.1 Energy landscape . . . 127
5.2.2 Forces . . . 128
5.3 Experiments to be performed . . . 129

Conclusions 135

Appendix A The geometry of helices 137
A.1 Helices . . . 137
A.1.1 Definition . . . 137
A.1.2 Tangent vector . . . 138
A.1.3 Other useful relationships . . . 139
A.2 Writhe of a braid . . . 140

Appendix B Theory of electrostatic interactions in DNA braids 141
B.1 Ideal DNA molecules . . . 142
B.1.1 General strategy . . . 142
B.1.2 Geometry . . . 145
B.1.3 Fourier transforms of potentials and charge densities for arbitrary surface charge patterns . . . 147
B.1.4 Zero-order image charges . . . 150
B.1.5 First-order images . . . 151
B.1.6 Structure factors . . . 153
B.1.7 Ideal double helices . . . 154
B.1.8 Energy calculation: direct term . . . 157

B.1.9 Energy calculation: image term 159
B.2 Non ideal DNA: torsional flexibility and adaptation 160
B.2.1 Non-ideal double helices 161
B.2.2 Energy functional 162
B.2.3 Variational approximation 163

Bibliography **164**

Acknowledgements **179**

Chapter 1

Kornyshev–Leikin theory of DNA interactions

In 1953, the double helix structure of DNA was proposed (Watson and Crick, 1953), based on the X-ray diffraction pattern that was measured shortly before (Franklin and Gosling, 1953). These studies gave birth to modern biology, and spawned an enormous amount of studies of the DNA molecule.

The double helix is well known for its remarkable chemical property of *self-complementarity*. The two strands that make a single DNA molecule are made of complementary base pair sequences (see (Watson and Crick, 1953)). This structural feature lies at the heart of the fundamental biological processes of DNA replication and transcription. In the years, this aspect of the DNA structure has become the single most famous example of the relationship between structure and function of a biological molecule.

The double helical structure of DNA may also responsible for other less well-known phenomena. In the mid 1990's, A. Kornshev and S. Leikin developed a theory of electrostatic interactions between DNA molecules that explained the phenomenon of DNA condensation, by taking into account the helical structure of DNA. In this chapter, we illustrate such theory, and show how it may explain the fascinating experimental observation of attractive forces between intact, double-stranded DNA molecules.

Before describing the theory, we will first illustrate the basic features of DNA-

DNA interactions (section 1.1). In particular, we will focus on the topic of cation-mediated DNA attraction, and the phase behaviour of aggregates of DNA molecules in the presence of a particular class of ions, named DNA condensers. We will hint that the explanation may well be attributed to the double helix structure, an idea which will be formulated and explored in detail in section 1.4. We will show that the intrinsic chirality of DNA plays an important role in determining the interactions of a DNA molecule with itself.

The Kornyshev–Leikin theory is based on a theoretical framework which is called Poisson–Boltzmann (PB) theory. We show how the PB theory can be applied to DNA, and how it is possible in that case to make an approximation which simplifies considerably the mathematical problem. Such approximation is known as the Debye–Hückel–Bjerrum (DHB) theory, which will be discussed in some detail. We will discuss these theories in section 1.2. The section ends with a brief discussion of how other theories dealt with the problem of DNA-DNA interactions.

In the remaining part of this chapter we will give an overview of the Kornyshev–Leikin (KL) theory of DNA-DNA interactions. The theory first appeared in a 1997 Journal of Chemical Physics paper (Kornyshev and Leikin, 1997), and dealt with the calculation of the electrostatic interactions of a pair of infinitely long, rigid, ideal, parallel molecules. It was extended to the case of skewed (Kornyshev and Leikin, 2000), non-ideal (Kornyshev and Leikin, 2001), finite-sized (Kornyshev et al., 2002), non-rigid (Cherstvy et al., 2004) DNA molecules. More recently, undulations of the molecular axes were taken into account (Lee et al., 2010), and the theory was applied to give account of some previously unnoticed features of the X-ray diffraction images of DNA (Kornyshev et al., 2011). It is beyond the scope of this thesis to give full details of the theory. Instead, this chapter will focus on explaining the basic physical picture emerging from the theoretical framework of the KL theory, and discussing its assumptions, approximations, and results.

In section 1.3 we illustrate the general formalism of the KL theory. We discuss the basic assumptions and approximations that stand behind it, write the equations that govern the system under those assumptions, and explain how transforming those equations in Fourier space allows for a solution of the problem.

We will then describe the case of parallel DNA molecules in section 1.4. The important extension of the theory to the case of non-ideal, non-rigid molecules will

be introduced in section 1.5. In section 1.6 we will shortly introduce the extension of the theory to the case of skewed molecules, which gave the original idea to apply the KL theory to the case of DNA supercoiling.

Most of the work presented in this chapter has been reviewed in (Kornyshev et al., 2007). Successive work is cited appropriately.

1.1 DNA-DNA interactions

DNA is a very long, highly charged molecule. Typically, eukaryotic DNA and chromosomal prokaryotic DNA is orders of magnitude longer (can go up to a few mm) than the average size of a cell (~ 1 μm). DNA molecules therefore must be tightly and efficiently packaged into the small cellular space and in the tiny viral capsids. Because of its high bare charge, DNA interacts very strongly with itself and other DNA molecules. In this section we will give a brief account of the basic features of these interactions.

We present here what is known from experiments about the interaction of intact, double-stranded B-DNA molecules. At first we describe the experiments and theories of intermolecular forces in multimolecular aggregates (section 1.1.1); we then describe the liquid-crystalline phase of DNA (section 1.1.2). We will then move to the topic of DNA condensation in section 1.1.3. The last two sections are devoted to protein-free homologous pairing of DNAs (section 1.1.4), and to some other relevant experimental results on the topic (section 1.1.5).

1.1.1 Multimolecular aggregates

The study on the forces between intact, double-stranded DNA molecules in multimolecular assemblies began with the works of Rau et al. (1984). Although the phenomenon of DNA condensation was already known (see section 1.1.3), the forces were measured there for the first time. The experiments were based on measuring the osmotic stress as a function of the intermolecular separation in fibers of short (~ 300 bp) DNA fragments. It was found that in DNA assemblies under osmotic stress, the intermolecular forces are repulsive, and deviate substantially from the "classical" double-layer electrostatic forces predicted in the Gouy–Chapman theory

(see 1.2 for details). The behaviour of the short-range repulsion (at intermolecular distances shorter than ∼15 Å) was found to be exponential, with a decay length of about 3.4 Å, roughly half of the Debye length in those particular conditions (the Debye length will be defined in section 1.2). Later experiments confirmed this observation (Podgornik et al., 1989; Todd et al., 2008; DeRouchey et al., 2010).

Further studies indicated the presence of a long-range attractive force, when low concentrations (order of a few millimolar) of the multivalent ions spermine^{4+}, spermidine^{3+}, and cobalt (III)-hexammine^{3+} were added in the mixture (Rau and Parsegian, 1992*a*,*b*; Podgornik et al., 1994; Todd et al., 2008; DeRouchey et al., 2010). Such forces lead to the well-known phenomenon of DNA condensation (Bloomfield, 1996). The force versus interaxial separation curves were parametrized using coefficients that depend on the ionic species, the ionic strength, the temperature, and the anion type. It was found that the coefficients strongly depend on the cationic species and strength, but only weakly on temperature and anionic type (Podgornik et al., 1994).

The observation of the existence of an attractive force between DNA molecules is surprising, because DNA molecules have the same charge. These studies therefore aroused great interest in developing a theory that could account for these observations. Several theories were proposed (Rouzina and Bloomfield, 1996; Kornyshev and Leikin, 1999; Shklovskii, 1999*b*,*a*; Grosberg et al., 2002; Zhang and Shklovskii, 2005), which were reviewed by Kornyshev et al. (2007). We will give a brief account of these theories in section 1.2.3.

1.1.2 Liquid crystalline phases

In a study of the phase behaviour of polyelectrolytes (Robinson, 1961), the liquid-crystalline phase of DNA was first reported. Successive studies (Iizuka, 1977) demonstrated that the type of liquid-crystalline phase for DNA was chiral nematic (cholesteric). Much work was done afterwards, reviewed by Yevdokimov et al. (1988). The existence of a phase transition to a columnar hexagonal phase was discovered later on (Livolant et al., 1989).

The concentration and type of counterions in solution affects the phase properties of DNA. Systematic studies of these effects proved the existence of a columnar-

to-cholesteric transition (Pelta, Livolant and Sikorav, 1996; Pelta, Durand, Doucet and Livolant, 1996), which was later fully characterized (Yang and Rau, 2005).

A very recent experimental work on ultra-short DNA fragments measured the handedness of the cholesteric twist (Zanchetta et al., 2010), and revealed that it is left-handed.

The cholesteric torque is thought of having a major role in the packaging of DNA inside viral capsids. Monte Carlo studies revealed that the inclusion of a cholesteric torque drives a transition between a random coil and an order toroidal-shape particle, inside phage heads (Marenduzzo et al., 2009). It is therefore clear that chiral interactions may play an important part in determining the shape of DNA *in vivo*.

1.1.3 DNA condensation

In 1971, sedimentation analysis of DNA in the presence of polymers and salt revealed a transition to a compact form (Lerman, 1971). This was the first observation of DNA condensation. It was found shortly after that spermine and spermidine could induce the same phenomenon (Gosule and Schellman, 1976). Much work was done afterwards to study the conditions under which condensation could occur, and the shape of the condensates. It is worth noticing that DNA condensation is distinguished from aggregation (which was dealt with in section 1.1.1). Condensation is a phenomenon in which the DNA exhibits a coil-globule transition, whereas aggregation is a multimolecular phenomenon (Bloomfield, 1996).

Circular dichroism (Gosule and Schellman, 1978) and electron microscopy (Chattoraj et al., 1978) were used to assess the dependence of DNA condensation on ion type and concentration. It was found that neither Mg^{2+}nor putrescine (a divalent polyamine) was able to induce condensation. It was also found that DNA stays in its B-form even in the condensed state.

Later studies using static and dynamic light scattering (Wilson and Bloomfield, 1979; Allison et al., 1981; Post and Zimm, 1982; Widom and Baldwin, 1983), gel electrophoresis of cleaved DNA condesates (Marx and Reynolds, 1982), and freeze-etch electron microscopy (Marx and Ruben, 1983, 1984) clearly proved that the shape of the DNA condensates was a toroid. The toroids were found to have an

inner radius of $\sim$ 185 nm, and an outer radius of $\sim$ 500 nm. It was also found that such values have high fluctuations.

Further details of the structure of the condensed DNA toroids was obtained via X-ray scattering (Schellman and Parthasarathy, 1984), electron microscopy (Arscott et al., 1990; Plum et al., 1990; Arscott et al., 1995), fluorescence microscopy (Yoshikawa et al., 1996), cryo-electron microscopy (Böttcher et al., 1998), and atomic force microscopy (Martin et al., 2000). All these studies revealed that for DNA condensation to occur, at least $\sim$ 90% of the DNA bare charge must be compensated by positive charges at its surface.

Many other studies confirmed and extended these observations. A comprehensive review of the topic may be found in several reviews (Bloomfield, 1996, 1997; Teif and Bohinc, 2011). Since the topic of DNA toroid formation is going to be explored in our analysis of single molecule DNA experiments, we defer a more detailed discussion of it to chapter 4.

1.1.4 Homologous pairing

Pairing of homologous DNA sequences (homologous meaning that their base pair text is identical) is a vital and not yet understood process. It stands behind the fundamental biological mechanisms of DNA recombination during meiosis, and DNA repair. The process of DNA recombination is rather well characterized (Zickler and Kleckner, 1999), and it is known to involve a large class of proteins, named recombinases. The first step of the process, however, still remains still a mystery (Barzel and Kupiec, 2008). During this first step, long homologous DNA sequences align with each others, allowing then for the intervention of the recombinases. How can two homologous sequences "find" each others in the immense genomic material? As an attempt to answer this question, several studies aimed at determining whether homologous pairing could occur in a protein-free environment, that is, to investigate whether DNA has an innate, structural ability that allows for sequence recognition.

In a study by Baldwin et al. (2008), protein-free homologous pairing was demonstrated in fibril assemblies of short, 150 bp-long fragments. This study used fluorescence resonant energy transfer (FRET) to assess the proximity of two differ-

ent species of DNA molecules. The first type of molecules, with one sequence, was labelled with one dye, and another species, with another sequence, were labelled with a different dye. After application of a very mild osmotic stress, coalescence of the two different species was observed.

A more recent study (Danilowicz et al., 2009) detected homologous pairing of double stranded DNA molecules in a protein-free environment at a quasi single-molecule level. DNA molecules were tethered at the surface of a capillary, and homologous and non-homologous molecules, attached to a magnetic bead, were introduced in the capillary. Application of a magnetic force in direction orthogonal to the capillary surface was able to detect paired molecules. It was thus demonstrated that pairing between homologous DNAs was promoted by sequence homology, and that such pairing could occur already in *monovalent salt*. None of the theories of DNA interactions, summarized in 1.2.3, gave a prediction of pairing of DNA molecules in monovalent salt. We anticipate here that we gave a plausible explanation of such unexpected result by the theory of DNA braiding (chapter 2).

1.1.5 Other studies

In this final subsection, we briefly describe other experiments that addressed the problem of DNA-DNA interactions.

The **diffraction of short DNA oligonucleotides** in crystalline aggregates was studied reported in (Timsit et al., 1989; Timsit and Moras, 1991, 1994). The analysis of the diffraction patterns in the presence of magnesium revealed important structural features of the intermolecular interactions. In fact, it was shown that two laterally displaced DNA molecules are locally skewed, in a right-handed configuration (for definition, see 2.1). Furthermore, the experiments showed that one magnesium atom acts as a bridge between the phosphate group of one molecule, and a solvent-exposed O6 and N7 group in a CG base step on the major groove of the other molecule. These experiments prove that in crystals of DNA molecules, a right-handed DNA crossing is stabilized by the presence of magnesium ions. The same conclusion was reported using all-atom molecular dynamics of short (14 bp) DNA molecules (Várnai and Timsit, 2010).

Small-angle X-Ray scattering yields precious information on the interac-

tions of pairs of DNA molecules. Such experiments were preformed on 150 bp-long DNA molecules in monovalent (Qiu et al., 2006) and divalent (Qiu et al., 2007) solutions. The data revealed that DNA molecules repel each others in monovalent salt, at all salt and DNA concentrations. However, in the presence of magnesium, it was shown that there is an attractive force between the molecules. Surprisingly though, at a given DNA concentration, the magnitude of the attraction increases with *decreasing* DNA length. If the attraction was due to lateral juxtaposition of oppositely charged groups, then the *opposite* trend would be expected. The results were therefore interpreted as an indication of the existence of a head-to-head attraction between DNA molecules, which is somehow mediated by magnesium ions. The end-to-end spontaneous association of short DNA fragments was also reported in a study of the liquid-crystalline phase of DNA (Nakata et al., 2007). Atomistic simulations very recently provided the evidence for such attractive forces also by means of *in silico* studies (Maffeo et al., 2012). Similar studies were also performed by Bai et al. (2005), which concluded that in the presence of magnesium there could be a small attractive force, but it is not sufficient to make short DNA fragments pair.

The **knotting probability of long DNA molecules** can be used to reconstruct the DNA interactions. Long, linear DNA molecules are ligated in the presence of a given amount of salt in solution. The equilibrium species are then run on a gel, and knots of different types are found. The knotting probability can be inferred by measuring the relative amount of the knot types in the gel. Monte Carlo simulations are then performed to have a theoretical estimate of the knotting probability. From the comparison of the experimental and simulated knotting probabilities, the *DNA effective radius* may be extracted. The effective radius is a measure of the amount of DNA-DNA repulsion. It was measured in this way as a function of the monovalent salt concentration (Rybenkov et al., 1993; Shaw and Wang, 1993), and in the presence of both mono- and di-valent ions (Rybenkov et al., 1997). However, the same study proved that in the presence of a concentration of $Mg^{2+} > 50$ mM, the effective DNA radius was negative, which is an indication of attractive DNA-DNA interactions.

1.2 General theoretical framework

The KL theory aims at the calculation of the electrostatic interaction energy between two DNA molecules. The theoretical framework on which it is based upon is the Poisson-Boltzmann (PB) equation, which is a non-linear partial differential equation that allows the calculation of the electric field due to a system of fixed charges immersed in an electrolyte solution. The non-linearity of the equation makes it very difficult to solve, unless the geometry is very simple. However, it is sometimes possible to simplify it to its linear form, known as the Debye–Hückel–Bjerrum (DHB) approximation. Here, we first describe the PB theory, then state the DHB approximation. Finally, we will discuss other theoretical approaches to the same problem.

1.2.1 Poisson–Boltzmann equation

The simplest approach to describe the electrostatic interactions within an electrolyte solvent is the Poisson–Boltzmann (PB) theory. Given a system of fixed charges described by the volume charge density function $\rho(\mathbf{r})$, the PB equation states:

$$\nabla^2\varphi(\mathbf{r}) = -\frac{4\pi}{\varepsilon}\rho(\mathbf{r}) - \frac{4\pi e_0}{\varepsilon}\sum_i n_i q_i \exp\left[-\frac{eq_i\varphi(\mathbf{r})}{k_B T}\right], \tag{1.1}$$

where ε is the dielectric constant of the solvent (assumed to be uniform here, but the theory can be extended to the case of non-local dielectic constant), $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplacian operator, e_0 is the elementary charge (which is taken to be a positive number), and the sum is performed over all the ion species present in solution, which have a number volume density of n_i and valence q_i. In this thesis, Gaussian units are always used.

The PB equation is based on an continuum, equilibrium, mean-field approach. To derive it, one assumes that the chemical potential of each of the ionic species present is spacially homogeneous in the absence of the fixed charges, and that there are no correlations between the ions. As a result, the equilibrium concentrations of the ions are related to the electric field via the Boltzmann factor.

It is then important to keep in mind that the PB theory should be applied only in the case of sufficiently low ionic concentrations (Kirkwood, 1934). The theory

in fact assumes that there are no correlations between the ions, which is a valid assumption only for dilute solutions. In the following, we will describe in a more quantitative way of stating the applicability of the theory.

Consider a polyelectrolyte with surface charge[1] $\hat{\sigma}$, in an electrolyte with counterions of valence q. The magnitude of the importance of counterion correlations is gauged by the parameter (Rouzina and Bloomfield, 1996)

$$\Xi = \frac{q^2 l_B}{l_{\hat{\sigma}}}, \tag{1.2}$$

where

$$l_B = \frac{e^2}{\varepsilon k_B T}. \tag{1.3}$$

is the Bjerrum length (in water at room temperature $l_B \approx 7$ Å), and

$$l_{\hat{\sigma}} = \frac{e_0}{2\pi l_B |\hat{\sigma} q|}. \tag{1.4}$$

The length $l_{\hat{\sigma}}$ is known as the Gouy–Chapman length. It was found that the PB equation is exact when $\Xi \to 0$. In the case in which the parameter Ξ is very large, one obtains the so-called "strong coupling" limit, which will be briefly discussed at the end of this section. The PB equation should not be used for high values of Ξ.

Also, the PB theory does not take into account any steric effects. That is, its applicability is limited to the case of small, point-like counterions. For the case of large, bulky ions such as the polyamines that cause DNA condensation (see 1.1.3), the theory is not expected to be valid.

Despite the limitations mentioned above, the PB equation is widely used in studies of biological molecules in physiological conditions (in which the ionic strength is ≈ 150 mM). It should be noticed that, because of the reasons stated here, the theory should not be considered to be applicable to this case. However, reasonable agreement between the results of the PB theory with those of atomistic computer simulations was obtained, and it was argued that this might be attributed to for-

[1] Usually the symbol used for the surface charge density is σ. Here, we use $\hat{\sigma}$ in order not to make confusion with the specific linking difference, discussed in chapter 3.

tuitous cancellations of opposite terms (for a review, see (Sharp and Honig, 1990), and the more extensive discussion reported in (Kornyshev et al., 2007)).

1.2.2 Debye-Hückel-Bjerrum theory

For sufficiently dilute solutions, and in the absence of highly charged molecules, one may expand the exponential in equation (1.1) in power series, and retain only the first term in expansion:

$$\nabla^2\varphi(\mathbf{r}) = -\kappa^2\varphi(\mathbf{r}) - \frac{4\pi}{\varepsilon}\rho(\mathbf{r}). \tag{1.5}$$

The zero-order term in the expansion vanishes because of the electroneutrality of the solution. In such way, one obtaines the so-called **Debye–Hückel equation** (DH). Here, the **Debye screening length** was defined, as

$$\lambda_D = \kappa^{-1} = \left(4\pi l_B \sum_i n_i q_i^2\right)^{-1/2}. \tag{1.6}$$

Compared to the PB equation, the DH equation has the advantage of being again *linear*.

For a point charge immersed in an electrolyte, the solution to the DH equation yields a potential that goes as $\varphi(r) \sim \exp(-r/\lambda_D)/r$. Compared to the case of a point charge in vacuum, it is clear that the main difference is the strong exponential decay of the potential, which effectively makes this interaction short-ranged. Physically, this is due to the screening of the electrostatic interactions because of the presence of the electrolyte.

For a highly charged molecule (such as DNA) immersed in an electrolyte solution, one may expect that there is a layer of strongly adsorbed or chemisorbed counterions at the surface of the molecule, usually referred to as the Gouy–Chapman layer. Within this nonlinear screening layer, the PB theory is not applicable. However, outside of the layer we may expect that the PB equation is again valid. Furthermore, given that the counterions within the nonlinear screening layer are expected to be strongly adsorbed on the highly charged surface, they may be treated as part of the fixed charge distribution at the surface. Then, the potential

outside of the nonlinear screening layer can be calculated using the DH equation, and incorporating the adsorbed counterions as part of the fixed charge distribution $\rho(\mathbf{r})$. This approach is known as **Debye–Hückel–Bjerrum theory** (DHB). The counterions are then distinguished into two types. The first type is the one present in the bulk of the solution, and contributes to the Debye screening length. The second type is strongly condensed or adsorbed at the charged surface. As we shall see, the Kornyshev–Leikin theory of DNA interactions relies on such approach. The successes of the theory in describing the electrostatic interactions between polyelectrolytes in solution was extensively discussed in (Kornyshev et al., 2007).

The thickness of the nonlinear layer may be evaluated by estimating the distance at which the potential energy close to the surface $e_0 q\varphi$ is comparable to the thermal energy $k_B T$. For DNA, one finds that such distance is shorter than the average roughness of the DNA surface (Kornyshev et al., 2007), and it is therefore very reasonable to apply the DHB approach.

The amount of charge that has to be considered as part of the fixed charges, and its spacial distribution, has to be estimated independently. We will see that this constitues a major challenge for the KL theory.

1.2.3 Other theoretical approaches

All other theories of DNA interactions consider the charge distribution on DNA to be homogeneously distributed over the molecular surface. Here, we give only a very short introduction to these theories. For a complete discussion, see (Kornyshev et al., 2007).

The amount of condensed charge on the surface of highly charged polyelectrolytes was calculated by Manning (1969). Based on the same theoretical framework, theories of DNA-DNA interactions that include multivalent counterion correlations were developed (reviewed in (Grosberg et al., 2002)). Such approach led to what is often referred to as "counterion condensation" theory.

In the limit of very high values of Ξ, it was argued that the counterions may form a 2D Wigner crystal at the molecular surface (Rouzina and Bloomfield, 1996; Shklovskii, 1999*b*,*a*; Grosberg et al., 2002). For DNA-like molecules, the formation of such structures would require a very high value of the valence ($q > 10$), but it

was found that already for tri-valent counterions such effect may become dominant. In fact, at close DNA-DNA separations, it might occur that a high amounts of positive charge accumulates in the space separating the DNA molecules. This effect would give rise to an effectively two-dimensional layer of mobile counterions. The correlated motions of the counterion cloud in the proximity of the two DNA molecules would give rise to the attractive forces observed in DNA condensation experiments. Under such conditions, it was theoretically predicted that also charge inversion of the polyelectrolyte may occur (Grosberg et al., 2002). The charge inversion was predicted to give rise to a "reentrant" behaviour (Zhang and Shklovskii, 2005): at high enough values of the condensing agent concentration, the DNA surface becomes overcharged and the amount of condensed DNA decreases. Such behaviour was experimentally observed in recent single-molecule experiments (Besteman, Van Eijk and Lemay, 2007). It is worth noticing that such reentrant condensation may also occur because of other reasons (Todd and Rau, 2008), which we will discuss in more detail in chapter 4.

1.3 The general problem

In this section we describe the formulation of the problem which the Kornyshev–Leikin theory deals with. We will first write general equations, within the Debye–Hückel–Bjerrum theory discussed in section 1.2.2. The choice of how to describe the charge patterns, which lies at the heart of the KL theory, will then be explained and discussed. We will then reformulate the problem in the language of Fourier transforms of charge densities, and show how the theoretical problem becomes much simpler in this way.

1.3.1 Formulation of the problem

The KL theory treats the interaction between DNA molecules as being purely electrostatic (only in the most recent development of the theory (Lee et al., 2010; Kornyshev et al., 2011), steric effects have been taken into account). The reason to neglect all other types of interactions is that the charge density on the DNA surface is very high, and that it was shown that the van der Waals interactions are

very weak compared to the electrostatic ones (Kornyshev et al., 2007).

To start, we write the the electrostatic interaction energy of a pair of DNA molecules:

$$E_{int}(R) = E(R) - E(\infty), \tag{1.7}$$

which is expressed as the energy required to bring one of the molecules from infinity to a distance R away from the first. We may write it as

$$E(R) = \frac{1}{2}\int \mathrm{d}\mathbf{r}\varphi(\mathbf{r})\rho(\mathbf{r}). \tag{1.8}$$

Here, $\varphi(\mathbf{r})$ is the electrostatic potential generated by the system of charges described by $\rho(\mathbf{r})$. Such potential is found by solving equation (1.5), with the correct boundary conditions (see below).

Let the surface of the two DNA molecules be described mathematically by S_1 and S_2. To solve the problem, we separate the electrostatic field into the one outside the dielectric cores ($\varphi_{out}(\mathbf{r})$) and the one inside the dielectric cores ($\varphi_{in}^{(\mu)}(\mathbf{r})$), where the index $\mu = 1, 2$ labels the molecules. Then the boundary conditions read as follows:

$$\begin{cases} \varphi_{out}(\mathbf{r})\big|_{\mathbf{r}\in S_\mu} = \varphi_{in}^{(\mu)}(\mathbf{r})\Big|_{\mathbf{r}\in S_\mu} & \text{(1.9a)} \\ \varepsilon_w \hat{\mathbf{n}}_{S_\mu}\cdot\nabla\varphi_{out}(\mathbf{r})\big|_{\mathbf{r}\in S_\mu} = \varepsilon_c \hat{\mathbf{n}}_{S_\mu}\cdot\nabla\,\varphi_{in}^{(\mu)}(\mathbf{r})\Big|_{\mathbf{r}\in S_\mu} \qquad \mu = 1,2. & \text{(1.9b)} \end{cases}$$

Here, $\hat{\mathbf{n}}_{S_\mu}$ is the unit vector normal to the surface S_μ, ε_w is the dielectric constant of water and ε_c is the dielectric constant of the molecular cores. The first equation is the continuity of the potential across the dielectric interfaces, the second equation is the continuity of the normal gradient.

Inside the dielectric cores of the DNA molecules the electrostatic field is given by the Laplace equation:

$$\nabla^2\varphi_{in}^{(\mu)}(\mathbf{r}) = 0 \qquad \mu = 1, 2. \tag{1.10}$$

We therefore have two different equations describing the electric field in space.

To solve the problem, we take into account that the presence of the dielectric

interface gives rise to a system of induced charges, which are sometimes referred to as "image charges". Physically, they arise because of the difference in dielectric constant inside and outside of the DNA core. From the mathematical point of view, they provide a convenient way of treating the boundary conditions.

We call the system of induced charges $\rho_{ind}(\mathbf{r})$. The electrostatic field may then be represented as only one function, which is analytically continued across the dielectric interface, which we call simply $\varphi(\mathbf{r})$. In this way the equation we have to solve is the following:

$$\nabla^2\varphi(\mathbf{r}) = -\kappa^2\varphi(\mathbf{r}) - \frac{4\pi}{\varepsilon}\left[\rho(\mathbf{r}) + \rho_{ind}(\mathbf{r})\right]. \tag{1.11}$$

The problem then is fully formulated. In the next subsection, we will discuss the choice of the fixed charges $\rho(\mathbf{r})$.

1.3.2 Fixed charge patterns in KL theory

As discussed earlier (see 1.2.2), the DHB theory assumes that the counterions can be distinguished into two classes: those that are bound to the oppositely charged surface, and those that are present in solution and contribute to the Debye screening length. In the Kornyshev–Leikin theory, it is assumed that the counterions bound to the DNA surface neutralize a total fraction θ of the total negative charge. Out of that amount of charge, a fraction f_1 is localized at the centre of the minor groove, a fraction f_2 at the centre of the major groove, and the remaining part $(1 - f_1 - f_2)$ is homogeneously smeared on the DNA surface. Notice that formally the charge density is fixed, but in reality the pattern of counterions at the DNA surface cannot be considered as fixed. The positive charge density must be interpreted as an *average* of the positions of the positive counterions. The most important assumption of the KL theory is that such density is primarily determined by the molecule to which the counterions bind to, and therefore follow the same helical path as the fixed negative charges. The charge pattern on a DNA surface is then described in terms of four charged helical lines.

The Kornyshev–Leikin theory does not explicitly give an estimate for the value of the parameters θ, f_1 and f_2, hereafter referred to as "charge compensation

parameters". Such values are expected to depend strongly on ion type, but weakly on ion concentration (Kornyshev et al., 2007). The reason for not attempting to estimate these numbers is that it would either require an extremely sophisticated theory, or to perform very heavy numerical simulations. In what follows, we briefly discuss the reasons for such high complexity.

The interaction of ions with DNA is complicated by many factors (see (Hud, 2009) for a complete discussion). Firstly, the interactions are strongly dependent on local, specific, and sterically asymmetrical binding potentials, which are experimentally difficult to probe. Small-angle X-ray scattering gives information on the distribution of ions that specifically bind to DNA (Qiu et al., 2006), but its use is limited in the case of Na^+ and K^+, which have a mobile distribution. The case of Na^+ is particularly complex because it is isoelectronic to water, and therefore very difficult to distinguish from the background. Second, it has been shown that in the presence of ions of different species and concentrations short DNA oligomers adopt different conformations (Leslie et al., 1980). This effect was shown to be dependent on the sequence and on the base pair composition.

The same specific ion-DNA interactions are difficult to extract also from simulation data, because of the quantum nature of the interactions themselves: any "classical", force-field-like approach is likely to give model-dependent results (see, e.g. (Mocci and Laaksonen, 2012) for a review).

Despite the difficulties mentioned here, there are a few cases in which it is possible to have a rough idea of the value of the charge compensation parameters. Namely, it is the case in which either the ions do not have any specific binding potential to DNA (such as Na^+or K^+), or the case in which the ions are known to have very strong preference for the minor or major grooves of DNA (such as spermidine and spermine).

For counterions that do not have specific binding potentials to the DNA surface, the theory of counterion condensation by Manning (1969) may be used. The Manning value in such case is around 70%, and we expect that the parameter θ may be of that order. This value is also close to the value that is obtained by extracting the effective DNA charge density in equivalent solvent conditions from DNA supercoiling experiments (Ubbink and Odijk, 1999), which is around 66%.

The case of divalent ions that do not have very strong binding affinities to

DNA is more difficult and delicate. Raman spectroscopy studies (Duguid et al., 1993) revealed an interesting dependence on the type of divalent ion on the binding properties of the ions to DNA. In fact, it was observed that apparantly Mg^{2+} and Ca^{2+} do not induce any detectable structural deformation of DNA, but transition metals such as Mn^{2+} and Cd^{2+} do. This is probably an indication of specific adsorption patterns.

The case of spermine and spermidine is more clear, since many different studies convincingly show that these ions bind to the DNA grooves (Feuerstein et al., 1986, 1990; Ruiz-Chica et al., 2001; Ouameur and Tajmir-Riahi, 2004). NMR studies revealed that although the binding constant of spermine and spermidine to DNA is very high, its mobility is effectively independent of the mobility of DNA (Wemmer et al., 1985). This indicates that the charge distribution in those cases is highly dynamic. In these cases however the validity of the PB theory is questionable, as pointed out in section 1.2.

1.3.3 Fourier transform formulation

The analytical problem we are dealing with here can be simplified if we write the equations in Fourier space. Using the definitions of the Fourier transforms given in Box 1, and the notation given therein, we may write the DHB equation as:

$$\hat{\varphi}(\mathbf{k}) = \frac{4\pi}{\varepsilon} \frac{\hat{\rho}(\mathbf{k})}{k^2 + \kappa^2}, \tag{1.12}$$

where $k = |\mathbf{k}|$.

The energy in equation (1.8) can be expressed conveniently in terms of the Fourier transforms, from straightforward application of the definition of the Fourier transform (b.1.1), integration using equation (b.2.15), and the expression for the electrostatic potential given in equation (1.12):

$$E = \frac{2\pi}{\varepsilon} \int \mathrm{d}\mathbf{k} \tilde{\varphi}(\mathbf{k}) \tilde{\rho}(-\mathbf{k}). \tag{1.13}$$

Box 1: Fourier transforms and cylindrical Fourier transforms

We define the Fourier transform of a three-dimensional function $f(\mathbf{r})$ as

$$\tilde{f}(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int \mathrm{d}\mathbf{k} \exp(i\mathbf{k}\cdot\mathbf{r}) f(\mathbf{r}). \tag{b.1.1}$$

The inverse Fourier transform is then given by

$$f(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} \int \mathrm{d}\mathbf{k} \exp(-i\mathbf{k}\cdot\mathbf{r}) \tilde{f}(\mathbf{k}). \tag{b.1.2}$$

We describe here also the **cylindrical Fourier transform** (Cormack, 1957), which is very important for our calculations. Given a two-dimensional function expressed in cylindrical coordinates $g(\phi, z)$, the cylindrical Fourier transform is defined as

$$\tilde{g}(n, q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}z \exp(iqz) \int_{0}^{2\pi} \mathrm{d}\phi \exp(in\phi) g(\phi, z) \tag{b.1.3}$$

The inverse is given by

$$g(\phi, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}q \exp(-iqz) \sum_{n=-\infty}^{\infty} \exp(-in\phi) \tilde{g}(n, q) \tag{b.1.4}$$

Taking then into account equation (1.11), we then have

$$E = \frac{2\pi}{\varepsilon} \int \mathrm{d}\mathbf{k} \frac{\tilde{\rho}(\mathbf{k}) \left[\tilde{\rho}(-\mathbf{k}) + \tilde{\rho}_{ind}(-\mathbf{k})\right]}{k^2 + \kappa^2}. \tag{1.14}$$

The main theoretical problem is then to calculate the Fourier transforms of the image charge systems, which are determined by imposing boundary conditions, and then calculating the integral in equation (1.14).

1.4 Parallel DNA molecules

In this section we will summarize the derivation, results and implications of the first builing block of the Kornyshev–Leikin theory of DNA interactions: the theory of rigid, ideal, infinitely long parallel DNA molecules. The material reported here

is described in greater detail in the original paper (Kornyshev and Leikin, 1997), and its follow-up (Kornyshev and Leikin, 1999).

The purpose of this section is to give a first qualitative description of the theory, in terms of its derivation, its physical predictions, and the experimental data that it reproduces. No technical details will be given here, since a much more detailed description of all of it will be given to chapter 2.

1.4.1 Model outline

The KL theory models DNA molecules as cylinders. The cylinders are filled with a dielectric material, with dielectric constant much lower than the dielectric constant of water. All the rest of the space is modelled as a continuum of solvent, which is described by the DHB equation (see section 1.2.2). All the charge distributions follow the helical symmetry of the DNA molecule. The physical-mathematical problem of calculating analytically the energy of interaction of such system may seem formidable. However, it was found that it is possible to find a fully analytical solution to it.

Let the charge distributions on the two DNA molecules be described by the density functions $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$. From the point of view of the technical task of performing the calculation, the basic simplification that allows us to perform the analytical calculation is to express the charge density of cylinder 1 as related to that of cylinder 2, through a simple space translation:

$$\rho_2(\mathbf{r}) = \rho_1(\mathbf{r} - \mathbf{R}), \tag{1.15}$$

where $\mathbf{R}$ is the vector that connects the axes of the two cylinders. Using equation (1.15), from the definition of the Fourier transform equation (b.1.1), is is easy to see that

$$\hat{\rho}_2(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{R}}\hat{\rho}_1(\mathbf{k}) \tag{1.16}$$

Since the formula that connects the two Fourier transforms of the charge distributions is so simple, the integrals in equation (1.14) become feasible. The remaining problem is to impose boundary conditions so that we can calculate the system of induced charges.

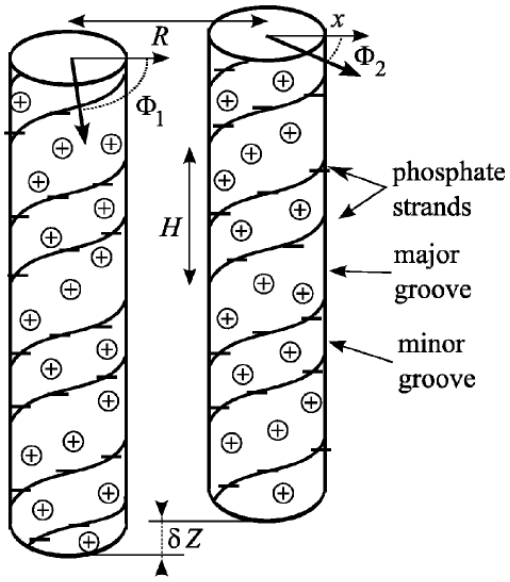


Fig. 1.1: Geometry of two interacting DNA molecules, within the description of the KL theory. The figure defines: the interaxial distance R; the azimuthal angles Φ_1 and Φ_2; the DNA pitch H; the axial shift δZ. The negatively charged phosphate strands and the positively charged adsorbed counterions are also shown. Figure taken from (Kornyshev et al., 2007).

In Fourier space, it is possible to impose boundary conditions on the surfaces of the molecules by introducing the cylindrical Fourier transform (see box 1).

1.4.2 Results

The formulas that describe the interaction of two parallel DNA molecules calculated within the KL theory will be given in chapter 2. Here, we illustrate the qualitative physics arising from the calculations.

Figure 1.1 shows two interacting parallel DNA molecules as described in the KL theory. The figure shows the geometrical variables that define the system: R, the interaxial distance, and Φ_1 and Φ_2, which are the azimuthal orientations of the centre of the minor grooves of molecule 1 and 2, respectively. The figure also shows that the positively and negative charges on opposing molecules may align, and stay in optimum alignment throughout the length of the molecular juxtaposition. This is precisely due to the helical nature of the charge distributions, and it results in the so-called "**electrostatic zipper**" effect. The attraction between DNA molecules

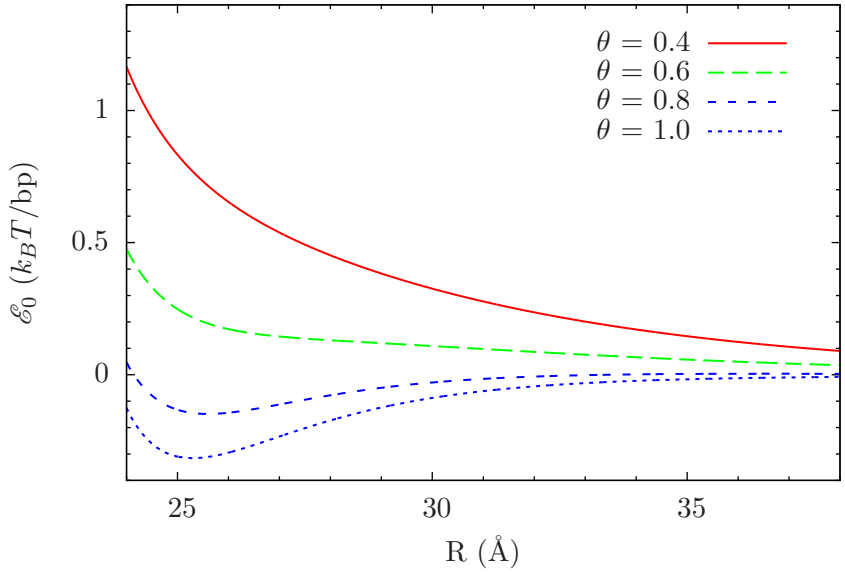


Fig. 1.2: Minimum energy per base pair as a function of the interaxial distance of two interacting ideal DNA molecules. The curves were plotted at fixed values of θ, using $f_1 = 0.4$, $f_2 = 0.6$, and $\lambda_D = 7.0$ Å. The minimum was calculated by numerical minimization of equation (2.3).

is explained by the KL theory in terms of this simple physical picture. Now we quantitatively explore the magnitude of this effect, as a function of the charge compensation parameters.

Figure 1.2 shows the dependence of the minimum energy as a function of R, at different values of θ. The figure was obtained from numerical minimization of equation (2.3), and at each point the optimum value of $\Delta\Phi$ was chosen (see section 2.2 for more details). This figure clearly shows the importance of the parameter θ in the context of the KL theory. In fact, different values of θ can switch between an attractive regime (in which there is a negative-energy minimum, shown in figure) and repulsive regime. Importantly, it should be noticed that the interaction energy can switch to attractive at values of θ smaller than 1. This means that attraction between DNA molecules are predicted by the theory even without full charge compensation.

An important feature of the interaction energy derived here is that it is made up of two different components: a direct electrostatic term, and an "image" term. The first term describes the interaction of the charges on a molecule with the real

charges on the other molecule and may be attractive or repulsive depending on the values of θ, f_1 and f_2. The "image" interaction term is the interaction of the charges on a molecule with the *induced* charges on the other molecule, and is always repulsive. The nature of this term implies that the image interaction component has half the decay length of the direct term. The energy minimum is therefore at the balance of the direct (attractive) term and the image (repulsive) term.

1.4.3 Discussion

The theory of parallel molecules derived in (Kornyshev and Leikin, 1997) was originally developed to describe the interaction of any long polyelectrolyte with a helical charge pattern on its surface. Applied to DNA, the theory successfully reproduced the experimental observation (Rau and Parsegian, 1992*b*; Todd et al., 2008) that the repulsive inter-DNA forces decay twice as fast as the attractive forces. This is due to the fact that the image forces are due to the propagation of the field that must travel exactly twice the distance in the solvent. This experimental observation is also consistent with the "hydration force" mechanism (Marčelja and Radić, 1976), which we will not discuss here.

1.5 Non ideal DNA

In the previous section we showed how the optimum alignment of positive and negative charges on opposing DNA surfaces may lead to attractive intermolecular forces. It is clear that such attraction heavily relies on the concept of *phase locking* of the two charge distributions. In fact, for the attraction to occur, positive and negative charges must remain aligned throughout the whole length of the lateral contact. Any thermal fluctuation and deviation of the DNA charge structure away from the ideal helical line will disrupt the phase locking, and result in a net repulsive force.

This effect was studied in (Kornyshev and Leikin, 2001) for rigid DNA. It was found that the interaction energy as a function of the DNA length decreased monotonically for homologous pairs, but eventually increased, and became positive, for

non-homologous pairs. It was soon realized that the DNA molecules may also adapt their structure to keep the positive and negative charges aligned. In fact, DNA is torsionally flexible, and it can invest some torsional energy to benefit from the favourable electrostatic interactions. A theory that incorporated this effect was developed by Cherstvy et al. (2004), and extended later by Wynveen et al. (2008) and Lee et al. (2010).

In this section, we first describe the idea of DNA non-ideality, and the loss of the favourable interactions when dealing with non-ideal DNA molecules (section 1.5.1). We then illustrate the theory of non-rigid DNA, and how to incorporate thermal and sequence-dependent fluctuations in the theory (section 1.5.2).

1.5.1 DNA non-ideality: homologous and non-homologous interactions

In their famous experiments, Dickerson and Drew (1981) proved that the DNA parameters that describe the tilt, roll, shift, slide, rise and twist between the base pairs are sequence-dependent, and deviate away from those of an ideal helical structure. The parameters of real DNA were quantified therein. It was proved that the most significant non-ideality is the one due to twist. Importantly, the effect is sequence-dependent in a non-symmetric way, that is, it depends on the direction of the strand.

Because of its non-ideal structure, the charge patterns on a real DNA molecule do not follow closely the pattern of an ideal helical line. An useful parameter to describe the global, macroscopic deviation away from an ideal helix is the **helical coherence length**. One can define the phase angle of the phosphate charges at a particular height, $\Phi(s)$, and then calculate the correlation function between two points. For two given heights s and s', provided that $|s' - s|$ is much larger than the axial rise per base pair l_c, it was shown that (Kornyshev and Leikin, 2001; Cherstvy et al., 2004):

$$< [\Phi(s') - \Phi(s)]^2 > \approx \frac{|s' - s|}{\lambda_c}, \tag{1.17}$$

that is, the mean square displacement accumulates in a random walk-like way.

The deviation away from the ideal helix has two different sources: thermal fluctuations and sequence-dependent effects. As to the coherence length, these two contributions add up like (Lee et al., 2010)

$$\frac{1}{\lambda_c} = \frac{1}{l_p^h} + \frac{1}{\lambda_c^{(0)}}, \tag{1.18}$$

where $l_p^h \approx 350$ Å is the helical persistence length, which is primarily determined by thermal fluctuations, and $\lambda_c^{(0)}$ is the intrinsic, sequence-dependent helical coherence length, in the absence of any thermal fluctuations (see appendix B for more details).

One can also define the same coherence length, but for a *pair* of DNA molecules. In such case, the result is exactly the same, but the meaning of the intrinsic coherence length is different. For homologous molecules (molecules that have the same base pair text, and juxtaposed in the same orientation), the intrinsic coherence length is infinite. On the other hand, the intrinsic coherence length for a pair of non-homologous molecules must be estimated from an ensemble average over all possible deviations away from the ideal helical structure. Such value was estimated from the analysis of NMR studies on DNA oligomers to be ≈ 150 Å (Wynveen et al., 2008). The result is the following:

$$\lambda_c = \begin{cases} l_p^h \approx 350 \text{ Å} & \text{Homologous molecules} \\ \frac{l_p^h \lambda_c^{(0)}}{l_p^h + \lambda_c^{(0)}} \approx 105 \text{ Å} & \text{Non-homologous molecules.} \end{cases} \tag{1.19}$$

The main physical result of this analysis is that the interaction between homologous DNA pairs and non-homologous ones is different. The theory predicts that under all circumstances homologous pairs will have stronger attractive forces. This effect was found in the analysis of the interaction of non-ideal, rigid DNA molecules (Kornyshev and Leikin, 2001), and the main physical result is shown here in figure 1.3.

1.5.2 Torsionally flexible DNA

DNA molecules are torsionally flexible, so that to gain advantage from the favourable electrostatic interaction, they may adapt their structure to keep the oppositely

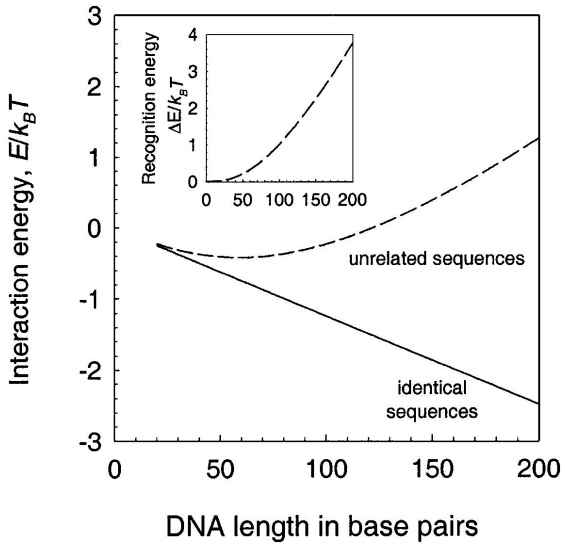


Fig. 1.3: Minimum interaction energy of a pair of homologous (solid curve) and non-homologous (dashed curves) DNA molecules, as a function of their length. Figure taken from (Kornyshev and Leikin, 2001)

charged groups in optimum alignment. To do so, the molecules must invest some torsional energy. Cherstvy et al. (2004) calculated the interaction energy of torsionally flexible DNA molecules that interact through the electrostatic energy function given in the KL framework.

To calculate the interaction of torsionally flexible DNA, one must take into account two different energy components, which are the torsional energy and the electrostatic energy of non-ideal DNA. In this section, we briefly explain the basic idea that stands behind the calculation. In chapter 2 we will give more details, and the full details of the calculation are given in appendix B.2.

The torsional energy cost of DNA structural deformations is taken into account using the elastic rod model, which we describe in detail in box 8. The electrostatic interaction energy of non-ideal DNA molecules may be calculated within a variational approximation, proposed by Lee et al. (2010), and described in section B.2.3. As a result, after calculating the partition function, one can write a free energy function which depends on an additional variable, $\lambda_h^\star$. This variable describes the length scale at which the electrostatic forces restore the optimum alignment of oppositely charged groups. The optimum value of $\lambda_h^\star$ is determined by the balance of the electrostatic energy and the torsional energy cost of adaptation (see section

2.2).

The basic physical picture emerging from this analysis is shown in figure 1.4 The figure shows that at all distances, (a) the lowest energy is the ideal case,

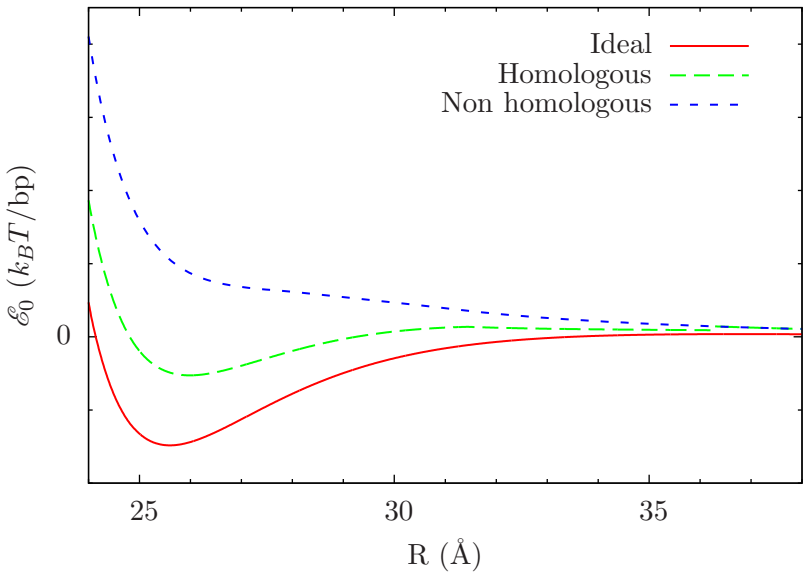


Fig. 1.4: Comparison between the minimum energy of a pair of homologous (), non-homologous (), and ideal () DNA molecules. All the curves were calculated by optimizing all other variables in the energy function, and with $\theta = 0.8$, $f_1 = 0.4$, $f_2 = 0.6$, $\lambda_D = 7.0$ Å. The interaction energy of non-ideal DNA was calculated using equation 2.10, whereas the energy of ideal DNA was calculated using equation (2.3).

(b) non-homologous molecules have higher interaction energy than homologous ones, and (c) at the chosen value of θ, homologous and ideal molecules have an attractive well, and non-homologous ones do not. It is also important to notice that the amplitude of the attractive well is almost double in the case of ideal molecules, compared to the case of homologous molecules.

It was speculated that this effect could stand behind the phenomenon of DNA homology recognition, discussed in section 1.1.4. The DNA molecule in fact was shown to have an innate ability to recognize sequence homology in a protein-free environment (Kornyshev and Wynveen, 2009).

These effects were directly observed by analysing the X-Ray diffraction pictures (Franklin and Gosling, 1953) and comparing them to the results of the theory (Lee et al., 2010; Kornyshev et al., 2011). Indeed, the diffraction patterns show signa-

tures of DNA flexibility and torsional adaptation, which were previously unnoticed in those experimental pictures.

1.6 Skewed molecules

In the previous sections our attention was limited to the case of straight, parallel DNA molecules. We now describe the theory and results of skewed DNA molecules.

The electrostatic interaction energy of two straight, infinitely long skewed DNA molecules was calculated in (Kornyshev and Leikin, 2000), and was later extended to the case of finite-sized molecules (Kornyshev et al., 2002).

We will only give a very short summary of the results. The derivation of the formulas and of the results will not be described in this thesis. Here, we want to present the qualitative results of the theory and discuss its application to describe the case of cholesteric liquid crystals of DNA.

1.6.1 Asymmetric crossings: the chiral torque

As was described in the previous sections, the KL theory takes into account the helical pattern of charges on the surface of the DNA molecules. It was shown that the helical charge motives result in attractive intermolecular forces. Taking into account the helical structure of the charge patterns has also another important consequence. In fact, as can be seen in figure 2.1, it results in an asymmetry in the geometry of DNA-DNA crossovers.

In right-handed DNA crossovers, the negatively charged phosphates align in an optimum geometry with the positively charged grooves in the opposing molecule. For left-handed crossovers the opposite happens: positive and negative charges tend to be perpendicular to each other, so that the electrostatic attractive components will be suppressed.

This qualitative idea was reproduced in the calculation of the electrostatic interaction of infinitely long (Kornyshev and Leikin, 2000) and finite-sized (Kornyshev et al., 2002) DNA molecules. It was found there that the chirality of the charge patterns results in a right-handed torque (see figure 2.1 for the definition of handedness of the crossings). In the case of straight, rigid molecules, as the skew

angle increases, so does the distance between distant portions of the molecules. Therefore, if there is attraction between the molecules, it is unfavourable to have a large skew angle. The optimum crossover angle is determined by the balance of the torque and the attraction. The more attraction there is, the smaller the crossover angle.

1.6.2 The chiral torque and the DNA cholesteric phase

The theory of finite-sized interacting DNA molecules (Kornyshev et al., 2002) was applied to the case of cholesteric liquid crystals of DNA. It was possible to predict the phase behaviour of the system, and the variation of the cholesteric pitch with interaxial distance and salt concentration. The theory correctly reproduced the non-monotonic behaviour of the pitch as a function of the interaxial distance (Stanley et al., 2005).

A recent experimental study on the liquid crystalline phase of DNA showed that the handedness of the cholesteric pitch is left-handed (Zanchetta et al., 2010). The KL theory predicts an opposite handedness, although it cannot be completely ruled out that it may still be valid (Kornyshev, 2010), since the image torque has not been calculated in the theory proposed in (Kornyshev et al., 2002). We notice that the correct handedness is predicted by a theory of Tombolato and Ferrarini (2005), which models the charges as protruding groups on the DNA surface.

1.7 Summary and outlook

In this chapter, we illustrated the basic features of the interactions between DNA molecules in multimolecular assemblies and in dilute conditions. We then explained the foundations of the Kornyshev–Leikin theory of DNA interactions, which gives an explanation for some of the experimentally observed features. In particular, three important effects predicted by the theory were discussed: (a) the attraction between parallel molecules, (b) the difference between homologous and non-homologous interactions, and (c) the asymmetry between right- and left-handed DNA crossovers.

A right-handed DNA crossover is more favourable for the electrostatic interac-

tions. DNA molecules are then predicted to have a tendency to keep this optimum alignment, without losing the favourable attractive energy due to their proximity. As a consequence, the **braid** is expected to be an optimum geometrical structure, which can be realized because of the bending flexibility of the DNA molecule. In the next chapter, we will discuss the development and the results of the extension of the KL theory to the case of braided pairs of DNA molecules.

Chapter 2

Free DNA Braids

In this chapter we will describe our model of two interacting double-stranded DNA molecules in a braided geometry. The development of this theory is a first necessary step to the description of closed loop DNA (chapter 3) and of single-molecule DNA manipulation experiments (chapters 4 and 5). The curved geometry of a molecule in a braid significantly increases the complexity of the mathematical treatment of the problem. It was then necessary to make simplyfing assumptions on the geometry, which nevertheless are suitable for the understanding of the physics that governs such a system.

In the previous chapter we showed that the attraction between DNA molecules, combined with the asymmetry of the DNA crossings, suggested to extend the KL theory to the case of DNA braids. In fact, the braided geometry (see figure 2.1) is such that the molecules stay at the same distance, and keep the local crossover angle constant, throughout the whole length of their juxtaposition. The braided geometry may be achieved at the cost of bending energy. We expect then that the optimum geometry is determined by the balance of these effects. Here, we formulate this idea in a rigorous way, by proposing an energy function that captures all of these interaction components.

We will start by giving a brief introduction on the basic structure of the model in section 2.1. We write an energy function (see equation (2.1)) which requires the calculation of the torque that two DNA molecule experience when coming in contact. The calculation of this term required a very sophisticated mathematical

analysis, and is only sketched in this chapter. The full details of the calculations, for the interested reader, are reported in appendix B. We will then describe in detail the results of the minimization of the free energy of the system in section 2.2, and finally we will compare the predictions of our theory with available experimental data (section 2.3).

2.1 Model

Here, we briefly describe our model for an interacting pair of double-stranded DNAs in a braided conformation. As outlined in the introduction to this chapter, there are three fundamental components of the interaction: (a) the interaction between parallel molecules, (b) the bending energy, and (c) the chiral torque between the molecules. The attraction between parallel molecules was already calculated (Kornyshev and Leikin, 1997, 1999) (and described here in section 1.4), and the bending energy is taken from a very straightforward calculation within the elastic rod theory (see box 2). The new, challenging part is the calculation of the chiral torque. The basic structure of our model, reported in this section, is based upon the original idea of Dr. Eugene Starostin.

The development of our model consists in two parts. In the first part, we derive the interaction energy of a very long, straight braid of *ideal* DNA molecules. The energy function derived in this way is then used in the next step to incorporate the effect of sequence-related and thermal fluctuations. The full details of the derivation of the formulas presented here are given in appendix B.

2.1.1 Geometry assumptions

The calculation of the electrostatic energy of a braided pair of interacting DNAs poses a hard theoretical challenge. In order to make the problem tractable, we developed a model of a braid which makes the following simplifying assumptions:

- **Straight braid**: the braid axis is straight.
- **Symmetric braid**: the two molecular centrelines can be obtained one from the other via a 180° rotation around the braid axis.

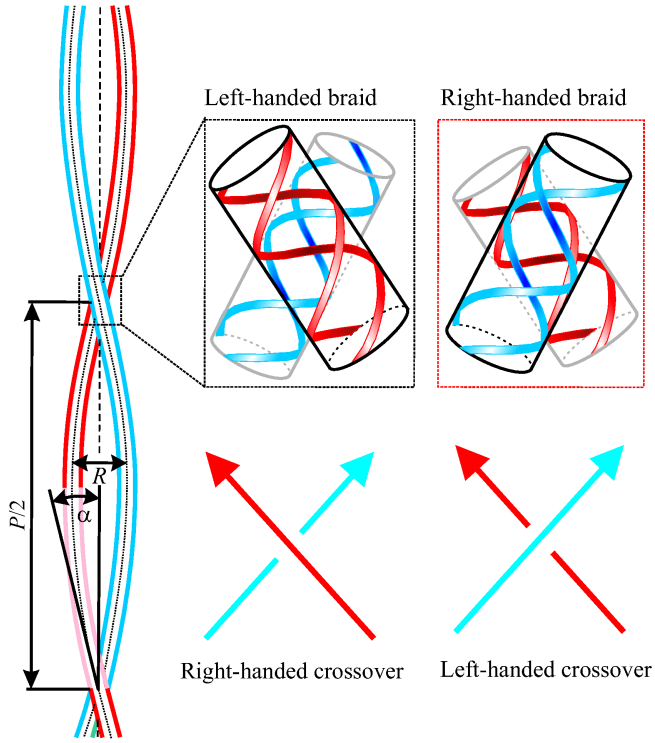


Fig. 2.1: Illustration of our model of free DNA braids. On the left, a pictorial representation of a braid of double-stranded DNA molecules, where the braid diameter R, and the braid tilt angle α, are defined. The upper right part shows the geometry of two DNA crossings of opposite handedness, and below we show the mathematical definition of the sign of the crossings.

- **Homogeneous braid**: the distance between the two molecules, and the local tilt angle, are constant throughout the whole length of the molecule.
- **Infinite braid**: the length of the braid is much longer than any relevant length scale of the problem. This allows to calculate the energy *density* in the braid.

The geometry is illustrated in figure 2.1, where also the quantities that characterize the system are illustrated. These are:

- R is the interaxial distance between the two DNA centrelines.
- α is the tilt angle that the two DNA molecules form with the braid axis.

- $\Delta\Phi$ is the average azimuthal angle between the centre of the minor grooves of the two molecules.

2.1.2 Energy function

Under the assumptions on the geometry listed above, we can write the following function for the energy per unit length function of an infinitely long braid:

$$\frac{\mathscr{E}_b(R,\Delta\Phi,\alpha)}{k_BT} = \frac{\mathscr{E}_0(R,\Delta\Phi)}{k_BT} + \alpha\frac{\mathscr{E}_1(R,\Delta\Phi)}{k_BT} + l_p^b\frac{4\alpha^4}{R^2}. \tag{2.1}$$

Here, k_B is the Boltzmann constant, T is the temperature, l_p^b is the DNA bending persistence length (see box 8). The last term in equation (2.1) is the DNA bending energy, when it is coiled in a superhelical geometry. A derivation of that term is reported in box 2. Given equation 2.1, we can immediately write a solution for the optimum value of α, by taking the derivative of equation (2.1) with respect to α, and setting it to zero:

$$\alpha_{min} = -\left(\frac{\mathscr{E}_1R^2}{16l_p^b}\right)^{1/3}. \tag{2.2}$$

This equation has a very straightforward interpretation. The optimum angle is determined by the balance of the chiral torque, described by $\mathscr{E}_1$, and the bending energy cost, which is proportional to l_p^b. If the chiral torque increases, so does the angle α. The optimum angle is inversely proportional to the bending persistence length. Hence we see that this simple energy function captures the physics that we want to describe.

The energy function given in equation (2.1) is given within one fundamental approximation: the tilt angle α must be small. In fact, the electrostatic interaction in a braid is expanded in the first power of α. The calculation of higher-order terms is extremely complicated, and has not been performed. We rely on the fact that the DNA bending persistence length l_p^b is very long (about 50 nm in physiological salt conditions (Hagerman, 1988)), so that the high bending energy cost will keep the angle α small enough to justify the linear expansion.

Box 2: Bending energy of a coiled DNA molecule

The bending energy of a DNA molecule given in equation (2.1) is calculated within the theory of elastic thin rods, explained in box 8. Such theory assumes that the rod has a cross-section diameter which is much smaller than the rod length. Within such approximation, the bending energy is of a curve described by a curve $\mathbf{r}(s)$ is given by

$$E_{bend} = \frac{l_p^b}{2}\int_0^L \mathrm{d}s\kappa_c^2(s). \tag{b.2.5}$$

Here, L is the length of the curve, and $\kappa_c = \mathrm{d}^2\mathbf{r}/\mathrm{d}s^2$ is the curvature . For a helix, the curvature is given by $\kappa_c = 2\sin\alpha/R$ (see appendix A). Therefore, using equation b.2.5, we obtain

$$E_{bend} = 2l_p^b\frac{\sin^4\alpha}{R^2}L. \tag{b.2.6}$$

If we expand equation b.2.6 into small α, take into account that there are two DNA molecules in a braid, and take the energy per unit length, we arrive at the expression in equation (2.1).

2.1.3 Calculation strategy

We attempt to calculate the electrostatic interaction energy of the system using the framework described earlier in section 1.2. In the case of a braid, the major difficulty lies in the fact that the dielectric interfaces are curved. Therefore, the calculation of the electrostatic image forces is very difficult. We can, however, still find an approximate solution by using the small-α approximation. Within such approximation, it is possible to think of the curved cylindrical cores of the DNA molecules as being locally straight (the curvature is in fact a second-order term in α, see appendix A). Such approximation simplifies the problem tremendously, and ultimately makes it possible to perform the calculation analytically.

The goal is to calculate the energy of interaction using equation (1.14). We start by calculating the potential of a point charge on the surface of one of the DNA molecules in the braid. Such calculation will be performed by assuming that the dielectric interfaces of both DNA molecules are *straight* cylinders, tilted one with respect to the other. The idea behind this approximation is that if the Debye screening length λ_D is much shorter than the superhelical pitch of the braid, then

a point charge on the surface of one molecule has a very small interaction with those that are located in distant portions of the molecules. The curvature can therefore be neglected.

Once the potential of a point charge is obtained, then it can be summed over all the charges in the braid, and finally the interaction energy of the pair can be calculated. The details of the calculation, as well as other important assumptions, are fully described in appendix B.

2.1.4 Sequence-related and thermal fluctuations

As was described in section 1.5, DNA molecules are not ideal, and the real structure of the charge distributions deviates from the structure of an ideal helical line. As a result, after a characteristic length scale (the helical coherence length, see section 1.5.1) the helical phase of the positive and negative charges of opposing molecules falls out of alignment. The mismatch then accumulates and results in a loss of the attractive force between the DNA molecules. However, because of the torsional flexibility of DNA, it can adapt its structure to keep the charges aligned, and benefit from the favourable attractive component of the interaction, at the cost of torsional energy.

To include these effects in the energy function of a braid, we used the same approach described in (Lee et al., 2010). That is, we write a partition function associated with sequence-related and thermal fluctuations, without, however, introducing any fluctuations of the molecular centrelines. This assumption is valid in cases in which the entropic component related to the undulations is significantly suppressed because of tension applied at the molecular tips, or because of a significant rigidity induced by the electrostatic interactions. The latter should be the case for us, as the model predicts a very signficant energy contribution from the chiral electrostatic term.

As a result, we obtain an expression which is very similar to the one that was obtained in the original paper (Lee et al., 2010). The idea to incorporate DNA non-ideality in DNA braids in this way was of Dr. Sergey Leikin.

2.1.5 Energy and free energy functions

After the calculation is completed, the following expressions are obtained for the electrostatic energy per unit length of the braid:

$$\mathscr{E}_0(R,\Delta\Phi) = \frac{2l_B}{l_c^2}\sum_{n=-\infty}^{\infty}\frac{\zeta_n^2}{(\kappa_n a)^2 K_n'(\kappa_n a)^2}[(-1)^n\cos(n\Delta\Phi)K_0(\kappa_n R)+ \\ +\Omega_{n,n}(\kappa_n R,\kappa_n a)] \quad (2.3)$$

$$\mathscr{E}_1(R,\Delta\Phi) = ag\frac{4l_B}{l_c^2}\sum_{n=-\infty}^{\infty}\frac{\zeta_n^2 n}{(\kappa_n a)^3\left[K_n'(\kappa_n a)\right]^2}[(-1)^n\cos(n\Delta\Phi)nK_1(\kappa_n R)+ \\ +\tilde{\Omega}_{n,n}(\kappa_n R,\kappa_n a)] \quad (2.4)$$

where

$$\Omega_{n,n}(x,y) = -\sum_{j=-\infty}^{\infty}\left[K_{n-j}(x)\right]^2\frac{I_j'(y)}{K_j'(y)} \quad (2.5)$$

$$\tilde{\Omega}_{n,n}(x,y) = \sum_{j=-\infty}^{\infty}\frac{K_{n+j+1}(x)K_{n+j}(x)}{\left[K_j'(y)\right]^2}\left[1+2jI_j'(y)K_j'(y)+\frac{j^2}{y^2}\right], \quad (2.6)$$

$$\kappa_n = \sqrt{g^2+\kappa^2 n^2}, \quad (2.7)$$

and

$$\zeta_n = \left[f_1+(-1)^n f_2+(1-f_1-f_2)\,\delta_{n,0}\right]\theta - \cos\left(n\tilde{\phi}_s\right). \quad (2.8)$$

Here, a is the DNA radius, taken to be 11.4 Å (Lee et al., 2010), and $\tilde{\phi}_s$ is the half-width of the DNA minor groove, taken to be ≈ 0.4 rad. Also, $I_n(x)$, $K_n(x)$, $I_n'(x)$ and $K_n'(x)$ are the modified Bessel functions and their derivatives, respectively. Numerical evaluation of the image torque (the term proportional to $\tilde{\Omega}$) shows that it is negligibly small. We neglect its contribution throughout the rest of this thesis.

The basic form of the energy function given in equation (2.1) is then used to calculate the *free energy* of the system. The detailed derivation of the following results is reported in appendix B.2. Despite the considerable technical difficulty in the calculation, the results are remarkably simple. To pass from the electrostatic

energy of ideal DNA molecules to the free energy of torsionally adapting non-ideal DNA, it suffices to substitute

$$\cos\left(n\Delta\Phi(s)\right) \longrightarrow \cos\left(n\overline{\Delta\Phi}\right) e^{-\frac{n^2\lambda_h^\star}{2\lambda_c}}, \tag{2.9}$$

where we introduced the variable $\lambda_h^\star$, which represents the length scale at which the optimum phase alignement is restored, driven by attractive electrostatic forces (see section 1.5.2). Also, the *constant* phase difference $\Delta\Phi$ which appears in equations (2.3) and (2.4) here is substituted by the *average* one, $\overline{\Delta\Phi}$. The final expressions for the free energy of a DNA braid are:

$$\mathscr{F}_0(R,\overline{\Delta\Phi},\lambda_h^\star) = \frac{2l_B}{l_c}\sum_{n=-\infty}^{\infty}\frac{\zeta_n^2}{(\kappa_n a)^2 K_n'(\kappa_n a)^2}[(-1)^n\cos(n\overline{\Delta\Phi})e^{-\frac{n^2\lambda_h^\star}{2\lambda_c}}K_0(\kappa_n R)+ \\ + \Omega(\kappa_n a,\kappa_n R)] \tag{2.10}$$

$$\mathscr{F}_1(R,\overline{\Delta\Phi},\lambda_h^\star) = ag\frac{4l_B}{l_c}\sum_{n=-\infty}^{\infty}\frac{\zeta_n^2}{(\kappa_n a)^3 K_n'(\kappa_n a)^2}[(-1)^n\cos(n\overline{\Delta\Phi})e^{-\frac{n^2\lambda_h^\star}{2\lambda_c}}K_1(\kappa_n R)] \tag{2.11}$$

The energy cost per unit interaction length associated with the torsional adaptation is given by (derivation reported in appendix B.2):

$$\mathscr{F}_{adapt}(\lambda_h^\star) = \frac{(\lambda_c + l_p^h)^2}{16\lambda_h^\star\lambda_c l_p^h}. \tag{2.12}$$

This term is inversely proportional to $\lambda_h^\star$, because shorter values of this variable are a sign of high adaptation, which comes at a higher energetic cost.

Summing all the terms reported so far, we obtain the following functional form for the total free energy density of a braid:

$$\frac{\mathscr{F}_b(R,\overline{\Delta\Phi},\alpha,\lambda_h^\star)}{k_BT} = \\ = \frac{\mathscr{F}_0(R,\overline{\Delta\Phi},\lambda_h^\star)}{k_BT} + \alpha\frac{\mathscr{F}_1(R,\overline{\Delta\Phi},\lambda_h^\star)}{k_BT} + \frac{\mathscr{F}_{adapt}(\lambda_h^\star)}{k_BT} + \frac{\mathscr{E}_{bend}(R,\alpha)}{k_BT}, \tag{2.13}$$

where $\mathscr{E}_{bend} = E_{bend}/L$ is the bending energy (see equation (b.2.5)) per unit length.

2.1.6 Model limitations

The limitations of our model arise from different sources, and can be classified broadly as (a) limitations due to the geometry, (b) limitations due to the inaccuracies of the model of DNA and (c) the ones due to the inaccuracies of the model of the solvent and counterions. We discuss these separately.

In deriving the free energy of a braided pair of DNA molecules, we *imposed* the geometry and calculated the interaction energy. In reality, the geometry will certainly be different from the idealized structure that we studied so far. We can imagine that:

- The braid axis is not straight. The assumption of a straight braid axis limits our analysis to the cases in which significant tension is applied at the ends of the molecule (but without causing topological constraints on the system) and the case in which the electrostatic interactions are strong enough to induce a very significant stiffness in the system.

- The braid structure is likely not to be homogeneous (R and α are not constant). By assuming a homogeneous braid, we limit our attention to a very restricted sets of conformations. The primary objective of our analysis is to study the qualitative physics of the problem, and stimulating new ideas to propose new experimental work.

- Real braids are likely not to be symmetric. We expect this effect not to be very important, because symmetric braids were found to be in a lower energy state than the asymmetric ones (D.J. Lee, unpublished calculations), unless there is a significant difference in length between the two molecules.

- Since the geometry of the interacting molecules is imposed, no entropic cost of the pair formation was included in the calculation.

The model of DNA that we used relies on the assumption that the counterions are located at the dielectric interface of DNA, at the centre of the DNA grooves. Such approximation was already discussed in some detail in section 1.3.2. Here, we want to emphasize that the predictions of the theory of free braids are robust with respect to the variation in the composition of the counterion distribution between

the major and minor grooves (see figure 2.7). On the other hand, this theory does not give a prediction of the values and variations of the parameters θ, f_1 and f_2. We may expect that the amount of counterions located at the DNA surface may change in response to the proximity of another DNA molecule. The pairing ability of the DNA molecule may then be determined by a complex balance between the entropic cost of adding the ions at the DNA surface and the enthalpic benefit of optimal alignment of charges. The physics arising from these additional effects is neglected in our model.

The assumption of bulk dielectric constant for water is at least questionable in case of small separation between the DNA surfaces. As we shall see in section 2.2, we predict that the optimum interaxial separation of a pair of braided DNAs is of the order of ≈ 24 Å. Such distance corresponds to about 1-2 layers of water in between the molecules. It is known (see, e.g., (Bopp et al., 1996)) that water has a dielectric constant of about 5 at such short separations. However, we will see that the basic qualitative features of the model do not change if we impose the interaxial distance to be up to ≈ 27 Å (see figure 2.3). Furthermore, at these distances it is likely that the hydration shells around DNA and around the counterions in solution might play an important role. Such effects are not included in our theory, for the sake of simplicity, and to have a basic model of the interaction of a braided pair of DNA molecules.

2.2 Results

In this section we describe the results of the analysis and minimization of the free energy function that we presented in the previous section (equation (2.13)). First we will illustrate the general features of the free energy landscape (section 2.2.2), and present the results of the minimization as a function of the values of the charge compensation parameters. We give an overview of the effect of changing the electrolyte concentration in section 2.2.3, and finally we will mention some other effects in section 2.2.4.

The infinite sums involved in the expressions of the electrostatic energy density were evaluated using the Levin $u-$transform method (Levin, 1973), to improve speed and accuracy of the computation. The numerical minimization of the energy

function was performed using a Nelder–Mead simplex method (Nelder and Mead, 1965). The code was written using the GNU Scientific Library implementation of these methods (Galassi et al., 2009).

2.2.1 Analysis of the model

The energy function written in equation (2.13) is minimized with respect to the four variational parameters R, $\overline{\Delta\Phi}$, α and $\lambda_h^\star$. To do this, we need to solve the equation

$$\nabla \mathscr{F} = 0, \tag{2.14}$$

where $\nabla \equiv (\partial/\partial R, \partial/\partial\overline{\Delta\Phi}, \partial/\partial\alpha, \partial/\partial\lambda_h^\star)$, so that equation (2.14) actually corresponds to four equations. Out of these four equations, the one for α can be solved easily, and the solution has already been given in equation (2.2).

We can also find an approximate solution for the optimum $\overline{\Delta\Phi}$. In fact, the only terms in the energy that depend upon $\overline{\Delta\Phi}$ are in the electrostatic free energy function. We follow here a procedure that is very similar to the one described in (Kornyshev and Leikin, 1997). The electrostatic term is expressed as an infinite series, which can however be truncated to indices $n \leq 3$ yielding accurate results. It is worth noticing that here we must use indices up to $n = 3$, since at $R \approx 24$ Å these terms are still important. Such approximation allows for an analytical solution of the equation for the optimum $\overline{\Delta\Phi}$. In fact:

$$\frac{\partial \mathscr{F}}{\partial \overline{\Delta\Phi}} \approx \frac{\partial}{\partial \overline{\Delta\Phi}} \sum_{n=0}^{3} (-1)^n \cos\left(n\overline{\Delta\Phi}\right) A_n(R, \alpha, \lambda_h^\star) = 0, \tag{2.15}$$

where the coefficients A_n are to be expressed via equations (2.10) and (2.11):

$$A_n(R, \alpha, \lambda_h^\star) = \frac{\zeta_n^2 e^{-\frac{n^2\lambda_h^\star}{2\lambda_c}}}{(\kappa_n a)^2 \left[K_n'(\kappa_n a)\right]^2} \left[K_0(\kappa_n R) + 2\alpha a g n^2 \frac{K_1(\kappa_n R)}{\kappa_n a}\right] \tag{2.16}$$

We calculate the derivative, use the formulas for the sine of the double and triple

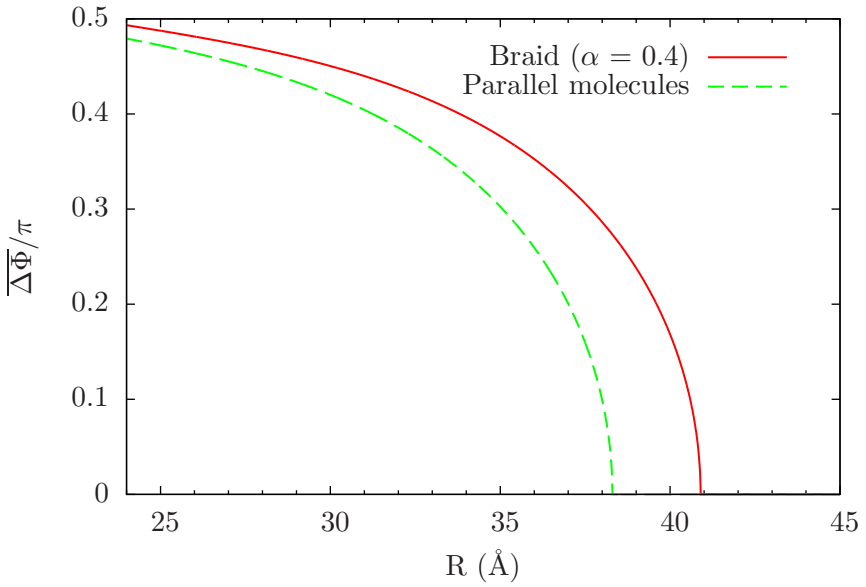


Fig. 2.2: Optimum azimuthal orientation $\overline{\Delta\Phi}$, for a braid with $\alpha = 0.4$, and parallel molecules. The figure was obtained with the values of $f_1 = 0.4$, $f_2 = 0.6$, $\theta = 0.8$, $\lambda_D = 7$ Å, $\lambda_c = 350$ Å, $\lambda_h^\star = 6$ Å.

of an angle, and solve the resulting second-order equation in $\cos\overline{\Delta\Phi}$:

$$\cos\overline{\Delta\Phi}_{min} = \frac{A_1 - \sqrt{A_1^2 - 3A_2(A_0 - 3A_2)}}{6A_2}. \tag{2.17}$$

Figure 2.2 shows the plots of the optimum $\overline{\Delta\Phi}$ as a function of the interaxial separation R, for parallel molecules, and for a braid with fixed α. Braiding is seen to have an effect of extending the range of values of the separation at which the optimum azimuthal orientation is different from zero.

The equations for the optimum values of R and $\lambda_h^\star$ do not have an analytical solution. We must proceed via a numerical minimization procedure in order to find the full energy minimum of the system.

2.2.2 Free energy landscape

Figure 2.3 shows the minimum free energy of an interacting pair of non-homologous DNAs, in physiological salt conditions ($\lambda_D = 7$ Å), with $\theta = 0.75$, which is close to the Manning value (see section 1.3.2). The figure shows that the energy at

$\alpha = 0$ is always positive: the theory predicts a repulsive force at all distances between parallel DNA molecules at this value of θ. Braiding, on the other hand, can provide additional favourable electrostatic energy, resulting in a net attractive force at this value of θ.

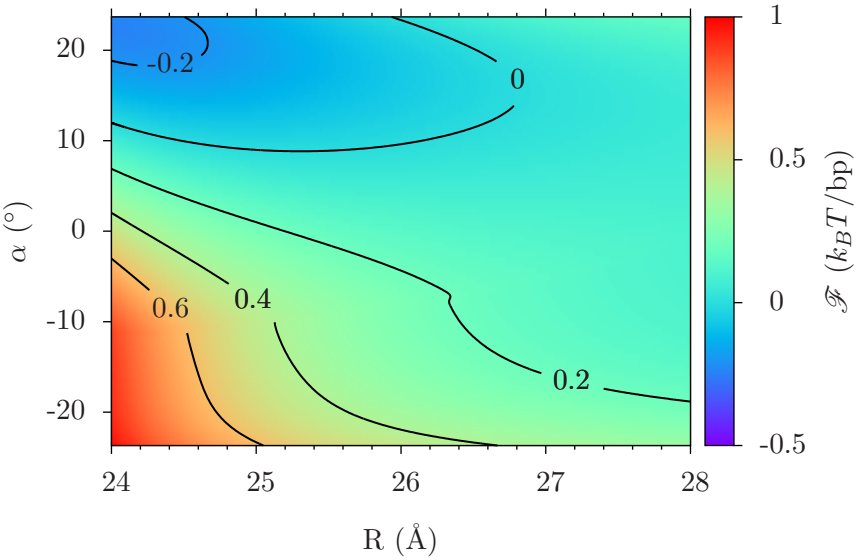


Fig. 2.3: Free energy landscape of a braid of DNA molecules. Values of the free energy were calculated using equation (2.13), at fixed values of R and α, and optimized values of $\overline{\Delta\Phi}$ and $\lambda_h^\star$. Parameters used were $f_1 = 0.4$, $f_2 = 0.6$, $\lambda_D = 7.0$ Å, $\lambda_c = 105$ Å (non-homologous molecules).

In figure 2.3 it is particularly important to notice that the attractive well (negative free energy minimum) is present also at values of R which are higher than the optimum value. This is particularly important because it shows that these results are robust with respect to the optimum interaxial distance. Our theory is in fact likely not to be quantitatively accurate in the prediction of the optimum value of the interaxial distance R, because of a number of reasons mentioned in the model limitations section (see section 2.1.6). This result shows that attraction can occur even if the interaxial distance is ≈ 27 Å.

The enhancement of the electrostatic interactions due to braiding is further demonstrated in figure 2.4. The figure shows the plot of the value of the minimum energy of a braid (red lines) and of a pair of parallel molecules (green lines). It is clear from the figure that braiding is not a small correction to the energy of

parallel molecules, and that it a very significant effect. In braids, attraction can occur at values of θ as low as $\sim 65\%$ for non-homologous molecules, and this value can be even lower for homologous pairs. Parallel molecules, on the other hand, need a much higher degree of charge compensation for pairing to occur ($\sim 80\%$ for non-homologous and $\sim 90\%$ for homologous pairs).

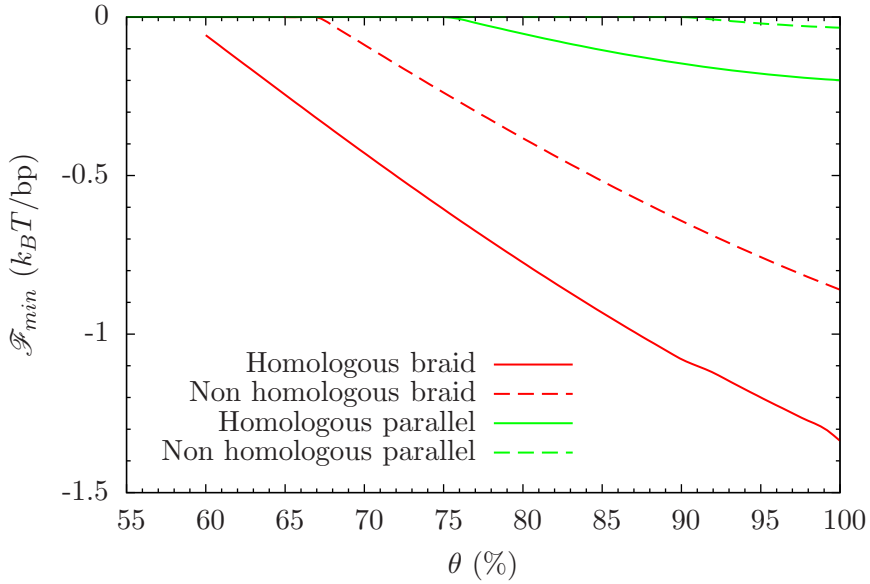


Fig. 2.4: Minimum energy of a pair of braided molecules (red curves) and parallel molecules (green curves), each plotted in the case of homologous (continuous curves) and non-homologous (dashed curves) pairs. The curves were obtained from minimization of equation (2.13). Parameters used were $f_1 = 0.4$, $f_2 = 0.6$, $\lambda_D = 7.0$ Å.

The optimum values of R, α and $\lambda_h^\star$ as a function of θ are shown in figure 2.5. The optimum value of R is around 24 Å, which corresponds to the DNA molecules almost coming into contact (the contact distance is $2a \approx 22.4$ Å), and it is nearly independent of θ, past the attraction threshold. Interestingly, the optimum R is the same within a few tens of Å in homologous and non-homologous molecules. In fact, the equilibrium value of R is for the most part determined by the image force interaction, which is independent of λ_c.

The optimum value of α is always positive, independently of θ. This is perhaps the most important result of this work. The theory predicts that chiral electrostatic interactions favour the formation of a left-handed braid, under all conditions. Figure 2.5 shows also that the value of the optimum α is larger for homologous

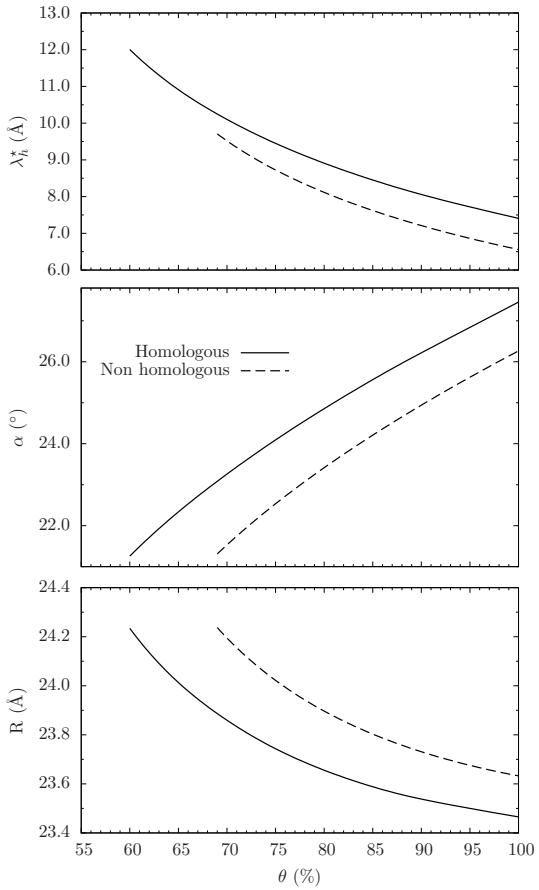


Fig. 2.5: Variables corresponding to the minimum energy of a pair of braided DNA molecules, each plotted in the case of homologous (continuous curves) and non-homologous (dashed curves) pairs. The curves were obtained from minimization of equation (2.13). Parameters used were f_1 = 0.4, f_2 = 0.6, λ_D = 7.0 Å.

than non-homologous molecules. In fact, in homologous pairs the attractive electrostatic components of the interactions are larger in magnitude. From equation (2.2), we can see that the value of the optimum α will increase accordingly.

The optimum $\lambda_h^\star$ decreases monotonically with increasing θ. In fact, as was pointed out in section 1.5.2, $\lambda_h^\star$ is smaller for stronger electrostatic forces. However, for homologous pairs the helical coherence length is longer, so that less torsional adaptation is required. The value of the optimum $\lambda_h^\star$ in therefore higher for non-homologous pairs.

2.2.3 Electrolyte effects

In this section we explore the consequences of changing the value of the Debye screening length. Before proceeding to the results it is worth making a short premise.

Changing the ionic concentration of the counterion species changes the Debye length (see equation (1.6)). However, this is not the only effect which is expected. It is likely that the amount of charge compensation at the DNA surface is also going to be affected. In other words, we should expect λ_D is coupled to θ, f_1 and f_2. In the analysis that follows, this should be kept in mind. The Manning theory (see section 1.2) suggests that the amount of condensed counterions should be relatively independent of the bulk ion concentration. The expected effect is that increasing the ion concentration reduces the amount of condensed counterions.

Figure 2.6 shows the dependence of the minimum free energy on the value of the ionic strength of the solution (see equation (1.6) for the definition of the ionic strength), for different fixed values of θ. The resulting curves show different qualitative trends, which we will describe in detail. It is important to keep in mind that increasing the ionic strength has the effect of decreasing the Debye length, which results in a screening of the electrostatic interactions.

At low values of θ, the minimum energy decreases with increasing ionic strength (see, e.g., $\theta = 0.7$ for non-homologous pairs). In fact, in such cases the dominant effect of increasing the ionic strength is to reduce the amplitude of the net repulsion between the DNA molecules (the term with $n = 0$ of equation (2.3)). Such repulsion is proportional to $(1 - \theta)^2$, so it is important when θ is significantly different from one.

At higher values of θ, the behaviour is more complicated. The net repulsion is weaker, but for relatively low values of the ionic strength, the reduction of the net repulsion still dominates the behaviour of the system, and the energy decreases. Increasing the ionic strength further however has the effect of suppressing also the attractive components of the intermolecular interaction. As a result, past a certain value of the ionic strength, the minimum energy increases again.

We now show how the minimum energy varies as a function of the distribution of counterions in the major and minor DNA grooves. To do so, in figure 2.7 we

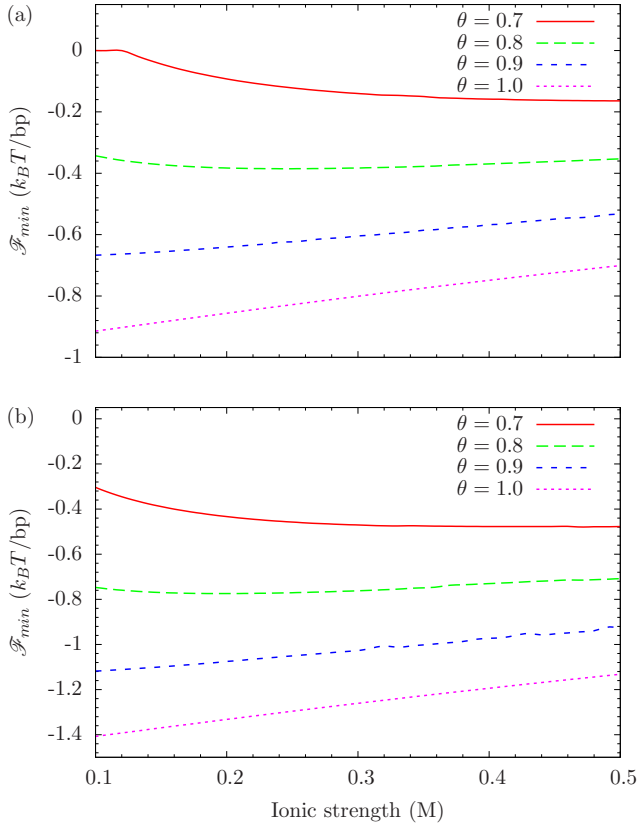


Fig. 2.6: Minimum energy of a braid as a function of the ionic strength, for different fixed values of θ, for non-homologous (a) and homologous (b) molecules. The curves were obtained from minimization of equation (2.13), keeping θ fixed and varying λ_D. The value of λ_D was expressed as a function of the ionic strength through equation (1.6). Parameters used were $f_1 = 0.4$, $f_2 = 0.6$, $\lambda_D = 7.0$ Å.

show a map of the minimum energy as a function of f_1+f_2 and of $f_1/(f_1+f_2)$. The parameter $f_1 + f_2$ represents the fraction of the total amount of adsorbed positive charge which is localized at the centre of the DNA grooves. The remaining fraction, $1-f_1-f_2$, is assumed to be homogeneously smeared on the molecular surface. The parameter $f_1/(f_1 + f_2)$ measures how much, out of the total amount of positive charges, is localized in the *minor* groove. Figure 2.7 shows that a pair of DNA braids is stabilized when most of the positive charges are located at the minor groove, and that the magnitude of the attractive well increases with increasing counterion localization at the DNA grooves. It is important to notice that our results are robust with respect to this parameter $f_1 + f_2$: it is in fact generally

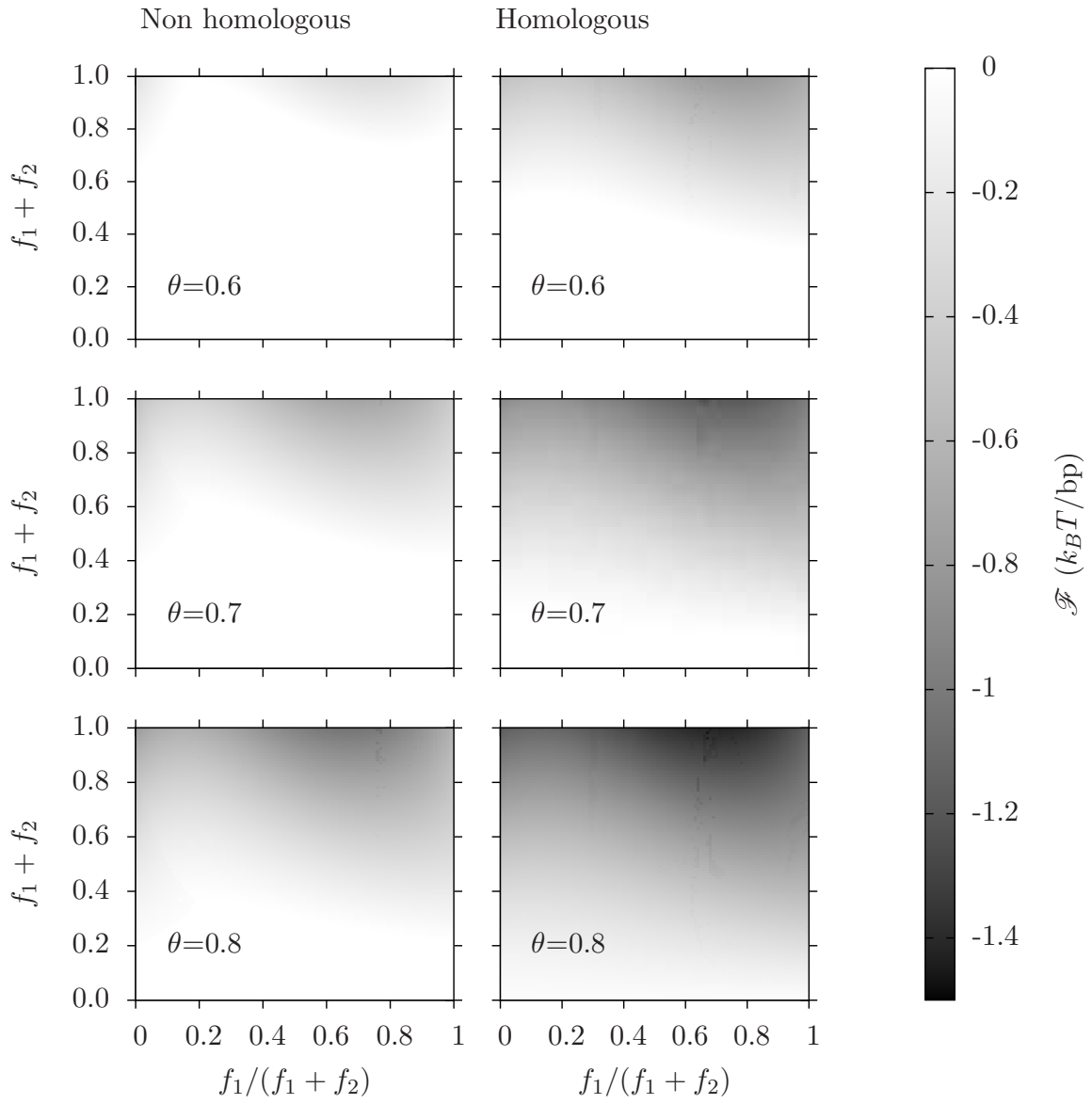


Fig. 2.7: Minimum free energy, as calculated from equation (2.13), as a function of the parameters f_1+f_2 (total amount of positive charge that is located at the centre of the DNA grooves), and $f_1/(f_1+f_2)$ (fraction of f_1+f_2 which is located at the centre of the minor groove. The function was plotted for $\lambda_D = 7$ Å (physiological salt conditions), for homologous and non-homologous pairs.

necessary to have as low as 30-40% of the charge localized at the centre of the grooves, to have sufficient attraction.

2.2.4 The effect of the DNA pitch

Figure 2.8 shows the minimum energy of the free braid system as a function of the DNA helical pitch. It is clear, from the figure, that changing the value of the helical pitch in the range of the values in which B-DNA exists does not affect significantly the value of the minimum energy. Interestingly, the free energy has a minimum at a value of the DNA helical pitch which is close to the value of the natural, unstressed B-DNA molecule.

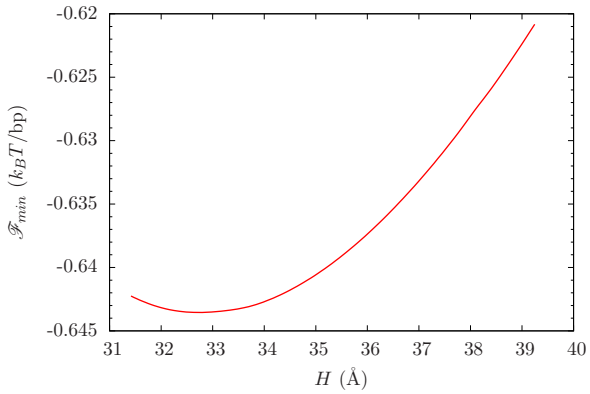


Fig. 2.8: Minimum energy of a pair of braided DNA molecules as a function of the DNA helical pitch. The curve was obtained for a non-homologous pair ($\lambda_c = 105$ Å) at $\theta = 0.9$, $f_1 = 0.4$, $f_2 = 0.6$.

2.3 Discussion

The results presented in the previous section allow for some interesting physical and biological speculation. We relate here the predictions of the theory to existing experimental data. In particular, we show how our theory of braiding may explain the surprising observations reported recently (Danilowicz et al., 2009) of pairing of homologous DNA molecules in monovalent salt. We will then discuss some experimental evidence that we found for the asymmetry in DNA crossovers.

The finding that left-handed braids are more stable than right-handed ones may have important biological consequences. The asymmetry is evident in the case of closed loop DNA supercoiling, which will be described in more detail in chapter 3. Here, we limit our attention to the discussion of experiments of other type.

Direct observation of free braids, to the best of our knowledge, has not been reported in the literature. Single-molecule braiding experiments have been performed (Charvin et al., 2005), but the experiments described there do not show any evidence for asymmetry in the braid handedness, because all of them were performed in low salt conditions.

2.3.1 The Prentiss experiments

In late 2009, it was reported that homologous DNA molecules were able to pair in monovalent salt, without any protein present in the solution. Neither the Kornyshev–Leikin theory nor the Wigner crystal and counterion condensation theories predicted the possibility of DNA condensation or aggregation in the presence of monovalent salt. In fact, all of these theories require that a sufficient amount of DNA charge has to be compensated by positive counterions present in solution. In monovalent salt, the expected amount of charge compensation is not enough to allow for condensation (see section 1.1.3).

In the experiments of Danilowicz et al. (2009), DNA molecules were tethered at one end by a streptavidin molecule, and allowed to pair to a biotin-covered surface of a capillary. In a successive stage, other DNA molecules, homologous and non-homologous to those tethered to the surface, were attached to a paramagnetic bead through a DIG–anti-DIG pair (digoxigenin). Such molecules were added to the capillary and were allowed to equilibrate. Then, a magnetic force was applied in an orthogonal direction to the surface of the capillary. Pairing between DNA molecules can be detected by visualizing the magnetic beads at the "diagnostic distance", which corresponds roughly to the contour length of the DNA molecules used (which were λ-DNA, 48 kbp long).

The results of the experiments were surprising. The number of tethered beads (taken as a measure of the strength of DNA pairing) was higher for homologous

molecules than for non-homologous molecules. This number increased also with the concentration of salt. The pairing was also observed to be temperature-dependent. It increased monotonically up to a temperature of 40°C, then decreased.

In our theory, as was shown in section 2.4, the threshold amount of charge compensation required for pairing of DNA braids is significantly lower than in the case of parallel molecules. This is due to the fact that the optimal alignment of negative and positive charges in DNA braids provides much less stringent conditions for the attractive forces to overcome their negative counterpart. As a result, for homologous pairs the threshold charge compensation is slightly below the Manning value for monovalent salt. For non-homologous molecules, the threshold θ is still below the Manning value, but it is higher than the one for homologous pairs. As discussed in section 2.1.6, the quantitative accuracy of our theory is expected to be limited, but the qualitative predictions are in line with the observations.

As was shown in figure 2.6, at a low value of charge compensation, the minimum energy amplitude (which we assume to be proportional to the number of tethered beads) increases with increasing salt concentration, which is in line with the observed trends. For homologous molecules, the minimum energy amplitude is always higher than the corresponding minimum energy for non-homologous molecules, also in line with the observations. More difficult, within our theory, is to explain the observed dependence of the pairing on temperature. In fact, to explain the decrease in pairing at a certain temperature, it would be necessary to include a full calculation of the entropic cost of pair formation, and of other entropic terms.

The experiments discussed here further support the notion of homology-mediated attraction, which was already found in other studies, discussed earlier (see section 1.1.4).

2.3.2 Evidence for asymmetry in DNA crossovers

Our study demonstrates that the interaction between two DNA crossovers is very different. The same conclusion was drawn on the basis previous theoretical work (Kornyshev and Leikin, 2000), and of all-atom molecular dynamics simulations (Várnai and Timsit, 2010), which found that right-handed DNA crossovers are more stable than left-handed ones, in the presence of magnesium ions. It was also

observed that in DNA crystals the crossovers are right-handed (Timsit and Moras, 1994), and that magnesium ions tend to be located at the N7 and O6 atoms of CG base pairs, "bridging" two DNA molecules.

Perhaps the most evident proof for the existence of a chiral torque between DNA molecule is the existence of the cholesteric phase (see section 1.1.2). The giant cholesteric pitch, observed in aggregates of DNA molecules in physiological salt conditions, demonstrates that the helical structure of DNA molecules must play a role in determining the supramolecular structures of DNA aggregates. However, we must note that the sign of the cholesteric torque observed in experiments is opposite to the one that is predicted by our theory. The geometry of the cholesteric phase (short rigid DNA fragments) is quite different to the case of DNA braids. In particular, it was suggested (Kornyshev, 2010) that the sign of the torque may depend on the sign of the image chiral torque. In the case of DNA braids, such torque was found to be negligibly small (see section 2.1), but this might not be the case for aggregates of short DNA molecules. Calculations in such direction are currently under way.

Chapter 3

Closed circular DNA

In the previous chapter, we showed the results of our model of a pair of free-ended DNA molecules in a braided configuration. It was shown that the magnitude of the predicted effects is large (see figure 2.4). That geometrical construction is somewhat artificial, and it is likely that a braid of two free DNA molecules in solution is not going to form, because of its entropic cost of pair formation. Here, we want to apply the theory to the more relevant case of closed circular DNA supercoiling, where braids of DNA are formed in closed circular DNA molecules *in vivo*, often referred to as *plectonemes*.

In this chapter, we will describe the fascinating biological phenomenon of DNA supercoiling, and illustrate our model of it. In section 3.1 we will introduce the concept of supercoiling, its mathematical and physical description, and its biological relevance. Section 3.2 will be devoted to the description of our model of a closed circular DNA molecule. Next, we will show the results of the minimization of the energy which we obtain. Finally, we will show how our results relate to the existence of tightly wound supercoiled structures (see section 3.1.7), and suggest a possible novel interpretation for well-known experimental facts.

The work presented in this chapter was published in our second research article (Cortini et al., 2012).

3.1 Introduction

Supercoiling is an essential feature of virtually every biologically active DNA molecule. It occurs in many important biological processes: (a) arising because of the torsional stress induced by the double helix unwinding during DNA transcription and replication (Liu and Wang, 1987) and (b) upon closure of circular DNA molecules in bacteria. It is deliberately introduced in bacterial and archaeal plasmids, where the circular genome is found in an interwound configuration (Drlica, 1992). In eukaryotes, DNA is found in the nucleus in a toroidal form, tightly wrapped around histones. Many studies aimed at understanding the role of supercoiling *in vivo*, the enzymes that control it, and the behaviour of closed circular DNA *in vitro*.

Here, we will give a brief overview of our current understanding of the occurence, regulation and role of DNA supercoiling *in vivo* (section 3.1.2), and of what is known from theory and experiment of DNA supercoiling in a protein-free system (section 3.1.5). Before doing that, we will make here a short introduction to the concept of supercoiling, and to its mathematical description.

3.1.1 Mathematical description of supercoiling

The closedness of a DNA molecule gives rise to a coupling between the local twisting deformations of the double helical structure, and the winding of its molecular axis. The precise mathematical formulation of this coupling was given in the work of White (1969), which extended the work of Călugăreanu (1959) (the latter was a general mathematical analysis valid in any number of dimensions). The connection between this formalism and DNA molecules was made by Fuller (1971) and resulted in the well-known formula

$$Lk = Tw + Wr. \tag{3.1}$$

Here, Lk is the linking number of the closed curve, which is a topological invariant, assigned at the moment of closure of the molecule; Tw and Wr are the twist (number of duplex winding around the molecule axis) and the writhe (roughly the number of the superhelical windings) of the molecule. The definitions of these

Box 3: The twist and the writhe of a curve

Consider a curve in space described by the function $\mathbf{r}(s)$, such that $s \in [0, L]$. The **writhe** of the curve is given by the "Gauss integral" of the curve:

$$Wr\{\mathbf{r}(s)\} = \int_0^L \mathrm{d}s_1 \int_0^L \mathrm{d}s_2 \frac{\left(\hat{\mathbf{t}}(s_1) \times \hat{\mathbf{t}}(s_2)\right) \cdot \left(\mathbf{r}(s_1) - \mathbf{r}(s_2)\right)}{\left|\mathbf{r}(s_1) - \mathbf{r}(s_2)\right|^3}. \tag{b.3.7}$$

where $\hat{\mathbf{t}}(s)$ is the unit tangent vector to the curve.
Consider then another curve, which precesses around $\mathbf{r}(s)$, at a constant distance away from it. Such curve may be described by two other vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, which, together with the tangent, form an orthonormal set. The precession of these two vectors around the curve $\mathbf{r}(s)$ is described by

$$\mathbf{\Lambda}(s) = \lambda_1(s)\hat{\mathbf{t}}(s) + \lambda_2(s)\hat{\mathbf{u}}(s) + \lambda_3(s)\hat{\mathbf{v}}(s) \tag{b.3.8}$$

The definition of the **twist** of the precessing curve is then

$$Tw = \frac{1}{2\pi} \int_0^L \mathrm{d}s \lambda_1(s). \tag{3.2}$$

quantities are given in box 3.

Because of the chemical nature of DNA, the linking number can only be integer, not half-integer. It is in fact not possible to attach the $5'$ end to another $5'$ end.

The topological state of closed circular DNA is often described by means of an *intensive* parameter, the **specific linking difference**:

$$\sigma = \frac{Lk - Lk_0}{Lk_0}, \tag{3.3}$$

where Lk_0 is the linking number of a linear, unstressed DNA molecule of the same length. This quantity is useful because it is independent of the DNA length, and describes the amount of elastic stress in a supercoiled molecule. Also, a positive value of σ means that the molecule is **positively supercoiled** (overwound) and negative σ corresponds to a **negatively supercoiled** molecule (underwound). It is found that DNA extracted from bacteria is negatively supercoiled, and $\sigma \approx -0.06$.

3.1.2 Supercoiling in vivo

The study of DNA supercoiling started in 1963 with the discovery by Dulbecco and Vogt (1963) that the DNA extracted from the polyoma virus found *in vivo* was in a closed circular form. Shortly after this study, analysis of the unusual sedimentation coefficient data on the same viral DNA led to the finding that the plasmid was in a negatively supercoiled form (Vinograd et al., 1965). Since then, a considerable effort has been made to elucidate the structural properties, the biological relevance, and the tertiary structure of closed circular DNA (ccDNA).

Supercoiling has different roles in eukaryotes and in prokaryotes, although some common features can be traced (Travers and Muskhelishvili, 2007). Both eukaryotes and prokaryotes possess a chromosomal component in their genome. In all chromosomal DNA, supercoiling arises as a consequence of unwinding the double helix during the transcription and replication processes. Positive supercoils form ahead of the helicase action fork, and compensatory negative supercoils form behind (Cozzarelli and Wang, 1990). In prokaryotes, DNA is also found in the form of **plasmids**, which are short, closed circular DNA molecules.

Plasmids are found generally in a negatively supercoiled state. The energy stored in this high-energy state can be used to help a wide variety of biological processes. It may be used, for example, to assist DNA replication and transcription, which require strand separation. These important processes are also facilitated because negative supercoiling promotes a more underwound double-helical configuration (or even H-DNA (Htun and Dahlberg, 1988), Z-DNA (Kłysik et al., 1981; Singleton et al., 1982), or PX-DNA (Wang et al., 2010)). Also, since supercoiled molecules are bent, processes that bend DNA often require supercoiling (Travers and Muskhelishvili, 2007).

DNA supercoiling is also thought of significantly enhancing the probability of interaction between distant genes in a plasmid (Craigie and Mizuuchi, 1986; Parker and Halford, 1991).

3.1.3 Control of supercoiling: topoisomerases

A special class of enzymes, called **topoisomerases**, control the amount of DNA supercoiling *in vivo*. Their function is multiple. One of the main tasks of these

enzymes is to to remove the topological problems arising during transcription and replication. The other function is to unlink daughter molecules after the replication of circular DNA (Witz and Stasiak, 2010).

Topoisomerases are generally classified into two broad categories: type I and type II, depending on the number of strands that are cut during their action. Further classification is made on the basis of topological conformations of their molecular domains. A detailed classification of all the known topoisomerases can be found in the review by Champoux (2001).

Type I topoisomerases transiently break one sugar-phosphate chain of the DNA molecule, thereby allowing it to find its minimum energetic configuration (actually, the product is a series of topoisomers, due to the fact that there are thermal fluctuations (Pulleyblank et al., 1975)). These class of enzymes operate in a ATP-independent manner, and change the linking number by one unit at a time (Wang, 2002).

Unlike type I topoisomerases, type II topoisomerases require ATP to function. These enzymes transiently cleave two strands of one segment of the molecule (gate segment, or G-segment), and transfer the other segment (the T-segment) through the gate. In the reaction, the linking number of the supercoiled molecule changes by ± 2.

In eukaryotes, where DNA is found wrapped toroidally around the histone complexes, the function of topoisomerases is mainly to relieve positive supercoils ahead of the transcription (or replication) fork, and to unknot the daughter molecules after transcription (Witz and Stasiak, 2010). In prokaryotes, negative supercoiling is actively introduced into plasmids by certain type II topoisomerases, named gyrases.

Some of the type II topoisomerases preferentially act on positive supercoils. It was found, for example, that Topoisomerase IV (Topo IV) has a marked preference to relieve positive supercoils (Charvin et al., 2003). The authors studied the rate at which Topo IV relieved a left-handed and a right-handed braid, by means of single molecule DNA manipulations. The data showed that the rate at which the braid was relaxed to the straight form was substantially higher for a left-handed braid than for a right-handed one (see figure 2.1 for definitions of the handednesses). The mechanism by which an enzyme can locally sense the global topological prop-

erty of a molecule has been subject of intense debate. Several mechanisms were proposed (Charvin et al., 2003; Randall et al., 2006), but then it was established unambiguously (Neuman et al., 2009) that topo IV recognises the local crossover geometry of the molecules. In the study of Neuman et al. (2009), single-molecule DNA manipulation was used to measure also the preferential crossover angle of the enzyme, which was found to be about 85 degrees. This angle is consistent with the optimum angle of a right-handed crossover of the already cited study of Várnai and Timsit (2010).

3.1.4 Hyperthermophiles

There is a notable class of organisms, in which portions of the genome are sometimes found in a *positively* supercoiled state. These organisms are the **hyperthermophylic** bacteria and archea. These organisms can survive at temperature of as high as 110 °C. Forterre et al. (1985) discovered the existence of *reverse gyrase* in the archeobacterium *Sulfobus acidocaldarius*. This topoisomerase has the unique ability to actively introduce positive supercoiling in plasmids. It was found that reverse gyrase is the only protein which is both specific and ubiquitous in the hyperthermophiles (Forterre, 2002).

Why is positively supercoiled DNA found in hyperthermophilic organisms? This question is currently open. The simple idea that the overwound DNA structure, which is associated with positive supercoiling, is more stable with respect to thermal degradation, is not supported by experiments (Marguet and Forterre, 1994). Instead, it was proposed (Timsit et al., 2010) that the higher stability of right-handed crossings, compared to left-handed ones, might prevent excessive thermal loosening of the plasmids. For a review of known and unknown aspects of positive supercoiling, see the recent reviews (Fogg et al., 2009; Valenti et al., 2011).

3.1.5 Structure of naturally occurring supercoils

The law stated in (3.1) allows for two distinct solutions of the topological problem imposed by the circularity, namely the *toroid* and the *plectoneme* (a word derived

from Greek, which means "braided string"). Numerous experimental studies aimed at the determination of which state was found in closed circular DNA in solution.

X-Ray scattering data analysis (Brady and Fein, 1976; Brady et al., 1987) suggested that the tertiary structure of supercoiled DNA was toroidal. On the other hand, dynamic light scattering (Langowski, 1987), neutron scattering in liquid crystals (Torbet and DiCapua, 1989), sedimentation coefficient analysis combined with electron microscopy (Rhoades et al., 1968), and electron microscopy (Sperrazza et al., 1984; Boles et al., 1990) studies suggested an interwound, plectonemic configuration. All these experiments were subject to some controversy, because it could not be ruled out that the preparation of the samples introduced some quantitative or even qualitative modification of the topological state of the molecules. Spengler et al. (1985) performed a study on the formation of knots and catenanes introduced by phage-λ integrases, which revealed that the molecules in solution could adopt the interwound form, but this study suffered the same limitations as the previous. It was only in the later study by Adrian et al. (1990) that the first direct observation of supercoiled plasmids in solution was achieved, by cryo-electron microscopy. This clearly revealed the interwound structure of the plasmids. The superhelical diameter was found to be 12 nm, decreasing to 4 nm (close contact between the segments) upon addition of 10 mM magnesium ions. The structure of the plasmids was found to be branched or unbranched, which is consistent also with later experiments, using scanning force microscopy (Samori et al., 1993), and atomic force microscopy (Lyubchenko and Shlyakhtenko, 1997; Tanigawa and Okada, 1998; Shlyakhtenko et al., 2003).

An early theory for the structure of plectonemic DNA predicted several interesting features of the ccDNA (Camerini-Otero and Felsenfeld, 1978). The results were obtained by a variational analysis of an energy functional that took into account bending and twisting energies, with the constraint of self-avoidance of the molecule. Taking also into account the experimental data of Bauer and Vinograd (1970), the model predicted severe restrictions for the allowed molecular configurations. Namely, for molecules with $\sigma < 0$: i) the number of superhelical turns cannot be greater than 1.2 times σ; ii) σ must be negative or zero; iii) the number of superhelical turns times the pitch is restricted to values between $0.707L$ and $0.86L$ (L is the total contour length of the molecule); iv) the pitch angle is

between 45° and 59°. The experimental results of Boles et al. (1990) are all in striking agreement with these requirements. Monte Carlo simulations also reproduced well the experimental features of DNA supercoiling (Vologodskii et al., 1992; Vologodskii and Cozzarelli, 1994; Hammermann et al., 1998).

3.1.6 Thermodynamics of supercoil formation

The thermodynamic properties of supercoiling were studied shortly after the discovery of the closed circular structure of viral DNA. The intercalative dye method (Bauer and Vinograd, 1970) yielded considerable information on the energetics of supercoiling. The Gibbs free energy of formation of supercoils was studied as a function of the specific linking difference σ, and was found to be of the form

$$\Delta G_\sigma / RT = 0.88\sigma^2 - 0.0038\sigma^3.$$

The small asymmetry was attributed to the properties of the dye intercalation method (see the review by Benham and Mielke (2005) for a detailed discussion).

Several studies aimed at finding the separate entropic and enthalpic contributions to the Gibbs free energy of supercoiling. It was found that the enthalpy contribution dominates, and that the entropy associated with supercoiling is positive. This result was obtained by several methods. Lee et al. (1981) used van't Hoff's method to measure the enthalpy of superhelicity. Seidl and Hinz (1984) used a microcalorimetric assay system, while Bauer and Benham (1993) used the transition competition method. Finally, Duguet (1993) studied the response of the helical repeat of closed DNA at different temperatures, thereby obtaining the temperature dependance of the Gibbs free energy. A very recent study used the isothermal titration calorimetry method (Xu et al., 2012). All methods cited showed that the entropy of supercoiling is positive, in stark contrast with intuition, theories and simulations. Statistical mechanical approaches allowed an estimate of the configurational entropy of supercoiling, all of which yield a negative value (Hearst and Hunt, 1991; Vologodskii et al., 1992; Marko and Siggia, 1994). A discussion on the possible explanation of this discrepancy has been recently reviewed (Benham and Mielke, 2005). The smaller conformational space associated with the supercoiled configuration is not the only entropic component to be considered. The other is

the structuring of water, but theories and simulations are much less precise in its determination. It is clear that further studies are required to have a clearer picture of the thermodynamics of DNA supercoiling.

3.1.7 The tighly wound supercoil enigma

The dependence of the superhelical parameters and structure upon salt type and concentration was studied in several works. Anderson and Bauer (1978) showed that the reduction of the electrostatic repulsion by adding positive ions led to a more tightly wound configuration, and generally increased the amount of positive supercoiling. Tightly wound supercoils (with no detectable space between the strands) were found in a cyro-electron microscopy assay (Bednar et al., 1994) above NaCl concentrations of 100 mM, or in 10 mM magnesium solutions. In atomic force microscopy experiments, the same tightly wound configurations were observed (Lyubchenko and Shlyakhtenko, 1997; Shlyakhtenko et al., 2003).

The structure of the observed collapsed structures could not be explained by their Monte Carlo simulations. In the simulations, only an repulsive DNA-DNA interactions were taken into account, incorporated via an "effective DNA diameter". Marko (1997) proposed a theory of closed circular DNA in which an attractive force between the segments was included, the nature of which was not specified. If such an attractive intersegmental component was included, the theory could reproduce the occurrence of the tightly supercoiled structures. In the electrostatic-undulatory theory of Ubbink and Odijk (1999), all data of Boles et al. (1990) were reproduced, but not those of the collapsed structured mentioned above. The authors studied the possibility that the van der Waals forces could be the source of the attraction, but it was found that they are too weak.

3.2 Model

Our model of a closed circular DNA molecule is based on the results of the theory of free braids that we presented in chapter 2. We consider a very simple model for the geometry of a closed DNA molecule, which is illustrated in figure 3.1. This section illustrates the details of our model. We will first describe the assumptions

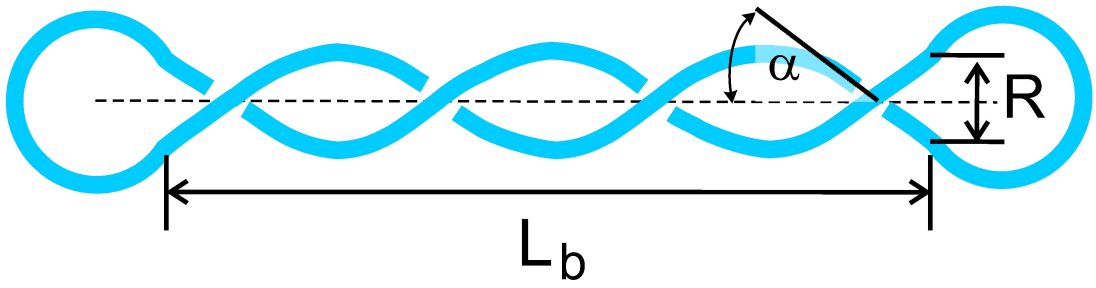


Fig. 3.1: Schematic illustration of our model of a closed circular DNA molecule. The model is characterized by the three variables R, α, and L_b, which are shown in figure.

and approximations that we make concerning the geometry of the system, as well as the consequences of the closedness condition of the molecule (section 3.2.1). The energy function that we use to describe the physical state of the system will then be presented in section 3.2.2. We will then discuss how we built a simple elastic energy function that describes the bending energy of the end loops of the molecule in section 3.2.3.

3.2.1 Geometry and topological constraint

We consider a closed DNA molecule of length L. Upon closure, the topological state is fixed by assigning the specific linking difference σ (see equation (3.3) for the definition). We assume that the molecule is made of (a) a braid and (b) two end loops connected to the ends of the braid. Once again, as in chapter 2, the braid is assumed to be straight, homogeneous, and very long. The axial length of the braid is L_b, so that we have the relationship

$$L = 2L_b + 2L_{loop}, \tag{3.4}$$

where L_{loop} is the length of the end loops, which we assume to be equal to each other.

The closedness of the molecule results in a coupling between the local torsional state of the molecule, and its global supercoiling state, as we explained in section 3.1. To formulate mathematically such coupling, and implement it into the model, we turn to equation (3.1). We need to express the twist and the writhe in terms

of the variables that describe the system.

To model the torsional state of the molecule, we assume that the twist is homogeneously distributed over the entire length of the molecule. Under this assumption, the twist of the molecule is given by

$$Tw = \frac{L}{2\pi} g. \tag{3.5}$$

Since $g = 2\pi/H$, it easy to see that this quantity is equivalent to the number of times that the DNA phosphates precess around the molecular centreline.

The writhe of a pair of helical curves, in the approximation of small tilt angle and very long helix, was calculated in appendix A (see equation (A.14)), and is given by

$$Wr \approx \frac{2\alpha}{\pi R} L_b = 2p, \tag{3.6}$$

where we defined

$$p \equiv \frac{L_b}{P} \approx \frac{\alpha L_b}{\pi R}, \tag{3.7}$$

where P is the superhelical pitch. The quantity p is the number of superhelical turns in the molecule. It is useful because it is connected to the writhe of the braid in a very simple way, which makes the analysis of the model easier.

We now substitute equations (3.5) and (3.6) into equation (3.1). Taking into account equation (3.3), we obtain the following expression:

$$g_\sigma(p) = g_0(1+\sigma) - \frac{4\pi p}{L}, \tag{3.8}$$

where we took into account that

$$Lk_0 = \frac{L}{2\pi} g_0. \tag{3.9}$$

Equation (3.8) is the relationship we were looking for that connects the local twisting deformations to the global writhing state of the molecule.

3.2.2 Energy function

We take into account three different energy components: the bending energy of the end loops, the torsional energy arising from the coupling due to the topological constraint, and the energy of the braided section of the molecule (electrostatic plus bending). Unlike what we did in the case of free DNA braids, we do not take into account here any torsional fluctuations and adaptation effects. In fact, taking them into account in the same way we did in the case of free braids, would have required also the inclusion of the undulations of the molecular axis of the molecule. The torsional state is in fact coupled to the geometry of the molecular axis, and therefore any local fluctuation of the twist has an equal and opposite fluctuation in the writhe. The inclusion of the effect of the undulations of the molecular axis is very difficult and was not attempted here.

The energy function that we write is given by the sum of the bending, torsional and electrostatic energies of the system. The zero energy reference point is taken to be the energy of an unconstrained circular molecule of length L, which is given by

$$E_0 = \frac{2\pi^2 l_p^b}{L}. \tag{3.10}$$

Here, we write the energy function, then we will explain the derivation of the single terms in the remaining part of this section:

$$\begin{aligned} &E_{CL}(R, L_b, p) = 2E_{loop}(R, L_b) + L_b \mathscr{E}_b^{(\sigma)}(R, p) + E_{tors}(p) - E_0 = \\ &= 2l_p^b \frac{\left[2\pi(L/2 - L_b) - \pi R/2\right]^2}{(L/2 - L_b)} + L_b \mathscr{E}_0^{(\sigma)}(R, p) + \pi R p \mathscr{E}_1^{(\sigma)}(R, p) + L\frac{l_p^t}{2}\left(g_0\sigma - \frac{4\pi p}{L}\right)^2 - E_0 \end{aligned} \tag{3.11}$$

The torsional energy of the system is given by the last term on the right-hand side of equation (3.11). Taking into account equation (3.8), and our "ansatz" for the DNA twist given in equation (3.5) it is then easy to see that

$$E_{tors}(p) = L\frac{l_p^t}{2}(g - g_0)^2 = L\frac{l_p^t}{2}\left(g_0\sigma - \frac{4\pi p}{L}\right)^2. \tag{3.12}$$

For the braided section, we use the model that we developed for free-ended

braids which was discussed in the previous chapter. The energy density function given in equation (2.1) is utilized here as a model for the energy density in the braided section of the closed circular DNA. We need however to make some modifications to be able to use this expression.

The most important modification which has to be made is that we have to take into account that we need to use equation (3.8) to express the value of g as a function of the other variables that we chose. This results in an energy density function which depends also on σ, as indicated in the notation of equation (3.11). As we discussed in section 2.2.4, the electrostatic energy density depends weakly on the value of g, but this coupling needs to be taken into account.

In the expression for $\mathscr{E}_b^{(\sigma)}$, we eliminate $\Delta\Phi$ using the expression for its optimum value, reported in equation (2.17). In the energy function the electrostatic term is the only one that depends on $\Delta\Phi$, so that the optimum condition is identical to the case of free braids (see equation (2.17)). Eliminating one variable in this way reduces the computational cost of finding the minimum.

The energy of the braided section can therefore be expressed as

$$E_b(R, L_b, p) = L_b \mathscr{E}_b \left(R, \Delta\Phi = \Delta\Phi_{min}, \alpha = \frac{\pi R p}{L_b} \right) \Big|_{g=g_\sigma(p)} = \\ = L_b \left[\mathscr{E}_0^{(\sigma)}(R,p) + \frac{\pi R p}{L_b} \mathscr{E}_1^{(\sigma)}(R,p) + l_p^b \frac{4\pi^4 p^4}{L_b^4} \right], \quad (3.13)$$

where we defined

$$\mathscr{E}_0^{(\sigma)}(R,p) = \mathscr{E}_0(R, \Delta\Phi_{min}(R))|_{g=g_\sigma(p)} \quad (3.14)$$

$$\mathscr{E}_1^{(\sigma)}(R,p) = \mathscr{E}_1(R, \Delta\Phi_{min}(R))|_{g=g_\sigma(p)} . \quad (3.15)$$

The terms in equation (3.11) are readily obtained from equation (3.13).

3.2.3 End loop energy

In our model of closed circular DNA, the end loops are supposed to contribute to the total energy of the molecule only with their bending energy. The main approximation behind this assumption is that the radius of the end loops has to

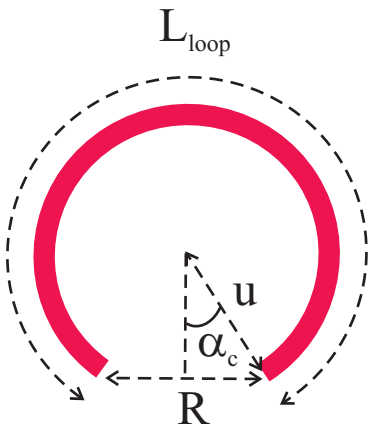


Fig. 3.2: Our model of an end loop. The angle α_c and the loop diameter u are defined in the figure.

be much larger than the Debye screening length.

There are numerous studies in the literature which have addressed the issue of modelling the elastic response of the end loops of closed DNA molecules (Benham, 1977; Tanaka and Takahashi, 1985; Tsuru and Wadati, 1986). Here, we propose a very simple model of an end loop, which has a planar geometry. The DNA braid has ends that point to opposite directions, so that it is not possible for an end loop to be both planar and not break the tangent vector when connecting to the braid. Our estimates indicate however that this is a good approximation (D.J. Lee, unpublished calculations).

Figure 3.2 shows our model of an end loop. The curve is entirely contained within a plane, and the energy of the loop is the bending energy which is calculated within the theory of thin elastic rods (see box 8). It is then very straightforward to calculate the energy of such a system. The curvature of a circle is simply $1/u$, where u is the radius of the circle. From equation (b.2.19) then, the bending energy cost is given by

$$E_{loop} = \frac{l_p^b}{2} \int_0^{L_{loop}} \mathrm{d}s \frac{1}{u^2} = l_p^b \frac{L_{loop}}{u^2}. \tag{3.16}$$

The expression in equation (3.16) can be simplified, taking into account that u and L_{loop} are coupled to each other, through the geometric relationship

$$L_{loop} = (2\pi - 2\alpha_c)\, u, \tag{3.17}$$

where $2\alpha_c$, depicted in figure 3.2, is the angle that the end points of the loop form with the centre of the circle. Such angle is related to the distance between the two loop ends R, through the following equation:

$$u \sin \alpha_c = \frac{R}{2}. \tag{3.18}$$

Eliminating r from equations (3.17) and (3.18), we obtain the following relationship that connects L_{loop} to R and α_c:

$$\sin \alpha_c = \frac{R}{2L_{loop}} \left(2\pi - 2\alpha_c\right). \tag{3.19}$$

Equation (3.19) is a trascendental equation in α_c, that does not allow for an analytical solution. However, we found a very good approximation to the analytical solution, which is based upon a linear interpolation of the exact solution.

Consider the two extreme cases in which $R = 0$ and $\pi R/2 = L_{loop}$. In the first case we must have $\alpha_c = 0$, and in the second one we have $\alpha_c = \pi/2$. It is easy to see that these two cases solve exactly equation (3.19). Then, we interpolate linearly between these two points, and obtain:

$$\alpha_c \approx \frac{\pi}{2} \frac{\pi R}{2L_{loop}}. \tag{3.20}$$

We show the difference between the numerical solution of equation (3.19) and the numerical interpolation given in equation (3.20) in figure 3.3. The data in figure shows that the numerical interpolation is very close to the exact solution (within a few hundredths of radians). The two curves merge at $\pi R/2L_{loop} = 1$.

We then express u in terms of α_c by inserting equation (3.20) into equation (3.17). Then, the resulting expression is used in (3.16) to eliminate u, and finally we obtain

$$E_{loop}(R, L_{loop}) = \frac{l_p^b}{2} \frac{(2\pi - 2\alpha_c)^2}{L_{loop}} \approx \frac{l_p^b}{2} \frac{\left(2\pi - \frac{\pi^2 R}{2L_{loop}}\right)^2}{L_{loop}}. \tag{3.21}$$

An important condition we impose is that $\alpha_c \leq \pi/2$. This assumption has a two-fold reason behind it. The first is to insure that equation (3.19) has only one solution. The second and more important reason is that this imposes a finite

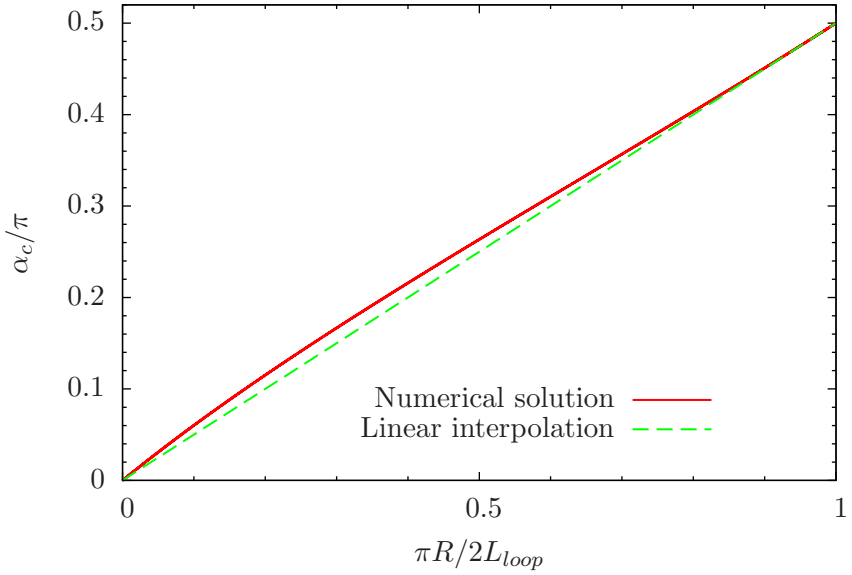


Fig. 3.3: The angle α_c as a function of the ratio $\pi R/2L_{loop}$. The curves were obtained by numerically solving equation (3.19), and using the linear interpolation given in equation (3.20). The numerical solution was performed using a Brent root-finding algorithm (Brent, 1971).

length to the end loop.

We expect that in conditions in which there is a strong electrostatic attraction in the braided, the equilibrium R will be small (as is in the case of free braids), such that $R \ll L_{loop}$ (or, equivalently, $\alpha_c \approx 0$). In such case we may use an approximate expression for the end loop energy:

$$E_{loop}(L_{loop}) = \frac{2\pi^2 l_p^b}{L_{loop}}. \tag{3.22}$$

3.2.4 Nicked molecules

A circular DNA molecule may also have single-strand nicks in some point. When the sugar-phosphate backbone breaks, the molecule may swivel freely around the other strand. This effectively removes the topological constraint, and the molecule may choose any configuration in both the twist and writhe, with no coupling between them.

The energy function of nicked molecules is then simpler, and is very similar to

the case of free braids:

$$E_{nicked}(R, L_b, p) = 2l_p^b \frac{[2\pi(L/2 - L_b) - \pi R/2]^2}{(L/2 - L_b)} + \\ + L_b \mathscr{E}_0(R, \Delta\Phi_{min}(R)) + \pi R p \mathscr{E}_1(R, \Delta\Phi_{min}(R)) + l_p^b \frac{4\pi^4 R^2 p^4}{L_b^3} \quad (3.23)$$

In the remaining part of this chapter, we will drop the explicit dependence of $\mathscr{E}_0$ and $\mathscr{E}_1$ from $\Delta\Phi_{min}$.

3.2.5 Model limitations

Our model of a closed loop DNA, as anticipated, is a simple two-state model that aims at the description of the interplay between electrostatic and elastic forces in closed loop DNA. It has several limitations, which we will briefly discuss here.

- The model inherits the limitations of our theory of free braids, which were discussed in section 2.1.6. In particular, we should mention here that we cannot apply our theory to the case in which the length of the braid is comparable to the Debye length or to the DNA helical pitch. Strictly speaking we should consider our theory to be valid only in the case in which it predicts values of the superhelical pitch that are much larger than any other length scale of the problem (DNA radius, DNA helical pitch, Debye length, bending persistence length).

- We did not take into account any torsional fluctuations in the braid. That is for the reasons that were discussed in the beginning of this section. As a result, we expect that the electrostatic interactions that were calculated here are of much higher magnitude than the one present in real DNA circles (see figure 1.4).

- Our model does not take into account the undulations of the molecular axis of DNA. This is a significant limitation, as molecules with low values of $|\sigma|$ are likely to have a significant amount of undulations present, and including these effects may alter the qualitative picture emerging from our analysis for

those values of σ. Our theory should therefore be applied strictly only to values of σ which are high enough.

- The model of the end loop that we made has disconnected tangents to the braid, which results in an underestimate of its elastic energy cost. However, we estimated this elastic energy cost to be small (D. J. Lee, unpublished calculations).

- The end loop shape that we chose has a planar geometry. Once again, this likely means that the energy cost is underestimated. Our goal however is to describe the situation in which the closed loop is dominated by the attractive electrostatic interactions. In that case, the principal contribution is the energy of the braid, and the end loop have only the role to stop the propagation of the braided section.

3.3 Results

In this section we illustrate the results of the analysis of the model that we described in section 3.2. The first part will be dedicated to the analyis of the equilibrium conditions of the closed circular DNA system. In section 3.3 we describe the results of the numerical minimization of the energy function given in equation (3.11), and discuss the effect of changing the energy parameters. We will present "state diagrams", which help us understand the role of the charge compensation parameters, the molecular length, and the supercoiling topological state.

3.3.1 Model analysis

The energy function given in equation (3.11) does not allow for much analytical treatment. None of the equations for the equilibrium values of the variables have an analytical solution. The system of equations that define the optimum energy of the system are given by

$$\nabla E_{CL} = 0, \tag{3.24}$$

where here $\nabla \equiv (\partial/\partial R, \partial/\partial L_b, \partial/\partial p)$.

We write the equation for the optimum p:

$$\frac{\partial E_{CL}}{\partial p} = 4\pi l_p^t \left(g_0\sigma - \frac{4\pi p}{L} \right) + L_b \frac{\partial \mathscr{E}_0}{\partial p} + \pi R \mathscr{E}_1 + \pi R p \frac{\partial \mathscr{E}_1}{\partial p} + l_p^b \frac{16\pi^4 R^2 p^3}{L_b^3} = 0. \quad (3.25)$$

This equation can be simplified if we neglect the derivative of the electrostatic density functions with respect to p (see figure 2.8). Unfortunately, even doing that, the result is a complete cubic equation in p, which has an analytical solution that is of little use due to its complexity.

The equilibrium equation for L_b reads:

$$\frac{\partial E_{CL}}{\partial L_b} = \mathscr{E}_0 - l_p^b \frac{12\pi^4 R^2 p^4}{L_b^4} - 2l_p^b \frac{\left[2\pi(L/2 - L_b) - \frac{\pi^2 R}{2}\right]\left[\frac{3\pi^2 R}{2} - 2\pi(L/2 - L_b)\right]}{(L/2 - L_b)^4}. \quad (3.26)$$

Again, this equation does not have an analytical solution, due to the fact that it is a fourth-order equation that contains all powers in L_b.

There is one case in which we can perform some analysis, which is the case of nicked molecules. In this case in fact the torsional energy cost vanishes, making the equations simpler. From our physical understanding of the problem, we expect that in the case of nicked molecules there are two different states: the open circle, and the tightly supercoiled state. In fact, the addition of a chiral attractive term in the energy function results in the formation of a writhed state. Unlike the case of topologically constrained molecules, we do not expect any "loosely" supercoiled state. Since there is no loose state (in which R becomes comparable to the length of the end loops), we may then use the approximate formula for the end loop energy (see equation (3.22)). To estimate the optimum length of the braid of nicked closed circles, it is convenient to rewrite the energy function back in terms of α, instead of p. Then, from the optimum condition, we obtain

$$L_{loop}^2 = -\frac{2\pi^2 l_p^b}{\mathscr{E}_b(R,\alpha)} \rightarrow L_b = \frac{L}{2} - \sqrt{\frac{2\pi^2 l_p^b}{|\mathscr{E}_b(R, \Delta\Phi_{min}(R,\alpha), \alpha)|}}, \quad (3.27)$$

which is a valid solution only if $\mathscr{E}_b(R) < 0$, i.e., there is a net attraction in the braid. In the case that the energy density in the braid is positive, the optimum configuration is the open circle.

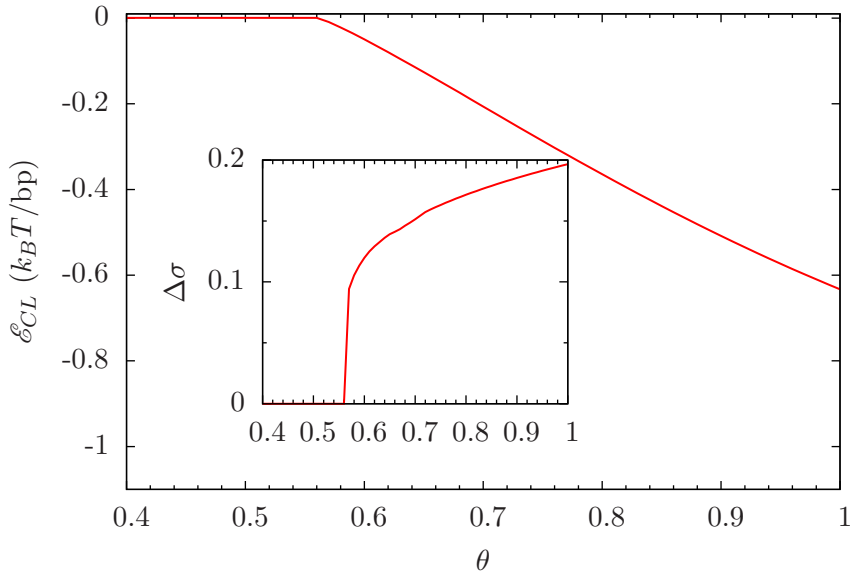


Fig. 3.4: Minimum energy of a nicked 2000 bp-long closed circular DNA molecule, as a function of the charge compensation parameter θ, as calculated by numerical minimization of equation (3.11). The inset shows the corresponding value of the excess linking number that would result upon ligation of such molecule, calculated using equation (3.28).

3.3.2 Nicked molecules energy

In this section we briefly illustrate the energy landscape of a closed circular DNA molecule in the simplest case, which is that of nicked molecules. As explained in section 3.2.4, there is no topological constraint in this case, so the molecule can adopt its optimum configuration without any torsional energy cost. The discussion in this section is limited, because the case of nicked molecules is very similar to the case of free braids.

Figure 3.4 shows the minimum energy of a nicked 2000 bp-long molecule, as a function of θ. The energy minimum as a function of θ has similar behaviour to the one of a braid of ideal DNA molecules. The energy is equal to zero (corresponding to the open circular state), until θ reaches the threshold value of ≈ 55 %, whereupon the molecules undergo a spontaneous transition to a writhed state.

The inset in figure 3.4 shows the value of the excess linking number resulting from the spontaneous writhing of a nicked circle. This can be estimated in the following way. From the definition of the excess linking number (see equation

(3.3)) we may write:

$$\Delta\sigma = \frac{Tw + Wr - Lk_0}{Lk_0} = \frac{Tw + Wr}{Tw_0} - 1 \approx \frac{Wr}{Tw_0}. \tag{3.28}$$

We took into account that $Lk_0 = Tw_0$, which is the twist of a linear DNA molecule in a particular ionic environment. We also took into account that upon writhing the twist of the nicked DNA molecule is roughly equal to the one of the linear molecule, as the electrostatic interactions do not significantly alter the DNA helical pitch (see figure 2.8). The value of the excess linking number resulting from our calculation is large: about $0.1 - 0.2$ for a 2000 bp-long molecule.

3.3.3 Energy landscape

The energy landscape of a closed circular DNA molecule in which chiral attractive forces are present is much more complicated than the case of free DNA braids. In fact, there are two additional parameters to take into account, which are the specific linking difference σ, and the DNA length L. Our model of the electrostatic interactions in a braid relies on the approximation of a very long juxtaposition length, as explained in chapter 2. Therefore, we can apply our theory only to the case of very long molecules. In this section, we will explore the effect of varying σ on the optimum configuration of the molecule.

As we shall see in this section, there are two distinct "phases" of a supercoiled closed circular DNA. One is a "loosely" supercoiled state, and one is a "collapsed" state. In the collapsed state, the dominant energy term is the electrostatic attraction in the braid, which, if strong enough, is able to overcome the energetic cost associated with creating that particular state. As we expect from the theory of free braids, the state that the electrostatic energy favours is the left-handed braid. Due to the topological constraint, however, the molecule must pay an energetic cost to form that state. We shall see how this interplay works in the case of closed circular DNA.

Figure 3.5 shows the variation of the minimum energy of the system, with varying σ, at fixed values of the charge compensation parameter θ. For low values of θ, the minimum energy curve is the "classical" parabolic curve (see section

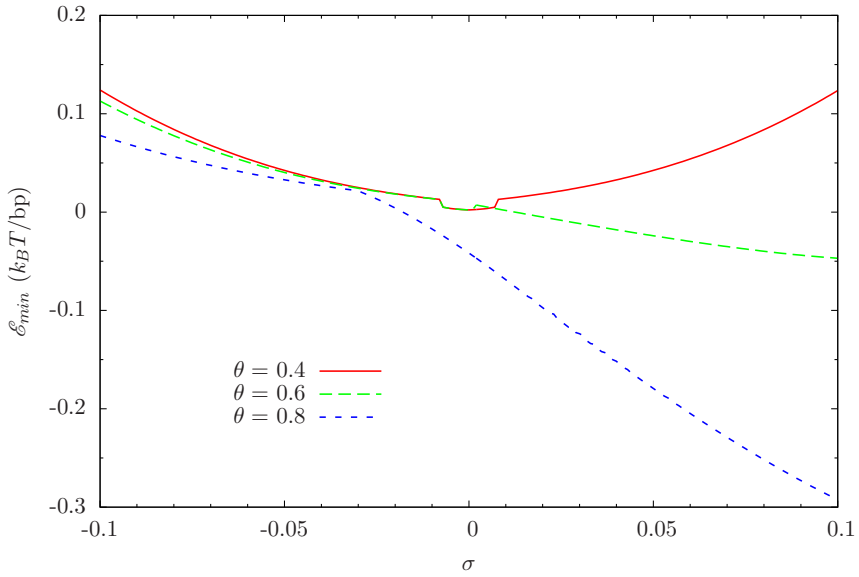


Fig. 3.5: The minimum energy of a closed circular DNA molecule (2000 bp-long), for varying excess linking number σ. The curves were obtained by minimizing equation (3.11), at fixed values of θ, and at $f_1 = 0.4$, $f_2 = 0.6$, $\lambda_D = 7$ Å.

3.1.6). When the inter-DNA interactions are purely repulsive, in fact, there is no impetus for spontaneous braiding, and there is always an energetic penalty associated with supercoiling. On the other hand, if the values of θ are such that electrostatic forces start to be attractive, there will be a propensity for left-handed braiding. As shown in figure 3.5, the curves deviate from the parabolic curve at some value of σ, which can be either positive, close to zero, or negative, depending on the value of θ.

For positive values of σ, the onset of attractive forces, which gives rise to a negative-energy state, is favoured by elastic forces. In fact, in the case of positive σ, elastic forces alone tend to form a left-handed braid. Therefore, there is a strong drive towards the collapsed state, in which there are tight intersegmental interactions. In such state the number of superhelical turns p increases, thereby alleviating the torsional stresses in the molecule. The collapse then may occur at relatively low values of θ (in the conditions chosen for figure 3.5, already at $\theta \approx 0.6$ there is collapse at positive σ).

The case of negative σ is more complicated. In fact, in such case the elastic forces alone form a right-handed braid, which is unfavourable for the chiral elec-

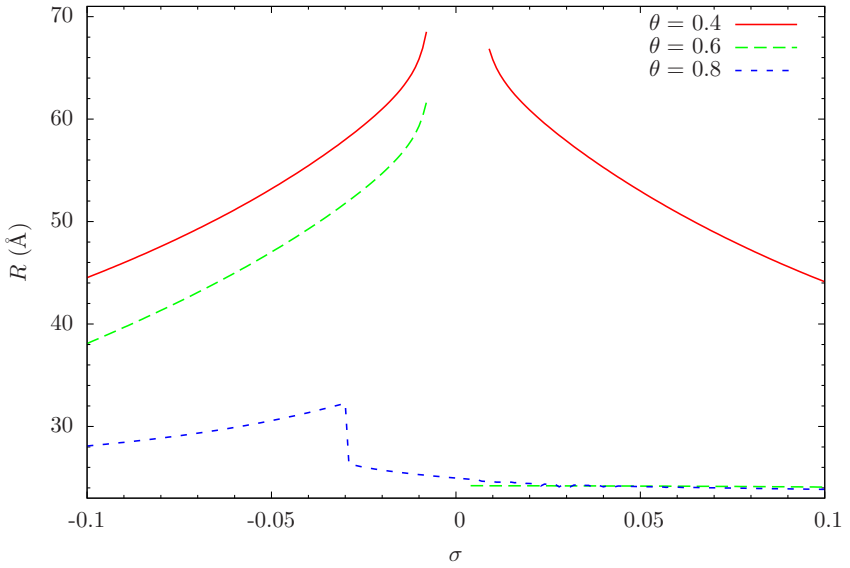


Fig. 3.6: Optimum value of R as a function of the excess linking number σ. The parameters chosen to plot these curves are the same as in figure 3.5. Where the curves are broken it means that the equilibrium radius is $R = L/\pi$, the diameter of an open circle.

trostatic interactions predicted by our theory. As was shown in section 2.2, the right-handed braids are always unfavourable. However, if there is a high value of θ, then there still might be a drive towards a collapsed state. At negative σ, there is a high torsional energy cost associated with the formation of a left-handed braid, due to the topological constraint. Then, only if there is a very strong inter-strand attraction the balance between the investment and the gain in energy favours the collapsed state. Figure 3.5 shows that this may occur at $\theta \approx 0.8$ for the particular values of the parameters chosen there.

Figure 3.6 shows the variation of the optimum R as a function of σ, for fixed values of θ. This figure reiterates the concepts which we just mentioned. It is interesting to notice that in the collapsed state, the optimum value of R stays almost the same, regardless of the values of σ and θ.

For $\sigma \approx 0$, and for low value of θ, the equilibrium configuration is the open circular state. This is shown in figure as a divergence in the optimum radius.

For values of σ and θ that do not result in a collapsed state, the equilibrium radius decreases with increasing $|\sigma|$. That is because higher values of $|\sigma|$ result in higher torsional stresses in the molecule, which the molecule attempts to relieve

by increasing its writhe.

3.3.4 State diagram

Here, we illustrate the state diagram of a closed circular DNA molecule. As mentioned in the previous section, there are three distinct states: the circle (C), the loose supercoil (L), and the tight (collapsed) supercoil (T). Figure 3.7 shows the state diagram as a function of the DNA length and the charge compensation parameter θ, at fixed values of σ. For $\sigma = 0.05$, most of the diagram is occupied by the T state. As explained in section 3.3.3, for valued of θ greater than $\approx 55\%$, the chiral forces readily drive the transition to the collapsed state. Only in a narrow region of values of L, the length of the molecule is too short to favour the highly writhed T state. For molecules shorter than ≈ 600 bp, the equilibrium shape is the circle, regardless of θ. The bending energy cost is in fact too large in that case to favour a writhed state.

The case of $\sigma = 0$ is simpler, because, similarly to the case of nicked molecules, there is no L state. The T state is favoured here only for sufficiently long molecules ($N_{bp} > 400$) and for sufficiently high values of the charge compensation ($\theta > 0.6$). For shorter molecules, a higher value of θ is needed for the T state to occur. Unlike the case of nicked molecules, writhing here is accompanied by a torsional energy cost.

For negatively supercoiled molecules, the access to the T state is more difficult (see section 3.3.3). The T state may occur here only for high values of charge compensation ($\theta > 0.8$), and for longer molecules ($N_{bp} > 500$).

3.4 Discussion

In this chapter, so far, we illustrated the formulation of our model of a closed DNA molecule, and our results for its equilibrium conformation. In this final section, we compare the results of our theory to experimentally known facts. Before proceeding with the analysis of the literature, it is worth reminding the reader about the assumptions of the model, and its limitations.

It should be kept in mind that we made the assumption of a regular, straight

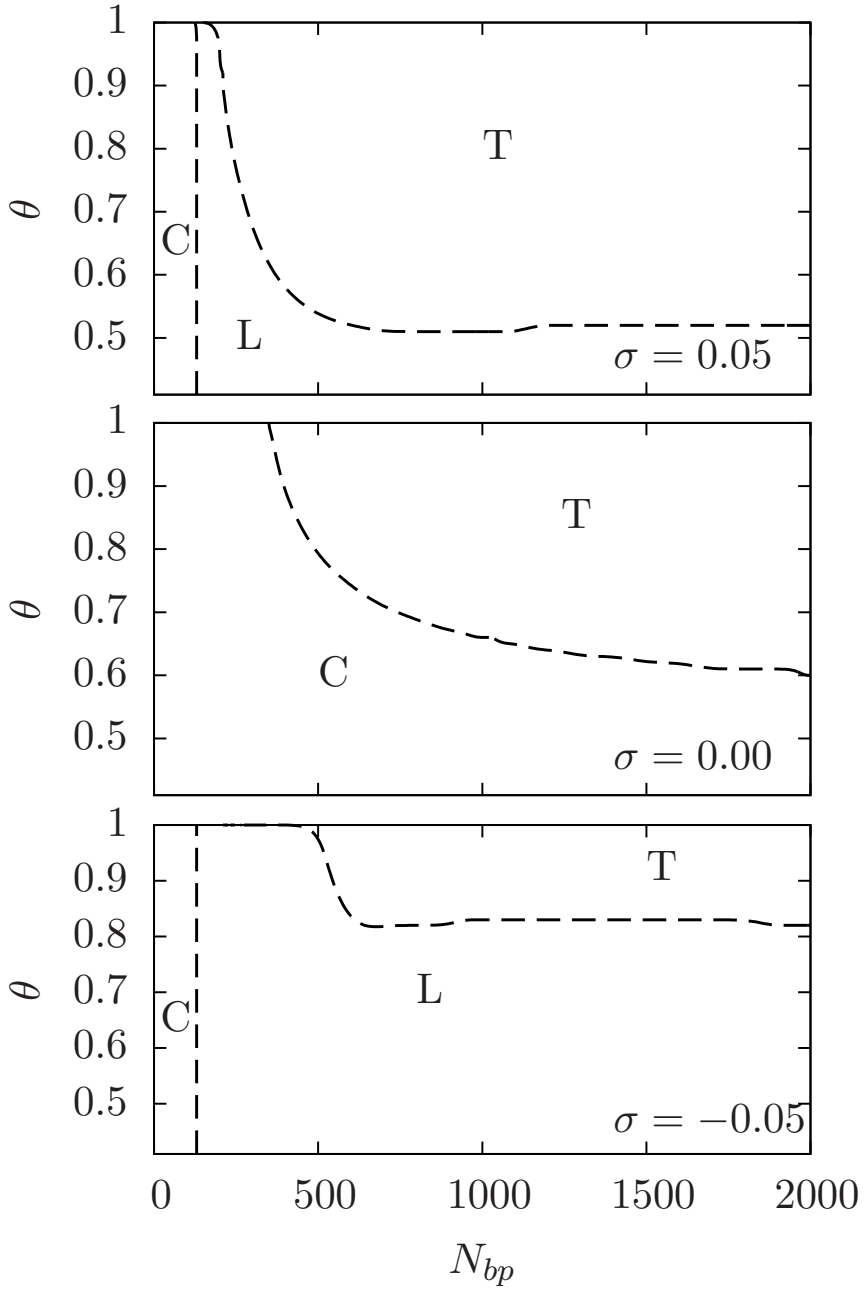


Fig. 3.7: State diagram of a closed circular DNA molecule as modelled by the energy function in equation (3.11). The state diagram is expressed as a function of the number of base pairs N_{bp} and of the charge compensation parameter θ. Three states are given: the open circular state (C), the loosely (L) and tightly (T) wound states. Numerical minimization of the energy function was performed, and the total electrostatic density in the braid was evaluated to establish whether the molecule was found in the tightly or loosely supercoiled state.

geometry of the molecule (see figure 3.1). It is known, e.g., from electron microscopy (Boles et al., 1990; Adrian et al., 1990; Bednar et al., 1994), that the shape of closed circular DNA molecules in solution is not straight. It is, in fact, rather irregular and often branched. The main reason behind our assumption of a straight plectonemic axis is that we want to account for the situation in which there are strong electrostatic forces that drive a transition to a tightly supercoiled state. In such case, we expect that the electrostatic forces will stiffen significantly the braid, resulting in a shape that is not far from the one we are assuming here. In fact, the cryo-electron microscopy images of DNA plasmids in high salt (Adrian et al., 1990; Bednar et al., 1994) do seem to be much straighter than the ones in low salt conditions.

It is also useful to recall that we assumed here that the DNA molecule is *ideal*. This means that the electrostatic interactions in this case are significantly stronger than in the case of non-ideal, torsionally flexible molecules (see section 1.5 for more details). As a result, we expect that the theory presented so far is not quantitatively accurate in terms of (a) the precise values at which the collapse transition occurs and (b) the precise values of the minimum energy and of the equilibrium parameters. Again, the primary focus of this model is to illustrate the consequences of including chiral attractive forces in the energy landscape of a closed circular DNA. We expect that the qualitative conclusions reached will not be altered if DNA non-ideality will be included in the model.

Perhaps the most significant shortcoming of our theory is that we did not include the effect of undulations of the molecular axis. The reason for this is that it would require an extremely difficult analysis of the undulations in the braided geometry. Undulations were taken into account in the KL theory in the case of fibers (Lee et al., 2010), but that approach was not applicable here, because there is no steric confinement term that helps in the analysis. Including the undulations might have a profound effect on the energy landscape described in the results of our theory. It is known from theory (Marko, 1997; Ubbink and Odijk, 1999) and experiments (Gebe et al., 1996), that the undulations are more pronounced for small values of $|\sigma|$, because in that case the system is looser and more able to undulate. Therefore, we expect that our theory may be far off the mark in that case. Even if strong electrostatic interactions are present there, undulations and

entropic effects might completely suppress the propensity for electrostatically stabilized writhes. With all this in mind, we start the comparison of the experiments to our theory.

3.4.1 Enthalpy of DNA supercoiling

Our theory successfully reproduces the parabolic (Bauer and Vinograd, 1970; Lee et al., 1981; Bauer and Benham, 1993; Seidl and Hinz, 1984; Xu et al., 2012) dependence of the enthalpy on σ, when θ is low enough. In this case, the repulsive interactions within the braid make the energy of supercoiling always unfavourable. This is a well known feature of DNA supercoiling, which was already reproduced in Monte Carlo simulations (Vologodskii et al., 1992). The energy stored in the supercoiled state is in fact used by enzymes that unwind DNA to assist their reactions.

Our model predicts that the sign of the enthalpic contribution to DNA supercoiling is negative, in conditions under which there is a high degree of charge compensation on the DNA surface. This fact has never been observed experimentally, and therefore it would be important to perform studies of the thermodynamics of supercoiling in the presence of DNA condensers, or in the presence of sufficient amounts of divalent salts.

3.4.2 The effect of divalent metal ions on the DNA secondary structure

It was observed several times that divalent ions overwind DNA (Anderson and Bauer, 1978; Xu and Bremer, 1997; Rybenkov et al., 1997; Cherny and Jovin, 2001; Vetcher et al., 2010). This conclusion was reached by analysing the mobility of nicked DNA molecules that were treated with divalent ions. A schematic representation of these experiments is depicted in figure 3.8. Negatively supercoiled DNA is extracted from bacteria and treated with topoisomerase I at a given concentration of divalent ions. Then, gel electrophoresis is performed, which shows that a moderate amount of positive supercoiling appears. These experiments were interpreted by assuming that divalent ions overwind DNA, which upon closure

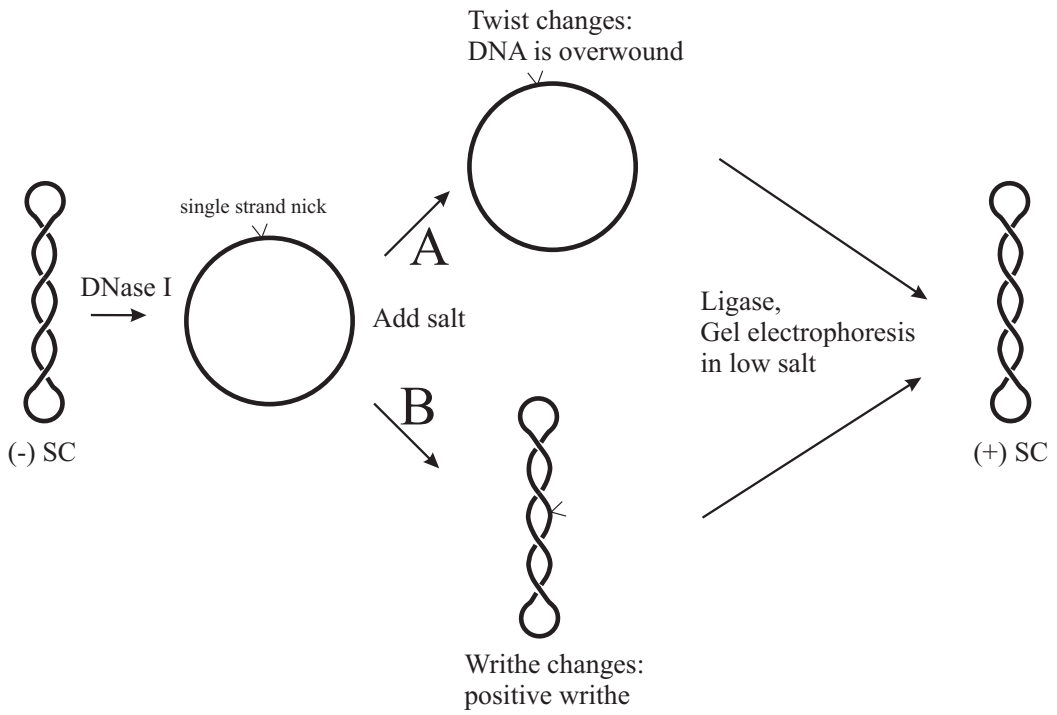


Fig. 3.8: Two different interpretations of the experiments reported in (Xu and Bremer, 1997). In the pathway of "A", the *twist* changes as the molecule is nicked in the presence of divalent ions. When closing it an running it in a gel in low ionic concentrations, positive supercoiling is observed. In our interpretation (path "B"), it is the *writhe* that changes when the molecule is nicked in the presence of divalent cations, which, in the same way, results in the final occurrence of positive supercoiling. Figure taken from (Cortini et al., 2012).

then results in the appearance of positive supercoiling.

These classical experiments can be also interpreted in a different way. In fact, exactly the same effect is expected by our theory of chiral electrostatic interactions. The presence of divalent cations may result in spontaneous positive writhing of the DNA circle, when the topoisomerase acts on it. As a result, after closing the circle, positive supercoiling is expected to arise. The magnitude of this effect may be estimated by our model of nicked DNA molecules (see figure 3.4). As discussed in section 3.3.2, the excess amount of σ predicted by our model is of the order of 10-20%, which is about ten times larger than the experimentally observed values (Rybenkov et al., 1997). The discrepancy may be due to the overestimation of the electrostatic energy and to the absence of undulations in our model.

Distinguishing between these two interpretations of the same experiments requires further investigation. We suggest that single-molecule studies of the DNA twist are ideal to put these effects to the test. For example, the method used by Gore et al. (2006) to measure the twist of a single DNA molecule would probably be suitable for the purpose. In closed circular molecules it is very difficult to distinguish between an effect on the secondary and on the tertiary structure, because they are connected together through the topological constraint.

Supercoiled DNA plasmids were also studied by electron microscopy in the presence of $MnCl_2$ (Ma and Bloomfield, 1994). The experimental observations indicated that the presence of manganese resulted in the formation of toroidal-shaped condensates in negatively supercoiled plasmids, but did not have any effect on linearized molecules. This effect was not caused by the presence of Na^+ or Mg^{2+}, nor did these two ions negate the effect of Mn^{2+}. These observations indicate that the effect reported there was not of electrostatic nature.

3.4.3 Tight supercoiling and writhing of relaxed circular DNA

In classical electron microscopy studies (Boles et al., 1990), supercoiled DNA was found to be quite irregular in shape. As mentioned earlier, this is a feature that our model does not take into account. However, the behaviour of supercoiled DNA was well accounted for by Monte Carlo simulation techniques (Vologodskii

and Cozzarelli, 1994). Here, we concentrate on the presence of tightly wound regions that were observed in cryo-electron microscopy (Adrian et al., 1990; Bednar et al., 1994) and atomic force microscopy (Lyubchenko and Shlyakhtenko, 1997; Shlyakhtenko et al., 2003).

In cryo-EM experiments in high monovalent salt (> 150 mM Na^+) or divalent salt (> 10 mM Mg^{2+}), negatively supercoiled DNA plamids are seen to straighten up and "collapse". The intersegmental distance between the ascending and descending branches of the supercoil falls below experimental resolution. Our theory suggests a possible explanation for these results. In fact, chiral electrostatic interactions, as illustrated in the results, drive a transition to a tightly wound state.

A similar observation was performed by *in situ* atomic force microscopy experiments. In this case, the plasmids were found to have *regions* of tightly wound DNA (as opposed to complete tightening observed in cryo-EM). The extent of the tightly wound regions was found to increase with the degree of negative supercoiling (increases with $-\sigma$).

Our theory suggests that chiral electrostatic interactions may play a role in driving or stabilizing the transition to the tightly wound state. However, it must be mentioned that our predictions are in contrast with some experimental observations mentioned here. First, in both AFM and cryo-EM, tight supercoiling was never observed in relaxed plasmids (the ones with $\sigma = 0$), and it increased with the degree of *negative* supercoiling. This is the opposite of what our thory predicts. In fact, as can be seen in figure 3.5, we expect that the electrostatic stabilization should favour states with positive σ, and be progressively less and less effective when going in the negative σ direction.

There are several arguments that can explain the difference between our theoretical predictions and the experimental reality. First, the observation of increased tightening with increasing $-\sigma$ is compatible with the assumption that increasing ionic strength overwinds DNA. In fact, if this is the case, an increasing excess twist must be compensated by extra negative writhes, which may result in tightly wound regions.

The second point is that we did not include undulations of the molecular axis in our model. We may expect that undulations are suppressed for molecules that have high values of $|\sigma|$, because of the increased elastic stress accumulated in them.

On the other hand, for relaxed plasmids undulations might be a dominant effect. This might be the reason that relaxed plasmids were never observed to undergo the collapse transition.

It should also be mentioned that the experiments of Bednar et al. (1994) have been heavily criticized (Gebe et al., 1996). It was argued that the structures observed in those experiments cannot be the equilibrium structures. In dynamic light scattering experiments (Gebe et al., 1996), the radius of gyration of supercoiled DNA plasmids was found to have an opposite dependence on the salt concentration than that reported in the cryo-EM experiments. It was then argued that the preparation of the sample (fast immersion in low temperature) might not be fast enough to insure that the structures observed are not an artefactual. Therefore, one might conclude that the cryo-EM experiments do not show equilibrium structures.

The results reported in the cryo-EM experiments are also in contrast with other observations. In equilibrium catenation experiments (Rybenkov et al., 1997), the gel mobility of linear DNA molecules that were ligated in different ionic conditions was measured. The products of ligation are linear DNA, closed circles, and catenated closed circles, each of which can be distinguished in the gel experiments. One might expect that if a collapse transition occurs, there will be a drastic decrease in the amount of catenated circles. Instead, such reduction was never observed, even in the presence of high concentrations of divalent salts. In the same experiments, the mobility of relaxed closed circles was always the one that was expected for an open circular state. This indicated that no spontaneous writhing occured in those conditions.

In an electron microscopy and scanning electron microscopy study of DNA supercoiling (Cherny and Jovin, 2001), it was confirmed that divalent ions overwind DNA. Plasmids with a mild level of positive supercoiling were found to relax to an open circular state when adding divalent ions, in contrast to the expectations of our theory. The statistical significance of those results is however limited.

3.4.4 Biological evidence for asymmetric DNA crossovers

From the biological point of view, there are several interesting speculations that we may put forward. We address the issue of the presence of positively supercoiled DNA in hyperthermophiles, following the original idea of Timsit et al. (2010). Positively supercoiled DNA might be stabilized in its tertiary structure by chiral electrostatic interactions. This effect might aid the survival of organisms which need to thrive in extreme temperature conditions.

In the work of Timsit et al. (2010), it was also speculated that the origin of the chiral preference of certain topoisomerases (see section 3.1.2), may also lie within the asymmetry in the energy of DNA crossovers. It was shown in the experiments of Charvin et al. (2003) that topoisomerase IV acts preferentially on positive crossovers with a tilt angle close to 90°. This angle is much larger than the one that our model predicts as the equilibrium angle.

As explained in the Introduction, DNA supercoiling arises during the processes of transcription and replication. In the study of Peng and Jackson (2000), the amount of positive supercoiling arising after transcription was studied as a function of the increasing concentrations of polyamines, *in vivo*. The presence of spermine and spermidine increase the level of positive stress ahead of the transcription complex. Since it is known that the polyamines increase the DNA flexibility (Feuerstein et al., 1990; Baumann et al., 2000), and cause aggregation as well, the authors wanted to discriminate between these two effects. When adding BSA (bovine serum albumin), the aggregation properties of the polyamines on DNA is drastically reduced. However, when adding BSA, the levels of trascription-induced positive stress stayed the same as before. The authors conclude that the major effect is due to increased flexibility rather than aggregation. These results show that the electrostatic effects predicted by our theory have a minor role, at least *in vivo*.

3.5 Summary and outlook

In this chapter, we described a simple model of closed circular DNA, that includes a braided section (plectoneme), the two end loops, and the topological constraint

arising from ring closure. The model, in its simplicity, captures the interplay between twist and writhe and describes the basic physical behaviour of a collapse transition due to electrostatic attraction in the braid. The results of our model indicate that positively supercoiled molecules are strongly driven to the collapsed state, in the presence of DNA condensers or groove-specific counterions. Our model further predicts the possibility for writhing of relaxed closed circular molecules, or even negatively supercoiled molecules in the presence of sufficient attraction in the braid.

The existing experimental data do not clearly indicate that the effects we theoretically predict do happen. Some experiments can be interpreted as a sign of the presence of attractive forces in the braid, but others indicate that the theoretical expectations are opposite to the one observed. The theory therefore awaits further experimental studies. Some experiments in this direction are currently under way in the laboratory of Dr. Tim Albrecht.

Chapter 4

Single molecule supercoiling

In this chapter we describe how we applied the theory of DNA braiding to the case of single molecule DNA manipulation experiments. In these experiments, a single DNA molecule is attached to a micron-sized bead at one end, and at a glass surface at the other end. Using magnetic or optical tweezers, it is possible to rotate and move the bead, while measuring the force acting on it. When the bead is rotated, the DNA starts forming plectonemes, which, as in the case of closed circular DNA, is a way to relieve torsional stress.

In the Introduction, we explain briefly the experimental method and illustrate its main results and capabilities. We will also explain the reason why we chose to model not only the formation of a plectoneme, but also of a solenoid. Section 4.2 gives the detailed derivation of our model for the plectoneme and the solenoid, as well as an explanation of our description of the straight, worm-like chain portions of the molecule. We then extract the theoretical predictions in section 4.3 and relate them to known experimental results in the final section of this chapter. The material presented in this chapter will be published in a paper currently in preparation.

4.1 Introduction

Single molecule DNA micromanipulations have proven to be an extremely valuable tool for the study of DNA biophysics. This method allows for unprecedented

insight into DNA mechanics (Smith et al., 1992), provided means to test polymer elasticity theories (Smith et al., 1992; Bustamante et al., 1994; Marko and Siggia, 1995; Marko, 2007), allowed for precise determination of the bending (Bustamante et al., 1994; Baumann et al., 2000) and torsional (Bouchiat and Mezard, 1998; Strick et al., 1999; Bryant et al., 2003; Forth et al., 2008) elastic moduli of DNA, and revealed important aspects of DNA structural transitions (Allemand et al., 1998; Sheinin et al., 2011).

In single-molecule experiments with magnetic beads, there are two classes of curves that can be measured (Strick et al., 1996). In one class, one keeps the force fixed, rotates the magnetic bead, and measures the distance between the bead and the surface. Extension-rotation curves are obtained in this way. The other class of curves are the extension-force curves: keeping the number of turns fixed, one varies the force, and measures the distance between the bead and the surface. By modelling the elastic and electrostatic behaviour of DNA in these conditions, one can predict the theoretical force-extension and extension-rotation curves. The theoretical curves may then be fitted to the experimental ones, thus allowing for the extraction of the elastic modulii of the molecule. This has been the approach of numerous theoretical studies (Marko and Siggia, 1995; Vologodskii, 1994; Moroz and Nelson, 1997; Bouchiat and Mezard, 1998; Marko, 2007; Clauvelin et al., 2008, 2009; Mosconi et al., 2009; Maffeo et al., 2010; Neukirch and Marko, 2011; Marko and Neukirch, 2012; Mazur, 2012).

So far, the studies on single molecule DNA mainly focused on the extraction of the DNA elastic parameters. In one theoretical study (Clauvelin et al., 2009), the differences between various theoretical models for the intra-DNA electrostatic interactions in the plectonemic region were studied. The difference between the Poisson-Boltzmann model (Ubbink and Odijk, 1999) and counterion-condensation model (Ray and Manning, 1994) was not found to have a significant impact on the theoretical prediction of force-rotation and force-extension curves. Those two interaction models are in fact both purely repulsive, and subtle differences between them do not have significant effects on the equilibrium DNA conformation. The model that we will introduce, however, differs from the ones mentioned in two fundamental aspects. First, it predicts that *attractive* inter-DNA forces will occur. Second, it is a *chiral* model, in the sense that the interaction energies in a left-

handed and in a right-handed plectoneme are different. As we shall see, inclusion of such interaction model significantly alters the predicted extension-rotation and extension-force curves.

As in the case of free braids and of closed circles, the effects that we predict are expected to be maximum in the presence of DNA condensing agents. DNA micromanipulation experiments have been performed also in their presence (Baumann et al., 2000; Murayama et al., 2003; Ritort et al., 2006; Fu et al., 2006; Besteman, Hage, Dekker and Lemay, 2007; Besteman, Van Eijk and Lemay, 2007; Todd and Rau, 2008; Battle et al., 2009). In was unequivocally determined that in those conditions the shape of the molecules was toroidal. It became necessary then to model the toroidal shape as well.

There have been some attempts to theoretically model DNA toroids (Marko and Siggia, 1994; Battle et al., 2009). When modelling DNA toroids, it must be kept in mind that toroids are generally regarded as non-equilibrium structures, in the sense that their final shape heavily depends the size of an initially formed loop (Shen et al., 2000; Hud and Downing, 2001; Conwell et al., 2003; Su et al., 2004). It was argued that the inner radius of the toroids in fact correlates with the size of an initial loop, around which the toroid growth proceeds. All the models of toroids will all have inherent limitations.

We therefore decided to propose the simplest possible model of a toroid, which captures its structural features: a solenoid. Such geometrical shape allows us to study the interplay between the bending energy cost of loop formation, and the benefit given by favourable intersegmental contacts. Since in DNA toroids it is known that DNA filaments are hexagonally packed (see, e.g. (Leforestier and Livolant, 2009)), it is likely that we will underestimate the latter component of the interaction.

Precise data from single molecule experiments in the presence of DNA condensers is still lacking, due to the technical difficulty, using magnetic tweezers, to measure the entire force-extension curve. In fact, upon condensation the bead easily escapes the trap, and the sample is lost. This technical inconvenience might be solved in experiments using the new "angular optical trap" technique (Bryant et al., 2003; La Porta and Wang, 2004; Forth et al., 2008). In fact, with such method it is the *extension* that one controls during the experiment, not the force.

The distance between the bead and the surface is changed, and the force is measured. This is a complementary approach to the one employed in magnetic tweezer experiments. To model such system, it is necessary also to modify the theoretical approach to the problem. In this chapter, we propose such novel approach, which was the idea of Dr. Dominic Lee.

4.2 Model

In this section we describe our model of single molecule DNA supercoiling. As said in the Introduction, we will model two different geometrical configurations of the system: the plectoneme and the solenoid. The two geometrical shapes are different in terms of variables that describe them, but share common features. For this reason, we start by presenting a general formalism that allows us to express the force acting on the bead in terms of variables common to the two cases. We then describe in detail our model for the plectoneme and for the solenoid.

4.2.1 General features of the system

We consider a molecule of length L which is thethered at one end to a surface and to the other end to a rotating bead. The molecule is kept at a fixed end-to-end distance z and it is anchored in such a way that it is not able to swivel around its anchoring points. This results in a topological constraint to the system, which couples the twist and the writhe of the molecule. We consider that the bead is rotated n times away from the straight relaxed molecular configuration. Therefore we have:

$$n = \Delta Tw + Wr. \tag{4.1}$$

Where ΔTw is the excess twist away from Tw_0, defined as the value of the twist of a straight molecule in a given ionic condition. It is given by

$$\Delta Tw = Tw - Tw_0 = \frac{1}{2\pi}\int_0^L \mathrm{d}s\,[g(s) - g_0]\,. \tag{4.2}$$

In what follows, as we have done for the case of closed circular DNA, we assume that the twist is homogeneously distributed throughout the length of the molecule,

so that $Tw = L(g - g_0)/2\pi$. The value of the writhe, Wr, is different in the case of the plectoneme and the solenoid.

In all the models we consider, there is a torsional energy contribution. The description of the torsional energy contribution is identical to the case of closed circular DNA (see section 3.2). Under such assumptions, and using equations (4.1) and (4.2), we can write a general expression for the torsional energy:

$$E_{tors} = l_p^t \frac{L}{2} (g - g_0)^2 = l_p^t \frac{2\pi^2}{L} (n - Wr)^2 , \tag{4.3}$$

where then the writhe Wr needs to be specified for the geometrical model under consideration.

It is useful, for the purpose of our theoretical analysis, to express the free energy of the system in a general way. In the models we shall consider, there is a writhed portion and a straight portion which is fluctuating in a worm-like chain fashion. Therefore, we can write:

$$L = L_w + \Delta L \tag{4.4}$$

where ΔL is the length of the worm-like chain fluctuating part, and L_w is the length of the writhed state. The free energy can be expressed then as

$$f(\mathbf{X}) = f_w(\mathbf{X}) + f_{WLC}\Delta L, \tag{4.5}$$

We define $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ as the vector of variational variables upon which the energy of the system depends upon. The term $f_{WLC}\Delta L$ is the free energy of the worm-like chain part, which we discuss in the next subsection, and f_w is the free energy of the writhed portion of the molecule.

4.2.2 Worm-like chain behaviour of the straight portions

To describe the contribution of these fluctuations to the free energy of the system, we use the framework developed by Marko and Siggia (1995). There, the authors write the free energy per unit length of a worm-like chain subject to a force F_{WLC}

with finite bending persistence length l_p^b as follows:

$$g_{WLC}(F_{WLC}) = \min_a \left[\left(\frac{ak_BT}{2l_p^b} - F_{WLC} \right) \left(\coth 2a - \frac{1}{2a} \right) \right]. \tag{4.6}$$

Once the minimum in this expression is obtained, the end-to-end distance is given by

$$\rho(F_{WLC}) \equiv \frac{z}{\Delta L} = -\frac{\partial g_{WLC}}{\partial F_{WLC}}, \tag{4.7}$$

where ΔL is the contour length of the molecule.

We want, instead, to find the value of the free energy in equation (4.6) as a function of the end-to-end distance, instead of the force. To do so, first we calculate $g_{WLC}(F_{WLC})$ through equation (4.6). After that, by using equation (4.7) we compute $\rho(F_{WLC})$, which we then can invert numerically and find $F_{WLC}(\rho)$. In such way, we obtain the value of the force acting on the worm-like chain as a function of its end-to-end distance. Once that is obtained, we can plug such value back into equation (4.6) and obtain a fully numerical expression for $g_{WLC}(\rho)$. We then need to switch back to the fixed length ensemble. We finally obtain the free energy contribution per unit length of the worm-like chain fluctuations:

$$f_{WLC}(\rho) = g_{WLC}(\rho) + F_{WLC}(\rho)\rho, \tag{4.8}$$

where $g_{WLC}(\rho)$ was calculated numerically as described above.

4.2.3 Force calculation

To calculate the force that is needed to keep the bead at a given height z, we calculate the value of the minimum free energy f_{min}, and then use the following expression:

$$F = \frac{\mathrm{d}f_{min}}{\mathrm{d}z}, \tag{4.9}$$

where f_{min} is the free energy at the minimum. If $\mathbf{X}^\star$ is the vector corresponding to the minimum energy, then we can write equation (4.9) as

$$F = \sum_i \left.\frac{\partial f_{min}}{\partial X_i}\right|_{X_i=X_i^\star} \frac{dX_i}{dz} + \left.\frac{\partial f_{min}}{\partial z}\right|_{\mathbf{X}=\mathbf{X}^\star} = \left.\frac{\partial f_{min}}{\partial z}\right|_{\mathbf{X}=\mathbf{X}^\star} \tag{4.10}$$

where the last equality is due to the fact that the gradient of the energy vanishes at the minimum. Substituting equation (4.5) into equation (4.10) we obtain, by applying the 2D chain rule for variable changes:

$$F = \left.\frac{\partial f_{min}}{\partial z}\right|_{\mathbf{X}=\mathbf{X}^\star} = \frac{\partial}{\partial z}\left[f_{WLC}\left(\frac{z}{\Delta L}\right)\Delta L\right]\Big|_{\Delta L=\Delta L^\star} = \frac{\mathrm{d}f_{WLC}(\rho)}{\mathrm{d}\rho} = F_{WLC}(\rho), \tag{4.11}$$

The last equality comes from straightforward differentiation of equation (4.8):

$$\frac{\mathrm{d}f_{WLC}(\rho)}{\mathrm{d}\rho} = \frac{\partial F_{WLC}}{\partial \rho}\frac{\partial g_{WLC}}{\partial F_{WLC}} + F_{WLC} + \rho\frac{\partial F_{WLC}}{\partial \rho} = F_{WLC} \tag{4.12}$$

where equation (4.7) is taken into account.

4.2.4 Plectoneme geometry and energy function

Our model of the plectonemic state is shown in figure 4.1. We start by listing the basic assumptions we make concerning the geometry of the system, and then write the free energy function of the system.

The molecule is assumed to be formed by three disconnected portions: a plectoneme, an end loop, and two fluctuating straight portions that connect the end of the braid to the anchoring points. Hence, we consider the energy of the system to be made of the energy of the braid, plus the energy of the end loop. Then, we have:

$$L = L_p + L_{loop} + \Delta L \tag{4.13}$$

The meaning of the variables of the system is the same as before. In addition, we define L_p as the total length of the plectonemic region of the molecule, and $p = L_p/2P$ as the number of superhelical pitches in the plectoneme. Coming back to the notation of equation (4.5), we have $\mathbf{X} \equiv (R, L_p, p, \Delta L, \lambda_h^\star)$. The superhelical

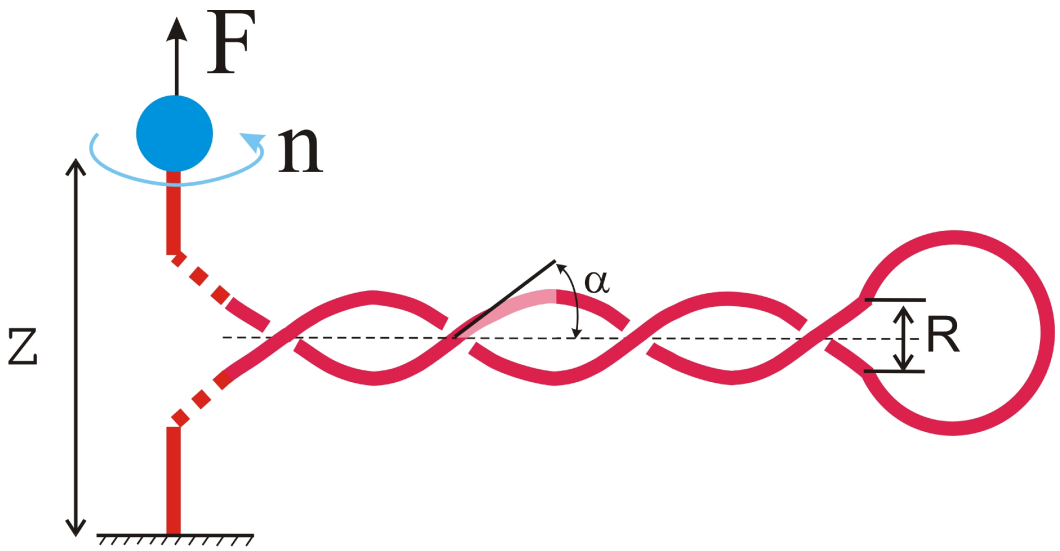


Fig. 4.1: Schematic illustration of the plectoneme geometry model. The DNA molecule is considered to be formed by two straight portions connecting its anchoring points to the braid, plus a braided section, plus the end loop. Dashed lines indicate that we do not consider the portions that connect the straight segments to the braided section.

pitch P and the braid tilt angle α are related to R, L_p and p through the following geometrical relationship (see equation (A.14)):

$$p = \frac{2P}{L_p} = \frac{L_p \alpha}{2\pi R}, \tag{4.14}$$

where we used the small angle approximation for α.

The writhe of the plectoneme can be expressed in very simple terms:

$$Wr = 2p. \tag{4.15}$$

The torsional energy equation (4.3) can then be reexpressed as

$$E_{tors} = l_p^h \frac{2\pi^2}{L} (n - 2p)^2 . \tag{4.16}$$

To describe the physics of the model, we write the following free energy func-

tion:

$$f_w^p(R, L_p, p, \Delta L, \lambda_h^\star) = \\ = f_{tors}(p) + f_{bend}(R, L_p, p) + f_{loop}(R, \Delta L) + f_{adapt}(\lambda_h^\star) + f_{es}(R, L_p, p, \lambda_h^\star) \quad (4.17)$$

where

- f_{bend} is the bending energy (see equation (b.2.5)).
- f_{loop} is the the end loop energy (see equation (3.21)).
- f_{es} is the electrostatic free energy in the plectonemic region.
- f_{adapt} is the adaptation free energy in the plectonemic region, due to sequence-dependent and thermal distorsions of the DNA helical structure (see equation (2.12)).

Taking into account equation (4.14), we may express the bending energy density of a braid (see equation (b.2.5)) in the following way:

$$f_{bend} = L_p l_p^b \frac{2\alpha^4}{R^2} = \frac{32\pi^4 R^2}{L_p^3} p^4, \quad (4.18)$$

where we used equation (4.15).

The electrostatic energy and adaptation energies were described in chapter 2. It is interesting here to show the shape of the electrostatic energy term, with our choice of variables:

$$f_{es}^p = \frac{L_p}{2} \left[\mathscr{F}_0(R, p, \lambda_h^\star) + \alpha \mathscr{F}_1(R, p, \lambda_h^\star)\right] = \frac{L_p}{2} \mathscr{F}_0(R, p, \lambda_h^\star) + \pi R p \mathscr{F}_1(R, p, \lambda_h^\star). \quad (4.19)$$

It is worth noticing that the magnitude of the chiral term $\mathscr{E}_1$ does not scale with L_p, but with p instead.

We conclude this section by describing the procedure of energy minimization. The energy function described here is a function of five variables, which makes it computationally heavy to minimize. Therefore, it is convenient to minimize the function separately for some of the variables. In particular, it is convenient to do

so for ΔL, as only the end loop and the worm-like chain terms depend upon this variable. We have

$$\frac{\partial f}{\partial \Delta L} = \frac{\partial}{\partial \Delta L}\left[f_{loop}(R, \Delta L) + f_{WLC}\left(\frac{z}{\Delta L}\right)\Delta L\right] = 0. \tag{4.20}$$

We can do some analysis of the last term. We have in fact that

$$\begin{aligned}\frac{\partial}{\partial \Delta L}\left[f_{WLC}\left(\frac{z}{\Delta L}\right)\Delta L\right] &= -\frac{z}{\Delta L^2}\frac{\mathrm{d}f_{WLC}(\rho)}{\mathrm{d}\rho}\Delta L + f_{WLC}\left(\frac{z}{\Delta L}\right) = \\ = -\frac{z}{\Delta L}F_{WLC}\left(\frac{z}{\Delta L}\right) + g_{WLC}\left(\frac{z}{\Delta L}\right) &+ \frac{z}{\Delta L}F_{WLC}\left(\frac{z}{\Delta L}\right) = g_{WLC}\left(\frac{z}{\Delta L}\right),\end{aligned} \tag{4.21}$$

where we used equations (4.8) and (4.12). The derivative of the end loop term is

$$\frac{\partial f_{loop}}{\partial \Delta L} = \frac{\partial L_{loop}}{\partial \Delta L}\frac{\partial f_{loop}}{\partial L_{loop}} = -l_p^b \frac{[2\pi L_{loop} - \pi^2 R/2]\,[3\pi^2 R/2 - 2\pi L_{loop}]}{L_{loop}^4}. \tag{4.22}$$

Using equations (4.21) and (4.22), the numerical solution of equation (4.20) is easy. By expressing this as a function of R and L_p, we can then eliminate one variable from the problem, and the minimization becomes easier.

4.2.5 Solenoid geometry and energy function

The basic geometric structure of the model is depicted in figure 4.2. We model the molecule as made of straight portions and a solenoidal part L_s long:

$$L = L_s + \Delta L. \tag{4.23}$$

The solenoidal part is made up of a number $|s|$ of equally shaped rings of DNA, which are assumed to interact only with their nearest neighbours. Each ring is at a distance R apart from the neighbouring ring, and it is assumed that the rings are disconnected from each others. We define D as the radius of each ring and l as the length of one ring. We have therefore:

$$L_s = |s|l = |s|2\pi D. \tag{4.24}$$

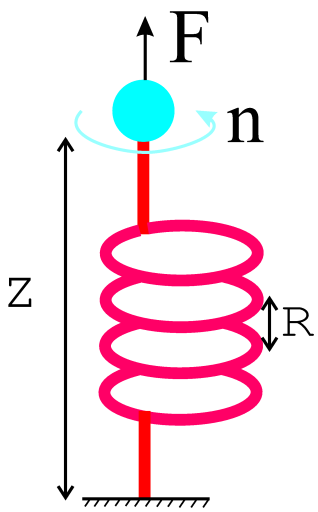


Fig. 4.2: Schematic illustration of our model of a DNA molecule in a solenoidal state. The molecule consists of s disconnected ringls ($s = 4$ in figure), at a distance R apart.

The writhe of the solenoid is given approximately by (Marko and Siggia, 1994):

$$Wr = s. \tag{4.25}$$

Using this expression, we can write the torsional energy of the solenoid using equation (4.3):

$$f_{tors} = l_p^h \frac{2\pi^2}{L} (n - s)^2 \tag{4.26}$$

To express the solenoid free energy, we need to calculate the bending energy and the electrostatic interaction contributions.

Each of the rings that make up the solenoid contributes with its bending energy to the total energy of the system. The bending energy of one ring can be calculated by a simple application of equation (b.2.19):

$$f_{ring} = l_p^b \frac{2\pi^2}{l}, \tag{4.27}$$

so that the total bending energy is given by

$$f_{bend} = |s| f_{ring} = l_p^b |s| \frac{2\pi^2}{l} = l_p^b \frac{2\pi^2}{L_s} s^2. \tag{4.28}$$

The electrostatic energy is evaluated under the assumption that the ring diameter is very large, i.e. $D \gg \lambda_D, a$. In this way, we can approximate the interaction between neighbouring DNA segments as that of two parallel DNA molecules. Within such hypothesis, the energy is simply given by the sum of $|s| - 1$ pairs of parallel DNA molecules:

$$f_{es}^s(R, L_s, s, \lambda_h^\star) = (|s| - 1)\, l \mathscr{F}_0(R, \overline{\Delta\Phi}_{min}, \lambda_h^\star) = \frac{|s| - 1}{|s|} L_s \mathscr{F}_0(R, \overline{\Delta\Phi}_{min}, \lambda_h^\star), \tag{4.29}$$

where, as done previously, we keep the value of the azimuthal angle $\overline{\Delta\Phi}$ to its optimum. It is interesting to notice that for large enough values of s, the electrostatic interaction energy is rather insensitive to the value of s, and is mainly dependent on the value of L_s.

The total energy of the solenoid is then expressed in the following way:

$$\begin{aligned} &f_w^s(R, L_s, s, \lambda_h^\star) = \\ &= \frac{2\pi^2}{L}(n - s)^2 + l_p^b \frac{2\pi^2 s^2}{L_s} + \frac{|s| - 1}{|s|} L_s \mathscr{F}_0(R, g, \lambda_h^\star) + f_{WLC}\left(\frac{z}{L - L_s}\right)(L - L_s) \end{aligned} \tag{4.30}$$

After energy minimization, we will always check that at the minimum it is still true that $D \gg \lambda_D$.

4.2.6 Model limitations

Our model has several limitations. We start by listing the ones that the model has in general.

- We assumed that the twisting persistence length does not depend on the applied force. This assumption is valid in the limit of a high force (Marko, 2007).

- The model does not take into account that the torque transfer to the plectoneme is not complete, due to rotational fluctuations of the bead (see (Mazur, 2012)).

- We did not include any structural transition in the model. It is known from the experimentals that DNA denatures at moderate-high values of negative n, and such propensity increases with the value of the applied force (see, e.g. (Mosconi et al., 2009)). The model should be applied only at values of n such that the elastic energy density per base pair is lower than the denaturation energy cost per base pair ($\approx 2.0\ k_BT$/bp (Marko, 2007)).

For the plectoneme model, we have:

- The limitations on the model of the braid that we listed in section 2.1.6. Once again, we should apply this theory only to the case of a very long plectoneme. It is also important to keep in mind that we did not take into account any undulations of the molecular centreline, which means that our theory should be strictly valid, as before, only in the limit of a high force.

- The model of the end loop has its own limitations, which were discussed in section 3.2.5.

- The plectoneme and the straight portions, which fluctuate in a worm-like chain fashion, do not have matching tangents. Therefore, it is likely that the elastic energy cost of the plectoneme is underestimated. For very long molecules dominated by the elastic and electrostatic interaction components, this contribution may be thought as a minor correction.

In the case of the solenoid, we may further add

- The equilibrium inter-strand distance R goes to infinity when there are repulsive electrostatic interactions in the solenoid. This is certainly not the case in reality, because there is a coupling between the total length of the solenoid and the value of R. However, this is not a very important limitation because there is still a competing term that keeps the value of the length of the solenoid to a finite value, which is the worm-like chain contribution to the energy.

- It is well known from experiments (see section 4.1) that a solenoid is not the equilibrium geometrical state in condensing conditions. The physical states reported here should be interpreted as a model of a toroidal state, which in turn has all the complications which were mentioned in the Introduction. In a toroid, the electrostatic interactions are enhanced by the presence of more neighbouring molecules. A rough estimate then is that the magnitude of the electrostatic energy in a toroid is three times larger than that of a solenoid. The energy of the solenoid may then be considered as an upper bound to the energy of a toroid.

4.3 Results

In this section, we illustrate the results of our model of single molecule supercoiling. We minimize the energy functions for the plectoneme (equation (4.17)) and for the solenoid (equation (4.30)), at fixed n and z. We then compute the force necessary to keep the bead at the height z, as prescribed in section 4.2.3.

We perform the minimization for two distinct sets of parameters, which we will conventionally call "high" and "low" salt cases. These have to be regarded as two extreme cases, which we show here to illustrate the basic physical behaviour of the system. With "low salt", we mean that the molecule is in a solution where only monovalent salt is present, so that we can assume that the value of θ is low, and that the ions have no propensity to bind in the DNA grooves. On the other hand, by "high salt", we mean that in the solution there are ions that have a strong affinity to the DNA grooves, and that completely neutralize the DNA bare charge. A summary of the parameters used in the two cases is given in table 4.1.

	Ion type	θ	f_1	f_2
Low salt	monovalent ions	0.4	0.0	0.0
High salt	DNA-condensing ions	1.0	0.4	0.6

Table 4.1: Summary of the parameters for the two different types of solution conditions described in the results.

4.3.1 Force versus extension experiments

Figure 4.3 shows the plots of the minimum energy against z/L (force-extension experiments), for different fixed values of n, for high and low salt, and for the plectoneme and the solenoid. Figure 4.4 shows the corresponding force acting on the bead. In order for our theory to be valid, we chose to plot values of z/L that go from 25% to 75% (see section 4.2.6). In low salt (left column), the curves are identical for $n = -60$ and $n = 60$. In fact, one does not expect any asymmetry between positive and negative values of n, in that case. For $n = 0$, the curves collapse onto the same curve, which indicates that no writhing occurs, and the curves correspond to pure worm-like chain behaviour. The worm-like chain behaviour can be seen in the corresponding curves for the force against z/L. The minimum energy increases as a function of z/L. When z/L increases, in fact, there is a lower fraction of the DNA contour length which is avaliable for both worm-like chain fluctuations, and for partitioning the linking number into writhe. The high torsional energy cost, plus the high entropic cost of stretching the molecule, results in a higher energy. This also results in a corresponding increase in the force.

In high salt, the situation is different. For $n = -60$, the plectoneme minimum energy is lower than the corresponding low salt case. A small electrostatic attraction is present in the system, which lowers the minimum energy. The electrostatic attraction is not sufficient to drive an inversion of the handedness of the plectoneme.

For $n = 0$ in high salt, the theory predicts that electrostatic attraction between portions of the molecule, in both the plectoneme and the solenoid geometry, is sufficient to generate a writhed state. Our model predicts that in this case the plectoneme geometry is still the most favourable geometric configuration. The solenoid exhibits a force plateau when the molecule is being pulled. This is due to the fact that the ratio between the length of the writhed portion and the end-to-end distance stays approximately constant as the molecule is pulled.

For $n = 60$, in high salt, positive supercoiling in the plectoneme results in a strong electrostatic attraction, so that the theory predicts a significant negative energy. The solenoid state however has a positive energy, as the electrostatic attraction is not able to overcome the bending and torsional energy costs.

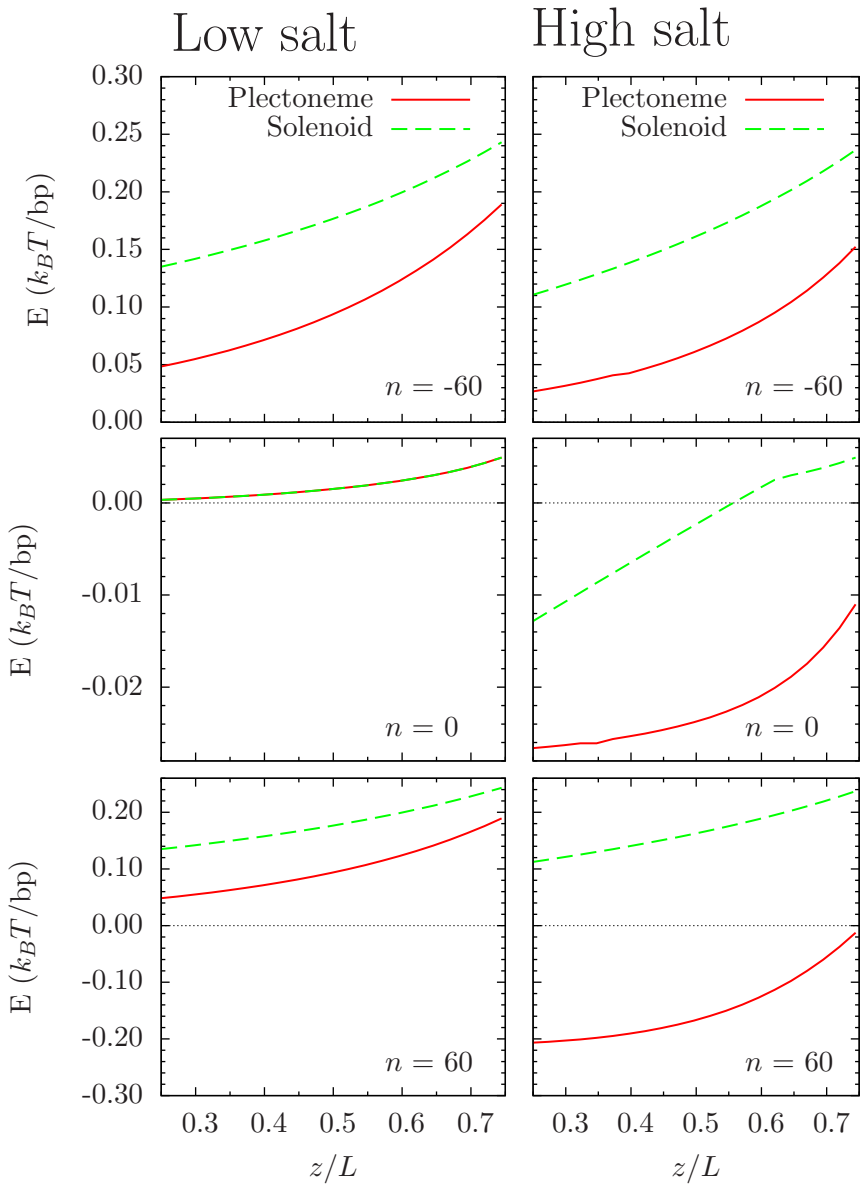


Fig. 4.3: Summary of the results of the minimum energy against extension of an 8000 bp molecule, at fixed number of turns n. The plots were calculated by minimizing the free energy given in equations (4.17) and (4.30). The parameters used for the "high" and "low" salt cases are given in table 4.1.

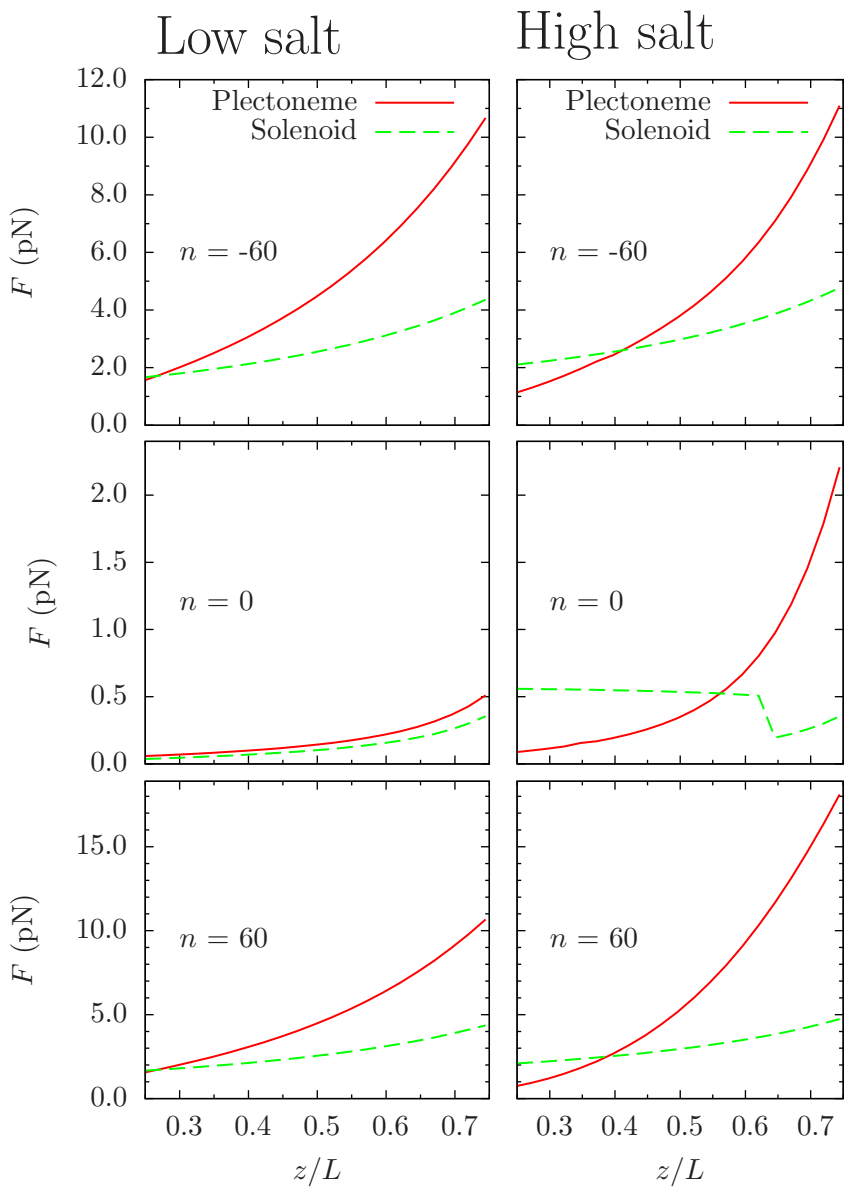


Fig. 4.4: Summary of the results of the force against extension of an 8000 bp molecule, at fixed number of turns n. The plots were calculated by minimizing the free energy given in equations (4.17) and (4.30), and then using equation (4.11) to calculate the force. The parameters used for the "high" and "low" salt cases are given in table 4.1.

4.4 Force versus rotation experiments

Figure 4.5 shows the minimum energy of the plectoneme and the solenoid as a function of the number of imposed turns n, at fixed bead height z. The curves are plotted for the low and high salt cases. Figure 4.6 shows the corresponding values of the force that are necessary to keep the bead at that particular height.

As in the case of force-extension experiments, in low salt the curves are all symmetric. No chiral term contributes to the total energy, as the electrostatic interaction term is purely repulsive, and there are no other terms in the energy that depend on the sign of n. The dependence of the minimum energy on n is parabolic.

In the case of high salt, there are significant deviations from the parabolic behaviour. The solenoid minimum energy keeps its parabolic dependence on n, since no chiral terms are present in the energy function (see equation (4.30)). Only for $z/L = 25\%$, and for small values of n, the solenoid minimum energy falls below zero. For the plectoneme, the transition to an electrostatically stabilized state is clearly visible in the plots, and is identified by a kink in the curves. The geometry of the plectoneme is very favourable for positive n, as the braid is strongly stabilized by the chiral term in the interaction. Moreover, a high fraction of the linking number can be effectively absorbed by the plectoneme. For $z/L = 25\%$ and $z/L = 50\%$, the minimum energy monotonically decreases from the transition point ($n \approx$ -5) up to the highest values of n that we decided to use. For $z/L = 75\%$, the minimum energy curve starts increasing after $n \approx 30$. Since the value of z/L is high, the system does not have enough contour length to be able to partition the linking number into writhing. As a result, continuing to add links (increasing n) has the effect of making the system less stable.

It is interesting to notice that the solenoid minimum energy is lower than the plectoneme one only when z/L is low, and for moderately negative values of n.

4.5 Discussion

We developed a theory of single molecule DNA experiments, by proposing two different models for the geometry of the molecule, and including the Kornyshev–

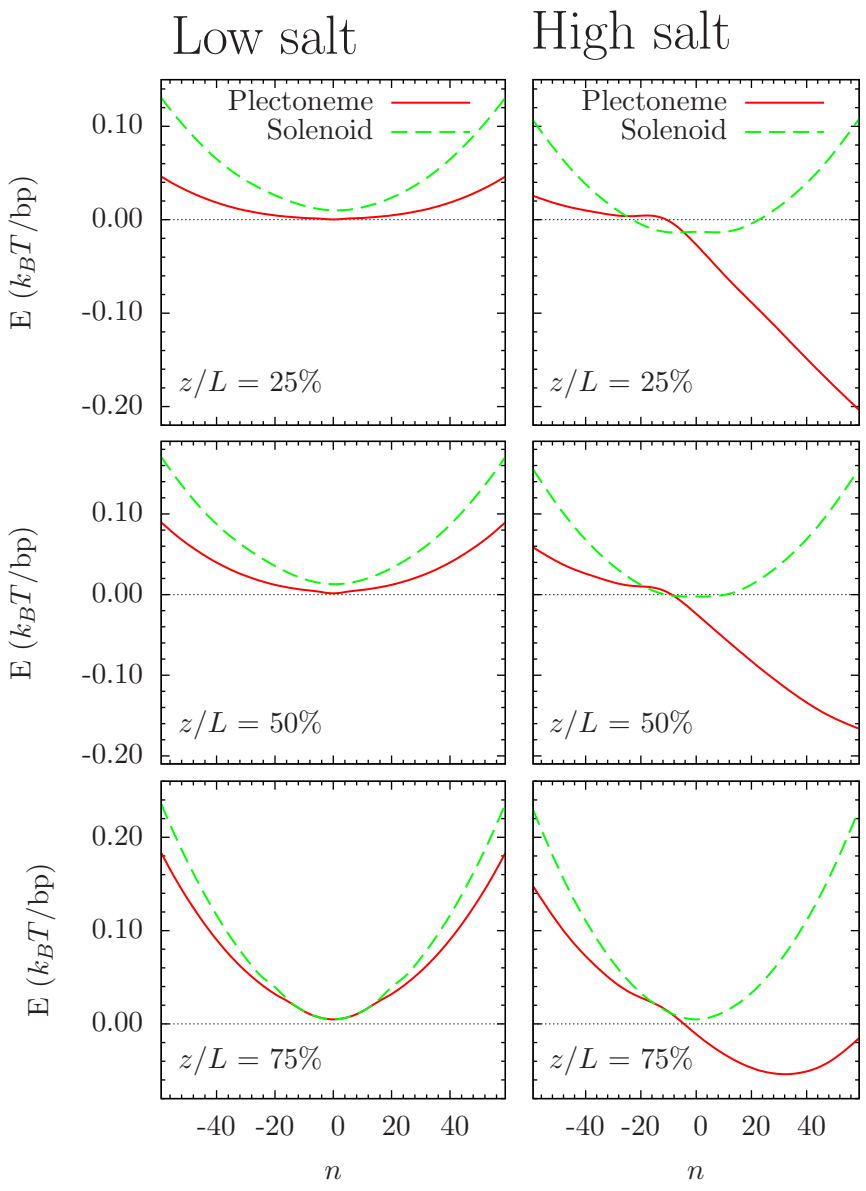


Fig. 4.5: Summary of the results of the minimum energy against number of turns of an 8000 bp molecule, at fixed bead height z. The plots were calculated by minimizing the free energy given in equations (4.17) and (4.30). The parameters used for the "high" and "low" salt cases are given in table 4.1.

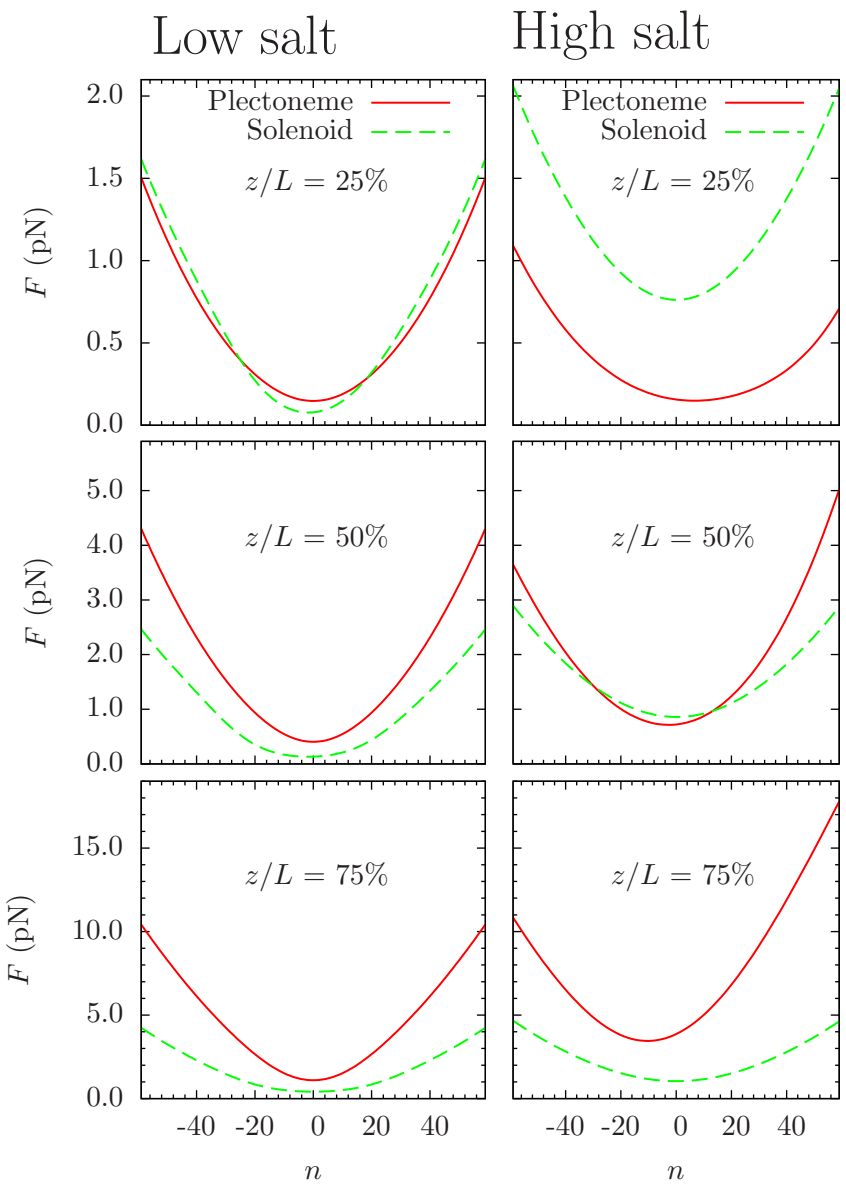


Fig. 4.6: Summary of the results of the force against rotation of an 8000 bp molecule, at fixed bead height z. The plots were calculated by minimizing the free energy given in equations (4.17) and (4.30), and then using equation (4.11) to calculate the force. The parameters used for the "high" and "low" salt cases are given in table 4.1.

Leikin theory of DNA-DNA interactions to model electrostatics. We analysed the results, which we now want to relate to the experimental reality.

The predictions of our theory reproduce fairly well the experimental curves in low salt (see, e.g. (Mosconi et al., 2009)). The case of low salt has already been studied in many theoretical works, and we now have a very clear understanding of the physics behind this system. For us, the most interesting case is the one in which ions that cause DNA condensation are present in the experimental buffer.

This section is divided into two parts. In the first part we will compare our theoretical predictions to the few experiments that were performed in the presence of DNA condensers. In the second section, we shall suggest ways of testing the results of our theory.

4.5.1 Single molecule experiments in the presence of condensing agents

Single molecule DNA experiments were performed in the presence of DNA condensers using optical traps (Baumann et al., 2000; Murayama et al., 2003; Ritort et al., 2006; Fu et al., 2006; Todd and Rau, 2008; Battle et al., 2009). This means that in those experiments it was not possible to rotate the bead and introduce extra linking number in the system.

The experimental force versus distance curves in that case have show a characteristic pattern. For short distances, the force increases sharply. At a certain distance, the force reaches a plateau. After another characteristic end-to-end distance, the force starts increasing again, with a typical worm-like chain pattern. The fine structure of the force plateau reveals the presence of peaks and drops in the force.

These curves were interpreted as being the signature of the presence of toroidal condensates. Such behaviour was modelled very effectively by a theory which describes the toroidal geometry (Battle et al., 2009). One can unequivocally conclude that the geometry of the $n = 0$ case is a toroidal one.

Our predictions show that when n is close to zero, the solenoid and the plectoneme have very similar energies. As stated in section 4.2.6, our model of a solenoid may represent an upper bound for the energy of an actual DNA toroid.

Because of the hexagonal packing, in fact, the number of DNA-DNA interactions in a toroid is increased by a factor of about three. If this is indeed the case, then our model reproduces the force plateau observed in the experiments (see figure 4.3, panel with $n = 0$ and high salt).

Single molecule DNA experiments in the presence of condensing agents have been performed also with rotating beads (Besteman, Hage, Dekker and Lemay, 2007; Besteman, Van Eijk and Lemay, 2007; Shao et al., 2012). In the case of the work of Besteman, Hage, Dekker and Lemay (2007), the experiments were carried out using magnetic tweezers in the presence of spermidine and cobalt-sepulchrate. As mentioned in the introduction to this chapter (see section 4.1), there is a significant technical shortcoming when performing such experiments. In fact, it is possible only to measure the *critical force*, under which the bead falls out of the trap and the DNA molecule condenses. It was possible therefore to measure the critical condensation force as a function of the number of applied turns. It was observed there that the critical force increases approximately linearly with positive n, and then there is a plateau for $n < -20$ (the number of base pairs was approximately 8000 in those experiments).

Our results cannot directly reproduce the behaviour of the critical force, as our electrostatic theory requires significant interaction lengths to be considered valid. However, the results of Besteman, Hage, Dekker and Lemay (2007) can be understood in terms of the qualitative physics presented here. In the positive n branch, our theory correctly reproduces the quasi-linear behaviour of the critical force as a function of n. The origin of the force plateau at negative n remains unclear, though.

Similar experiments were carried out by the same group (Besteman, Van Eijk and Lemay, 2007). The authors studied the variation of the condensation force as a function of the concentration of condensing agent. The experimental curves show that the condensation force increases up to a concentration threshold, and then decreases afterwards. Such behaviour was predicted by the theory of counterion condensation (Nguyen et al., 2000; Grosberg et al., 2002; Zhang and Shklovskii, 2005), and it was observed in those experiments. This suggests that the strong coupling effects discussed in section 1.2.3 are responsible for the observed behaviour. The theoretical framework built by us may however still explain the observations.

In fact, Todd and Rau (2008) argued that the same behaviour is predicted under different assumptions. The Oosawa–Manning theory of counterion condensation (Manning, 1969) successfully predicts the observed behaviour, provided that the interactions between the cations in solution and their corresponding anions is taken into account. In fact, the cation species may resolubilize past beyond a certain concentration of the ions in solution.

We conclude this section by mentioning a very recent experimental work (Shao et al., 2012). In this work, extension-rotation experiments were carried out in the presence of the polyamines spermine and spermidine. In the buffer, also about 200 mM KCl were added, which is known to hinder the condensation properties of polyamines (Krasnow and Cozzarelli, 1982). The main effect that was observed was that DNA denaturation occurred less readily at negative n. Increasing the concentration of polyamines had the effect of stabilizing the base pairing of DNA. A slight increase in the slope of the extension-rotation curve was seen in the positive n branch of the experimental curves. The authors interpreted this as an increase in writhing, which may be due to softening of DNA, or to the plectoneme radius becoming smaller.

Our theory clearly predicts that in the presence of cations that have high affinity to the DNA grooves there is a tendency of the molecule to writhe in the positive direction. In the experiments of Shao et al. (2012), this was not observed, but it cannot be ruled out that the reason is that the presence of a high concentration of monovalent salt competes with the polyamine binding to DNA. We are currently investigating on possible interpretations of these experimental results within the context of our theory.

4.5.2 Experiments to be performed

The theory presented in this chapter awaits experimental verification. As stated in the introduction (section 4.1), the ideal instrument to test the predictions of the theory is the rotating optical tweezer (Bryant et al., 2003; La Porta and Wang, 2004; Forth et al., 2008). Using this setup, it is possible to probe the full force-extension and force-rotation curves in the presence of DNA condensers. In fact, the very reason we chose to formulate the theory in the fixed z ensemble is that

we wanted to give predictions for those kind of experiments.

Ideally one should perform a set of experiments measuring the force-extension and force-rotation curves in the presence of spermine, spermidine, cobalt-hexammine, and varying their concentrations. It would be also useful to measure whether divalent ions such as Mg^{2+} or Ca^{2+} are able to drive the system to the collapsed state.

Chapter 5

Four-bead DNA braiding

This final chapter illustrates a model we developed to describe the behaviour of a pair of DNA molecules in a four-bead micromanipulation experiment (see figure 5.1). The four-bead system was first presented in (Noom et al., 2007), and was used there to study the interaction of proteins with DNA. The theory we present here aims at the description of experiments of a similar kind, which are currently under way in the laboratory of Prof. Gijs Wuite, in Vrije Universiteit Amsterdam. In such experiments, a single DNA braid is formed, by attaching two different DNA molecules to four beads, which are trapped in four optical tweezers. Once the braid is formed, the force acting on one of the beads is measured as a function of the separation between the beads.

The model we present here for the description of the system is simplified to its maximum degree. It is needed to give a first estimate of the order of magnitude of the effects due to chiral electrostatic forces. Such estimate has been given to Prof. Wuite and his co-workers to design the initial experimental setup of the system.

A much more involved and rigorous theory is currently being developed and analysed jointly by Dr. Dominic Lee, and our theoretical partners of University College London (Prof. Gert van der Heijden, Dr. Eugene Starostin, Dr. Anthony Korte). The simple theory presented in this chapter needs to be refined to take into account many other effects, which we will discuss in section 5.1.2.

Section 5.1 will illustrate the basic structure of the model we made for the four-bead assay. In section 5.2 we will show force versus inter-bead distance curves

predicted by the model, both in the case of homologous and non-homologous interactions. In the final section of this chapter we list a number of experiments that have to be performed to test the results of the theory. This chapter lacks a "Discussion" section, since currently there are no experiments published in the literature.

5.1 Model

In this section we briefly illustrate a simple model of the four-bead system. The experimental setup we have in mind is of the type described in reference (Noom et al., 2007). Figure 5.1 shows a graphical representation of our model. We propose a two-state model the system (similarly to what we have done for the closed loops) as being composed of (a) a braid and (b) the four straight segments that connect the beads to the braid.

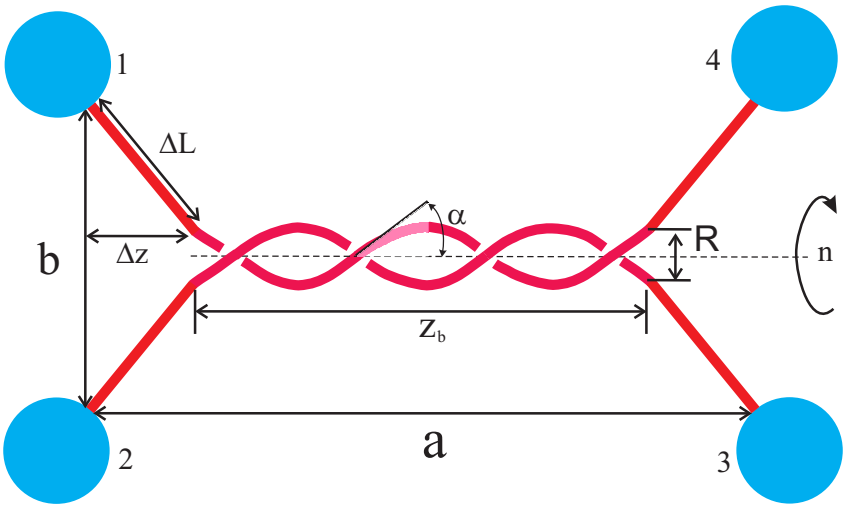


Fig. 5.1: Graphical representation of our model of the four-bead experiments, including the definitions of the various geometrical quantities mentioned in the text.

Box 4: Geometry of the four-bead system

First, we write the basic geometrical relationships between the quantities defined in figure 5.1. They read as follows:

$$L = L_b + 2\Delta L. \tag{5.3}$$

$$a = z_b + 2\Delta z \tag{5.4}$$

$$\Delta L^2 = \Delta z^2 + \left(\frac{b}{2} - \frac{R}{2}\right)^2 \tag{5.5}$$

In equation (5.4) we have to substitute a relationship between the quantity z_b and L_b. This is given by $z_b = L_b \cos\alpha \approx L_b$, for geometrical reasons. The above equations then can be solved for L_b, and equation (5.1) is obtained.
We then have to impose that there is a fixed number of turns applied on the system. To do so, we fix the linking number to be equal to n. Then we may write

$$n = Lk = Tw + Wr \tag{5.6}$$

where Tw and Wr are the twist and the writhe of the *braid*, respectively. Now, since we impose that the braid axis has to be straight, the writhe of the braid is zero. The twist of the braid can be calculated easily, and is equal to (see appendix A)

$$Tw = \frac{\sin(2\alpha)L_b}{2\pi R} \approx \frac{\alpha L_b}{\pi R}. \tag{5.7}$$

Then, substituting equation (5.7) into equation (5.6), and rearranging for α, we arrive at equation (5.2).

5.1.1 Geometry and energy function

We suppose that bead number 4 has been turned around the others n times (positive n is in the direction indicated by the arrow in figure, negative n the opposite way). The following equations relate the geometrical parameters to each others:

$$L_b = \frac{L^2 - a^2 - (b-R)^2}{2(L-a)} \tag{5.1}$$

$$\alpha = \frac{n\pi R}{L_b} = \frac{2n\pi R(L-a)}{L^2 - a^2 - (b-R)^2}. \tag{5.2}$$

These equations are derived in the box 4.

We assume that the only contribution to the energy of the system is the energy of the braid. Therefore, we write our energy function in the following way:

$$f_{4B}(R, \lambda_h^\star) = L_b(R)\mathscr{F}_b(R, \overline{\Delta\Phi}_{min}(R), \alpha(R), \lambda_h^\star) \tag{5.8}$$

Here, $\mathscr{F}_b(R, \overline{\Delta\Phi}, \alpha, \lambda_h^\star)$ is the energy density of free braids (see equation (2.1)). The energy function written in equation (5.8) can be re-expressed by using the equations that give α and L_b as a function of R. Therefore, it will depend only on the variable R, and additionally on $\lambda_h^\star$.

5.1.2 Model limitations

A list of the model limitations follows:

- The segments connecting the end of the braid to the beads are considered to be straight. In reality, they are fluctuating in a worm-like chain fashion, which could be described in a way similar to what was done in the case of single molecule supercoiling (see section 4.2.2). The main drawback from this assumption is that we are introducing an artificial effect in the system. In fact when the beads are pulled all of the length is transferred from the braid to the straight portions of the molecule, whereas in reality one would expect that there is an interplay between the braid energy and the worm-like chain entropic terms. Since we are assuming that the segments are straight, this means that the model will likely be more valid in the limit of a high force applied on the beads.

- The braid and the straight segments have discontinuous tangents. We expect that, similarly to the case described for the closed loops case (see section 3.2), when the braid length is very long, the transition region (in which the end tilt angle relaxes to the braid tilt angle) is likely to be small, and gives a negligible contribution to the total energy.

- The tilt angle within the braid is constant. As before, when there is a substantial electrostatic attraction, it may be expected that the tilt angle is primarily determined by a competition between the linking number con-

straint, and the bending and electrostatic energies in the braid. Therefore, in the limit of mild elastic stresses within the braid (small number of imposed turns), we expect this to be a good approximation.

- No buckling transition was taken into account. It may be expected that, similarly to the case of single molecule braiding experiments (Charvin et al., 2005), there is a buckling of the braid axis at high number of turns. The buckling transition occurs in conditions where our theory is not expected to be valid.

5.2 Results

We minimize the free energy function given in equation (5.8) to obtain the equilibrium conformation of the pair of DNA molecules. In this section we show the results of such minimization.

5.2.1 Energy landscape

Figure 5.2 shows the minimum energy of the four-bead system as a function of θ, at different values of the number of turns n. At low values of θ, there is always repulsion between DNA molecules (see also chapter 2). In repulsive regime, the equilibrium distance between the two molecules is such that the electrostatic energy contribution is negligible, and the only relevant energy term is bending.

As θ increases, the electrostatic energy gradually becomes attractive. When there is attraction, we expect that the equilibrium value of R will suddenly drop to about 24 Å, which is the optimum distance for the electrostatic interactions. We call such phenomenon "collapse" of the system.

The threshold value of θ at which attraction occurs depends on n and on whether the braid is formed by homologous or non-homologous molecules. For positive n the angle α is positive (see equation (5.2)), then an increasing value of positive n makes the electrostatic interactions in the braid become more favourable, and the threshold value of θ decreases. On the other hand, if n is negative then the braid tilt angle is negative, which is always unfavourable for the electrostatic

energy component. Therefore, increasing the value of negative n results in an increase of the threshold value of θ.

5.2.2 Forces

The forces acting on the beads may be calculated by differentiating the minimum energy of the system with respect to a or b. The theoretical estimate that one can make in this way addresses experimental results in which *two* beads are moved at the same time (symmetric pulling). If moving only one bead (asymmetric pulling), then the estimate must be made differently, and calculations in this direction are currently being performed by D.J. Lee and A. Korte. We show here the results for the force along the direction of b:

$$F_b = -\frac{\mathrm{d}f_{4B}}{\mathrm{d}b}. \tag{5.9}$$

Here, we show the behaviour of the force acting on the beads in the symmetric case. We start by showing its dependence on θ, keeping a and b fixed.

In the presence of attractive inter-DNA forces, the behaviour of the force changes discontinuously. The behaviour depends on the value of n, and on whether the molecules are homologous or non-homologous. Figure 5.3 shows a summary of the behaviour of the force along the direction of b, as a function of n, and for homologous and non-homologous molecules. At values of θ lower than the threshold, the force is small and negative (i.e., the energy decreases if the distance b decreases). In fact, as explained in the previous section, the only relevant energy component here is the bending energy, which is proportional to the tilt angle α. As b becomes lower, α becomes lower, so that the total energy decreases.

When the regime switches to attractive, the behaviour of the force curves is more complicated. One should keep in mind that α is proportional to b, but L_b is *inversely* proportional to b. Therefore, when increasing b, two opposite effects occur. The first is that the tilt angle in the braid becomes larger (in absolute value), and the second is that the braid becomes shorter. Because the tilt angle increases in magnitude, the energy *density* in the braid changes, and it becomes more or less favourable depending on the sign of n: at positive n, it becomes more

favourable; viceversa for negative n. The two opposite tendencies originate from the form of the energy function which is $E \sim L_b \mathscr{E}_b$.

For positive values of n, at values of θ slightly higher than the threshold, the force jumps to a positive value. Immediately after the threshold for attraction, the energy density in the braid is small, so the main tendency is to increase the energy density rather than to scale up the length of the braid. As θ increases, the force gradually decreases, until it becomes negative. In fact, for high values of θ the energy density in the braid is high, and it is more favourable to increase the length of the braid rather than increase the energy density in the braid. The height of the jump in the strength of the electrostatic attraction. Therefore, it increases with increasing positive n, and is higher for homologous molecules.

If n is negative, then the force decreases monotonically with θ, even after the threshold for attraction. That is because it is always unfavourable to increase the magnitude of the tilt angle in the braid.

5.3 Experiments to be performed

In this section we briefly illustrate a series of experiments that should be performed in order to test the predictions of the model. This section is deliberately left incomplete. In fact, the theory predicts that much information can be gained by experiments on the four-bead system. The first priority, however, is to ascertain whether the "collapse" that was discussed in the "Results" section occurs or not. Once the collapse is detected, it is also important to determine which are the conditions at which it occurs.

The hallmark of the collapse of the molecules is the presence of an unusually high value of the forces in the x and y directions, acting on the beads (see figure 5.3). Whether it will happen or not depends on the parameters θ, f_1 and f_2, of which we know very little. Therefore, it would be desireable to start by performing experiments in the conditions in which the chiral effects we are discussing are maximized: in the presence of DNA condensing counterions such as spermine, spermidine or cobalt-hexamine. In such conditions we expect that $f_1 + f_2 \approx 1$ and θ is close to 1. Then, a large force signal should be detected on the beads.

Once that the first experiment using DNA condensers establishes the validity

of the model, then a systematic study of the behaviour of the force curves in the presence of other ion types should be performed. For the reasons explained in the previous chapters, a very interesting case is the one of magnesium and calcium.

Experiments of this type also have the potential to fully study the difference between homologous and non-homologous DNA-DNA interactions. Our theory predicts that homologous molecules will always be more stable than non-homologous molecules, when they attract each others. If two identical DNA molecules are put on the beads, and *they are allowed to slide perfectly in register with each other*, then this will be a configuration probing the homologous DNA-DNA interactions. On the other hand, if the same identical DNA molecules are put in an antiparallel configuration, this will be probing non-homologous interactions (Kornyshev and Wynveen, 2009).

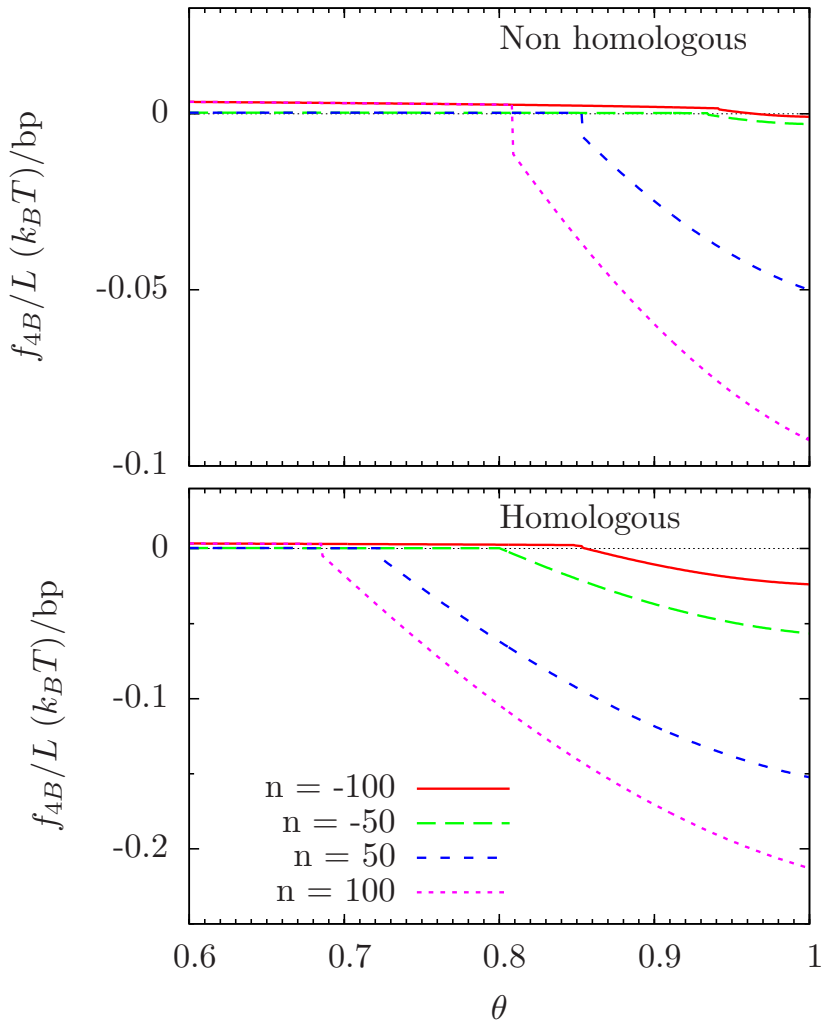


Fig. 5.2: Minimum energy per base pair of the system as a function of the charge compensation parameter θ. The plots were obtained for $\lambda_D = 7.0$Å, $f_1 = 0.4$, $f_2 = 0.6$.

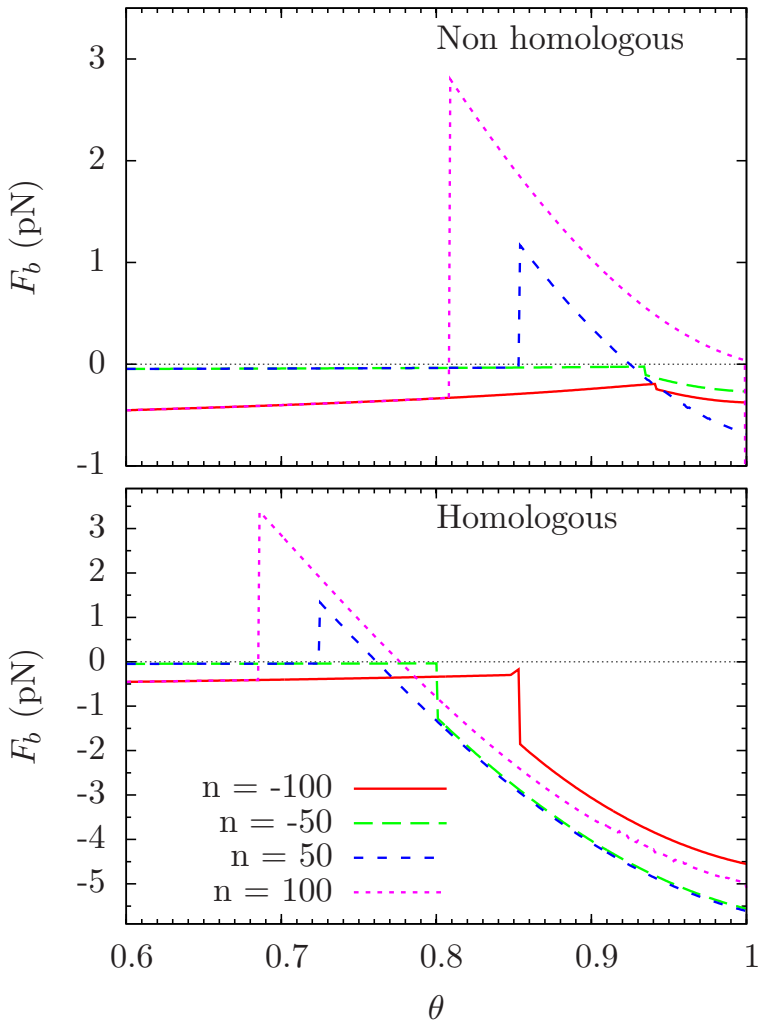


Fig. 5.3: Predicted values of the force along the y direction acting on the beads. The forces were calculated using the same parameters as in figure 5.2.

Conclusions

In this thesis, we presented the derivation and the predictions of a chiral theory of DNA supercoiling. First, we developed a theory of braiding of two DNA molecules, which was based on the Kornyshev–Leikin theory of DNA-DNA interactions, which takes into account the helical patterns of charge distributions on the molecular surfaces. We then applied the theory to the cases of closed loop DNA supercoiling, and to two different kinds of single-molecule DNA manipulations.

The results of the theory clearly indicate that the chirality of the charge distributions has a significant effect on the energy landscape of two interacting DNA molecules. The right-handedness of the DNA double helix results in the electrostatic stabilization of left-handed braids, which occur in positively supercoiled molecules. We proposed that the electrostatic stabilization of braids of DNA molecules stands behind the recent observation of homologous pairing in monovalent salt. We speculated that these effects might play an important role in several biologically relevant cases: the chiral discrimination of the topoisomerase enzymes; the presence of positively supercoiled molecules in hyperthermophilic bacteria and archea; the occurrence of tightly supercoiled structures observed *in vitro*. Such effects also suggest novel interpretations of experiments that proved that DNA overwinds in the presence of divalent metal ions.

Careful comparison of the theoretical results with available observations reported in the literature reveal the at present more experimental studies are needed to assess the validity of the theory. In some cases, the available data is inconclusive; in other cases, it can be interpreted as a sign of the existence of significant chiral effects; finally, some observations are opposite of our theoretical predictions.

More experimental data will soon be available from the laboratories of Dr. Tim Albrecht at Imperial College London, and of Professor Gijs Wuite in Vrije

Universiteit Amsterdam.

Our theoretical estimates suffer from the limitation that they heavily depend on unknown parameters. At the moment, simulation studies are being performed in collaboration with Professor Jeremy Smith and Dr. Xiaolin Cheng at the Oak Ridge National Laboratories.

Appendix A

The geometry of helices

This appendix derives all the basic formulae concerning the helical geometry that are used throughout this thesis. We will start by describing the helical curves and their properties, such as pitch and curvature. In the final part we report a sketch of the derivation of the formula for the writhe of a braid.

A.1 Helices

A.1.1 Definition

Given a set of cartesian coordinates in $\mathbb{R}^3$, the following curve

$$\mathbf{r}(s) : [0, L] \to \mathbb{R}^3$$

$$\mathbf{r}(s) = \begin{pmatrix} r\cos(\omega s + \psi) \\ r\sin(\omega s + \psi) \\ \frac{P\omega}{2\pi} s \end{pmatrix} \tag{A.1}$$

is a helix with the z axis as the helical axis. Here,

- L is the total arc-length of the curve.
- s is the curve parameter. It will be shown in the next subsection that the parametrization of the curve is chosen in such a way that s is the so-called "natural" parametrization ($\int_0^L \mathrm{d}s|\mathrm{d}\mathbf{r}/\mathrm{d}s| = L$).

- ω is the angular frequency of precession of the curve around the helix axis.
- ψ is the angle that the curves form with the xy plane at $z = 0$.
- r is the helix radius.
- P is the helical pitch.

The definition of the curve is made in such a way that when $s = 2\pi/\omega$ (one full rotation around the z axis) the value of z is P.

The same curve can be obtained through a rotation of the vector

$$\mathbf{r}_0(s) = \begin{pmatrix} r \\ 0 \\ \frac{P\omega}{2\pi} s \end{pmatrix}. \tag{A.2}$$

Consider, in fact, the rotation given by

$$\mathbf{H}(s) = \begin{pmatrix} \cos(\omega s + \psi) & -\sin(\omega s + \psi) & 0 \\ \sin(\omega s + \psi) & \cos(\omega s + \psi) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{A.3}$$

It is readily seen that

$$\mathbf{r}(s) = \mathbf{H}(s)\mathbf{r}_0(s). \tag{A.4}$$

It is important to notice that there is a sign definition here. For positive values of ω, the helix is right-handed, and it is left-handed for negative values of ω.

A.1.2 Tangent vector

The tangent vector to the curve defined in equation (A.1) is given by

$$\hat{\mathbf{t}}(s) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \begin{pmatrix} -\omega r \sin(\omega s + \psi) \\ \omega r \cos(\omega s + \psi) \\ \frac{P\omega}{2\pi} \end{pmatrix}. \tag{A.5}$$

The requirement that norm of the tangent vector is unitary gives

$$|\hat{\mathbf{t}}|^2 = \omega^2 r^2 + \frac{P^2\omega^2}{(2\pi)^2} = 1 \Rightarrow \omega = \pm\left(r^2 + \frac{P^2}{(2\pi)^2}\right)^{-1/2}. \tag{A.6}$$

Now we show that the same tangent vector can be obtained through a rotation of the unit vector

$$\hat{\mathbf{t}}_\alpha = \begin{pmatrix} 0 \\ -\sin\alpha \\ \cos\alpha \end{pmatrix}. \tag{A.7}$$

We can see easily that

$$\hat{\mathbf{t}}(s) = \mathbf{H}(s)\hat{\mathbf{t}}_\alpha. \tag{A.8}$$

In this way, by setting $\psi = 0$, we can easily identify

$$\omega r = -\sin\alpha \tag{A.9}$$

and

$$\frac{P\omega}{2\pi} = \cos\alpha. \tag{A.10}$$

The angle α has a clear geometrical interpretation, as being the angle that the tangent vector forms with the z axis, at any height. It thus coincides with the α angle defined in figure 2.1. Notice that the sign convention for α is opposite of the one used for ω: α is positive for right-handed helices and negative for left-handed ones. The reason for this choice is to be consistent with the definition of positive and negative *crossovers*, which will be more clear in section 2.1.

A.1.3 Other useful relationships

From equation (A.10), it is seen that the pitch of the helix is given by

$$P = \frac{2\pi\cos\alpha}{\omega}. \tag{A.11}$$

It is also useful to notice that, from equations (A.9) and (A.10), we obtain

$$\frac{P}{2\pi r} = \tan\alpha. \tag{A.12}$$

This is the fundamental identity that relates all the geometrical parameters of the helix to each others.

Given a curve defined by $\mathbf{r}(s)$, the *curvature* is given by $\kappa_c = \mathrm{d}^2\mathbf{r}/\mathrm{d}s^2$. We can calculate the curvature of the helix by using equations (A.5) and (A.9):

$$\kappa_c = \left|\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}s^2}\right| = |\omega^2 r| = \frac{\sin^2\alpha}{r}. \tag{A.13}$$

The curvature is a second-order term in α.

A.2 Writhe of a braid

In this section we want to calculate the writhe of a symmetric, homogeneous braid formed by two helices with $\alpha \ll 1$, $R \ll L_b$, $R \ll P$. Here, R is the braid diameter.

First, we need to define the curve. We suppose that the curve is formed by two helices, Γ_1 and Γ_2, that are very long and are connected to each other by a segment of negligible length, that contributes with a negligible contribution to the writhe. We start from the mathematical definition of the writhe, given in equation (b.3.7), and within these approximations, the writhe is given by

$$Wr_b = \frac{2\alpha}{R} \tag{A.14}$$

Appendix B

Theory of electrostatic interactions in DNA braids

In this appendix we describe how the calculation of the electrostatic interaction energy of a pair of braided DNA molecules was performed. The energy function of a braid consists of three terms: the parallel molecule term, $\mathscr{E}_0$, which was originally calculated in (Kornyshev and Leikin, 1997), the chiral torque, $\mathscr{E}_1$, which we will calculate here, and the bending energy term, which is described in box 2.

The calculation of the interaction energy of a braid was sketched by Dr. Dominic Lee, and performed by myself. Most of its description has been published in our research paper (Cortini et al., 2011) (particularly in its web-appended Supplementary Information), with the exception of how the image torque (see equation (2.6)) was derived. The calculation of the $\mathscr{E}_0$ term, which coincides with the energy of a braid at $\alpha = 0$, was first performed in (Kornyshev and Leikin, 1997), and the result will be recovered here using a different theoretical method.

This appendix is divided into two sections. The first section describes the derivation of the energy function of a pair of *ideal* DNA molecules in a braid. The second part will be devoted to the calculation of the interaction energy for nonideal DNA molecules.

B.1 Ideal DNA molecules

We will describe here the derivation of equations (2.3) and (2.4). The basic idea of the calculation was described in section 2.1. We shall proceed in several steps. First, we give the general idea of how we performed the calculation, and what are the relevant assumptions that we make. Then, we write the equations that describe the geometry of the system (section B.1.2). We then derive equations for the Fourier transforms of the induced charge densities and potentials for arbitrary sets of surface charge patterns. We then specify the charge pattern to the one that describes DNA. Once the the calculation of the Fourier transforms of the real and image charge patterns are known, we proceed with the calculation of the energy, using equation (1.14).

B.1.1 General strategy

Our objective is the calculation of the energy integral which was given in equation (1.14). We therefore need to express the Fourier transforms of the real and induced charge densities. The induced charges (often referred to as "image" charges) are located at the dielectric boundaries of the molecules, and ensure that the boundary conditions (see equations (1.9a) and (1.9b)) are fulfilled. The Fourier transform of the real charges is relatively straightforward to calculate, and the calculation was already performed by Crick (1953), and reported in the Supplementary Material to (Kornyshev et al., 2007). The challenging part is the calculation of the image charges.

It was found in the calculation of the interaction energy of parallel DNA molecules (Kornyshev and Leikin, 1997) that the image charge densities are in the form of infinite series. It was also found, however, that such series can be truncated to the first few terms (for complete details, see box 5):

$$E(R) = E_{1,2}(R) + E_{1,1}(R) + E_{2,2}(R). \tag{B.1}$$

Box 5: Image charge series

The image charges located at the dielectric boundary of two interacting molecules may be represented by the following series:

$$\rho_{\nu}^{ind} = \sum_{n=0}^{\infty} \left[\rho_{\nu,\nu}^{2n} + \rho_{\nu,\mu}^{2n+1}\right]. \tag{b.2.9}$$

The term $\rho_{\nu,\nu}^{(0)}$ represents the image charges located on the surface of molecule ν, which are created in the absence of the dielectric core of molecule μ. When introducing the dielectric boundary of molecule μ, the system of surface charges on molecule ν, given by $\rho_{\nu} + \rho_{\nu,\nu}^{(0)}$, creates a system of induced charges which we call $\rho_{\mu,\nu}^{(1)}$. The field given by such charges then induces again a system of induced charges on the first molecule, which we call $\rho_{\nu,\nu}^{(2)}$. This is the idea expressed in equation (b.2.9).

Here, the term

$$E_{1,2}(R) = \frac{4\pi}{\varepsilon} \int \mathrm{d}\mathbf{k} \frac{\left[\tilde{\rho}_1(\mathbf{k}) + \tilde{\rho}_{1,1}^{(0)}(\mathbf{k})\right]\left[\tilde{\rho}_2(-\mathbf{k}) + \tilde{\rho}_{2,2}^{(0)}(-\mathbf{k})\right]}{k^2 + \kappa^2} \tag{B.2}$$

represents the interaction energy of fixed charges on the surface of molecule 1 in the field created by the fixed charges on the surface of molecule 2. Such energy is equal to the energy of fixed charges on molecule 2 in the field created by the charges on the surface of molecule 1. The terms

$$E_{\nu,\nu}(R) = \frac{2\pi}{\varepsilon} \int \mathrm{d}\mathbf{k} \frac{\tilde{\rho}_{\nu,\mu}^{(1)}(\mathbf{k})\left[\tilde{\rho}_{\mu}(-\mathbf{k}) + \tilde{\rho}_{\mu,\mu}^{(0)}(-\mathbf{k})\right]}{k^2 + \kappa^2} \qquad \nu \neq \mu \tag{B.3}$$

represent the energy of the dielectric core of molecule ν in the electric field created by molecule μ.

For parallel molecules, it was found that the image charge series admits a fully analytical solution. Here, due to the fact that the two molecules are braided, an analytical solution cannot be found for any value of α. Therefore, we express the

interaction energy as an expansion in small α:

$$E(R) \approx E_0(R) + \frac{\partial E(R)}{\partial \alpha}\alpha \equiv E_0(R) + \alpha E_1(R) \tag{B.4}$$

where $E_0(R)$ is the interaction energy of a braid at $\alpha = 0$, and $E_1(R)$ is the chiral torque.

To calculate the energy, we then need to express the image charges given by $\rho^{(0)}_{\nu,\nu}$ (zero-order charges) and $\rho^{(1)}_{\nu,\mu}$ (first-order images). To find the system of charges induced on the molecular surface, we need to impose boundary conditions. That is, we must solve the system of equations given by

$$\begin{cases} \tilde{\varphi}_\nu \left[R_\nu, k'_z, m; \tilde{\rho}_\nu(\mathbf{k}) + \tilde{\rho}^{ind}_\nu(\mathbf{k})\right]\big|_{R_\nu=a} = \tilde{\varphi}_{\nu,in}(R_\nu, k'_z, m)|_{R_\nu=a} \\ \varepsilon_w \dfrac{\partial}{\partial R_\nu}\tilde{\varphi}_\nu \left[R_\nu, k'_z, m; \tilde{\rho}_\nu(\mathbf{k}) + \tilde{\rho}^{ind}_\nu(\mathbf{k})\right]\Big|_{R_\nu=a} = \varepsilon_c \dfrac{\partial}{\partial R_\nu}\tilde{\varphi}_{\nu,in}(R_\nu, k'_z, m)\Big|_{R_\nu=a} . \end{cases}$$

(B.5a)

(B.5b)

where we adopted the following notation: $\tilde{\varphi}_\nu \left[R'_\nu, k'_z, m; \tilde{\rho}(\mathbf{k})\right]$ is the cylindrical Fourier transform of the field due to the charge density $\tilde{\rho}(\mathbf{k})$. The subscript ν in the potential φ denotes that the field is expressed in the frame of reference of cylinder ν. The above system of equations was written for the total field, generated by the total charge density. We can use the linearity of the Debye-Hückel equation to write the same equations separately for the components of the field generated by successive orders of image charge distributions.

The potential inside the core of molecule ν, given by $\varphi_{\nu,in}$, is found by solving the Laplace equation (see equation (1.10)). For the cylindrical Fourier transform, it is possible to prove that

$$\tilde{\varphi}_{\nu,in}(R'_\nu, k'_z, m) = C(k'_z, m) I_m(k'_z R'_\nu), \tag{B.6}$$

where $C(k'_z, m)$ is an arbitrary coefficient, which needs to be found by imposing boundary conditions (Abramowitz and Stegun, 1964).

Therefore, to find the zero- and first-order images, we proceed as follows:

1. Define "moving" frames of reference locally associated with each of the two molecules.

2. Express the coordinates of a point charge at the surface of molecule ν, in the local frame of reference.

3. Calculate the Fourier transform of the field generated by the point charge.

4. Impose boundary conditions on the surface of molecule ν, thereby finding the zero-order image charges for a point charge.

5. Transform the field due to real and zero-order images located on the surface of molecule ν to the frame of reference of molecule μ, and impose boundary conditions on the surface of molecule μ. We find the first-order images for a point charge in such way.

6. Transform the Fourier transforms of the charge densities into the lab frame, and sum over all charges of the system, thereby obtaining these expressions for arbitrary patterns of surface charges. We obtain the total zero- and first-order image charge densities in this way.

7. Specify the charge density to the case of an ideal DNA molecule.

Once the Fourier transforms of the charge densities are obtained, we will proceed with the calculation of the integrals in equations (B.2) and (B.3), by making the small-α approximation.

B.1.2 Geometry

We define multiple reference frames. The **lab frame r**, with the braid axis being coincident with the z axis, with Cartesian coordinates. The **local frames**, associated with the two molecules, $\mathbf{r}'_{\nu,j}$, associated with each charge j on the surface of molecule ν, with Cartesian coordinates. We then use also a cylindrical coordinate frame associated with the local frames, defined by $(R', z', \phi')_{\nu,j}$.

The local frames are defined so that:

- the origin lies on the molecular centreline.

- the z' axis is tangential to the molecular centreline.
- the charge j lies on the $z' = 0$ plane.
- the x' ($\phi' = 0$) axis points away from the braid axis.

Unlike what was done for the calculation of the interaction energy of two skewed cylinders (Kornyshev and Leikin, 2000), it is more convenient to generate one system of reference from the other through a *rotation* rather than a translation. In such way, the two cylindrical coordinate frames are equivalent, and the potential calculated in one frame has the same form than the one calculated in the other frame.

We now derive the transformation laws between the coordinate frames just defined.

$$\mathbf{r} = \mathbf{r}^c_\nu(s_j) + \mathbf{T}_{\nu,j}\mathbf{T}_{-\alpha}\mathbf{r}'_{\nu,j}, \tag{B.7}$$

where we defined:

$$\mathbf{r}^c_\nu(s_j) = \begin{pmatrix} b\cos\omega_{\nu,j} \\ b\sin\omega_{\nu,j} \\ s_j\cos\alpha \end{pmatrix}, \tag{B.8}$$

$$\mathbf{T}_{\nu,j} = \begin{pmatrix} \cos\omega_{\nu,j} & -\sin\omega_{\nu,j} & 0 \\ \sin\omega_{\nu,j} & \cos\omega_{\nu,j} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{B.9}$$

and

$$\mathbf{T}_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}. \tag{B.10}$$

The vector $\mathbf{r}^c_\nu(s_j)$ is the curve that describes the centreline of molecule ν. We defined also $b = R/2$. For a homogeneous braid, the angle $\omega_{\nu,j}$ is given by

$$\omega_{\nu,j} = Q_c s_j + \psi_\nu \approx Q s_j + \psi_\nu. \tag{B.11}$$

Here, we defined the angle ψ_ν, which is the azimuthal orientation of the centreline of molecule ν when intersecting the xy plane of the lab frame, $Q_c = 2\pi\cos\alpha/P$, and $Q = 2\pi/P$. Also, in equation (B.11) we defined P as the helical pitch of the

molecular centrelines (see appendix A for definitions and additional details). In the second term on the right-hand side of equation (B.7), the two rotation matrices realign the axes of the local frames so that they are always consistently defined.

The vector $\mathbf{r}_\nu^c$ may also be written in a form which will be convenient to use later:

$$\mathbf{r}_\nu^c(s_j) = \mathbf{T}_{\nu,j}\mathbf{r}_0^c, \tag{B.12}$$

where

$$\mathbf{r}_\nu^c(s_j) = \begin{pmatrix} b \\ 0 \\ s_j \cos\alpha \end{pmatrix}. \tag{B.13}$$

Using equation (B.7), we can write:

$$\mathbf{T}_{\nu,j}\mathbf{r}_0^c + \mathbf{T}_{\nu,j}\mathbf{T}_{-\alpha}\mathbf{r}'_\nu = \mathbf{T}_{\mu,j}\mathbf{r}_0^c + \mathbf{T}_{\mu,j}\mathbf{T}_{-\alpha}\mathbf{r}'_\mu. \tag{B.14}$$

Taking into account the symmetric braid condition, which states that $\psi_\mu = \psi_\nu + \pi$, this equation leads to the transformation law between the two cylindrical systems of reference:

$$\mathbf{r}'_\nu = -\mathbf{T}_\alpha\mathbf{R} + \mathbf{T}_\alpha\mathbf{S}\mathbf{T}_{-\alpha}\mathbf{r}'_\mu, \tag{B.15}$$

where

$$\mathbf{S} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{B.16}$$

and $\mathbf{R} = (R, 0, 0)$ (recall that $R = 2b$).

B.1.3 Fourier transforms of potentials and charge densities for arbitrary surface charge patterns

The Fourier transform of a system of arbitrary surface charges $\sigma_\nu(z'_\nu, \phi'_\nu; j)$ is conveniently expressed in cylindrical coordinates $\mathbf{k} \equiv (K, \phi_K, q)$. After the calculation is performed, the result is the following:

$$\tilde{\rho}_\nu(K, \phi_k, q; j) = \frac{a}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} i^n J_n(Ka)\tilde{\sigma}_\nu(n, q; j)e^{-in\phi_k}, \tag{B.17}$$

where $J_n(x)$ are the Bessel functions (Abramowitz and Stegun, 1964). The derivation of this formula is reported in Box 6.

For a point charge located at the surface of the cylinder, the surface charge pattern is expressed as

$$\sigma_\nu(z', \phi'; j) = \frac{e_0 q_j}{a} \delta(z') \delta\left(\phi' - \phi'_{\nu,j}\right), \tag{B.18}$$

where e_0 is the elementary charge (taken to be a positive number), q_j is the valence of the charge, and $\phi'_{\nu,j}$ is the azimuthal orientation of the point charge[1]. The cylindrical Fourier transform of such charge distribution is given by

$$\tilde{\sigma}_\nu(k'_z, m; j) = \frac{e_0 q_j}{2\pi a} e^{im\phi'_{\nu,j}}. \tag{B.19}$$

Substituting equation (B.19) into equation (B.17), we can express the Fourier transform of the charge density of a point charge at the surface. Since the charge distribution is point-like, the Fourier transform does not depend upon k'_z.

To end this part, we provide equations that are used to relate the Fourier transforms of a scalar field that are calculated in different reference frames. For scalar functions we have

$$f_\mu(\mathbf{r}'_{\mu,j}) = f_\nu(\mathbf{r}'_{\nu,j}). \tag{B.20}$$

Then, applying the definition of the Fourier transform (equation (b.1.1)), equation (B.20), and equation (B.15), we obtain

$$\tilde{f}_\mu(\mathbf{k}) = \exp(i\mathbf{S}\mathbf{T}_{-\alpha}\mathbf{k} \cdot \mathbf{R}) \tilde{f}_\nu(\mathbf{T}_\alpha \mathbf{S}\mathbf{T}_{-\alpha}\mathbf{k}). \tag{B.21}$$

[1]The a^{-1} in equation (B.18) derives from the normalization condition $\int \mathrm{d}\mathbf{r}\rho(\mathbf{r}) = e_0 q_j$.

Box 6: Derivation of equation (B.17)

To derive equation (B.17), we first need to express the scalar product between $\mathbf{k}$ and $\mathbf{r}$ in their respective cylindrical components. After straightforward algebra, we obtain this expression

$$\mathbf{k}\cdot\mathbf{r} = Kr\cos(\phi-\phi_K)+qz. \tag{b.2.10}$$

We then express the Fourier transform in cylindrical coordinates, using the definition (b.1.1) and equation (b.2.10):

$$\begin{aligned}\hat{\rho}(\mathbf{k}) &= \frac{1}{(2\pi)^{3/2}}\int \mathrm{d}\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r}) = \\ &= \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\mathrm{d}z e^{iqz}\int_0^{2\pi}\mathrm{d}\phi\int_0^{\infty} r\mathrm{d}r e^{iKR\cos(\phi-\phi_K)}\delta(r-a)\sigma(\phi,z)\end{aligned} \tag{b.2.11}$$

Here we can calculate the integral over R easily, using the Dirac delta integration. We also use the Jacobi-Anger identity:

$$\exp(iA\cos\alpha) = \sum_{m=-\infty}^{\infty} i^m J_m(A)e^{-im\alpha}, \tag{b.2.12}$$

where A and α are real numbers. We then obtain:

$$\hat{\rho}(\mathbf{k}) = \frac{a}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\mathrm{d}z e^{iqz}\int_0^{2\pi}\mathrm{d}\phi\sum_{m=-\infty}^{\infty} i^m J_m(Ka)e^{-im(\phi_K-\phi)}\sigma(\phi,z). \tag{b.2.13}$$

Once this is obtained, we represent the function $\sigma(\phi,z)$ in terms of its cylindrical Fourier components (see equation (b.1.4)). Having done that, the integrals become trivial, as the only integrals to be calculated are of the types:

$$\int_0^{2\pi}\mathrm{d}\phi\exp(i(n-m)\phi) = 2\pi\delta_{n,m}, \tag{b.2.14}$$

where $\delta_{n,m}$ is the Kronecker delta, and

$$\int_{-\infty}^{\infty}\mathrm{d}z\exp(i(q-k_z)z) = 2\pi\delta(q-k_z). \tag{b.2.15}$$

Using equations (b.2.14) and (b.2.15), equation (B.17) is obtained.

Box 7: Derivation of equation (B.22)

We apply the definition of the cylindrical Fourier transform (equation (b.1.3)) to the field generated by the charge distribution described by equation (B.17):

$$\tilde{\varphi}_\nu(R_\nu, q, n; j) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi'_\nu e^{in\phi'_\nu} \int_{-\infty}^{\infty} \mathrm{d}z'_\nu e^{iqz'_\nu} \int \mathrm{d}\mathbf{k} e^{-i\mathbf{r}'_\nu\cdot\mathbf{k}} \frac{4\pi}{\varepsilon}\frac{\tilde{\rho}_\nu(\mathbf{k}; j)}{k^2+\kappa^2} \quad \text{(b.2.16)}$$

We then express $\tilde{\rho}_\nu(\mathbf{k}; j)$ in terms of equation (B.17), and calculate the trivial integrals and sums in a similar way to what was described in box 6. We are left with this expression:

$$\tilde{\rho}_\nu(R'_\nu, q, n; j) = \frac{4\pi a}{\varepsilon}\tilde{\sigma}_\nu(q, n; j)\int_0^{\infty} \mathrm{d}K \frac{K}{K^2+q^2+\kappa^2} J_n(Ka)J_{-n}(KR'_\nu) \quad \text{(b.2.17)}$$

The integral here may be solved with the help of formula 8.11.51 from (Erdélyi and Bateman, 1954):

$$\int_0^{\infty} \mathrm{d}K \frac{K^n J_n(Ka)J_m(Ka)J_{n-m}(Kb)}{K^2+a^2} \quad \text{(b.2.18)}$$

The resulting expression is equation (B.22).

B.1.4 Zero-order image charges

To find the zero-order images, we need to impose boundary conditions to the potential due to the point charge on the surface of cylinder ν. The cylindrical Fourier transform of the field generated by the point charge distribution $\tilde{\sigma}_\nu(q, n; j)$ is given by (box 7):

$$\tilde{\varphi}_\nu\left[R'_\nu, q, n; \tilde{\rho}_\nu(\mathbf{k}; j)\right] = \frac{4\pi a}{\varepsilon} I_n(\kappa_q a)K_n(\kappa_q R'_\nu)\frac{e_0 q_j}{2\pi a} e^{in\phi'_{\nu,j}}, \quad \text{(B.22)}$$

where

$$\kappa_q = \sqrt{\kappa^2+q^2}. \quad \text{(B.23)}$$

To obtain the zero-order image charges, we add a system of charges $\tilde{\sigma}^{(0)}_{\nu,\nu}$ to the surface of molecule ν. Then, we solve equations (B.5a) and (B.5b) for the total potential given by the system of real and zero-order image charges. We then obtain a system of two linear equations, with $C(k'_z, m)$ and $\tilde{\sigma}^{(0)}_{\nu,\nu}$ as unknowns.

After straightforward algebra, we obtain the following solution:

$$\tilde{\sigma}^{(0)}_{\nu,\nu}(k'_z, m; j) \approx t_m(k'_z)\tilde{\sigma}_\nu(k'_z, m; j) \tag{B.24}$$

where

$$t_m(x) = \frac{K_m(\kappa_m a)I'_m(\kappa_m a)}{I_m(\kappa_m a)K'_m(\kappa_m a)}. \tag{B.25}$$

This equation was obtained under the approximation that $\varepsilon_w \gg \varepsilon_c$. This completes the calculation of zero-order images.

B.1.5 First-order images

The calculation of first-order images is slightly more involved than the one needed to calculate the zero-order images. We need to transform the field generated by $\sigma_\nu + \sigma^{(0)}_{\nu,\nu}$ to the frame of reference of molecule μ, then impose boundary conditions on its surface. To express the Fourier transform of this field in the other local frame of reference, we use equations (1.12), (B.17), and (B.21):

$$\tilde{\varphi}_\mu\left[\mathbf{k}; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right] = \frac{4\pi}{\varepsilon}\frac{\tilde{\rho}_\nu(\mathbf{T}_{-\alpha}\mathbf{S}\mathbf{T}_\alpha\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{T}_{-\alpha}\mathbf{S}\mathbf{T}_\alpha\mathbf{k}; j)}{k^2+\kappa^2}e^{i\mathbf{S}\mathbf{T}_{-\alpha}\mathbf{k}} =$$
$$= \frac{4\pi a}{\varepsilon\sqrt{2\pi}}\sum_{m=-\infty}^{\infty}\frac{i^m J_m(\hat{K}a)e^{im\hat{\phi}_K}}{K^2+k_z^2+\kappa^2}\left[1 + t_m(\hat{k}_z, m)\right]\tilde{\sigma}_\nu(\hat{k}_z, m)e^{i\tilde{\mathbf{k}}_\alpha\cdot\mathbf{R}} \tag{B.26}$$

where we defined:

$$(\tilde{K}, \tilde{\phi}_K, \tilde{k}_z) \equiv \tilde{\mathbf{k}}_\alpha \equiv \mathbf{S}\mathbf{T}_{-\alpha}\mathbf{k} \tag{B.27}$$

and

$$(\hat{K}, \hat{\phi}_K, \hat{k}_z) \equiv \hat{\mathbf{k}}_\alpha \equiv \mathbf{T}_\alpha\mathbf{S}\mathbf{T}_{-\alpha}\mathbf{k} = \mathbf{T}_\alpha\tilde{\mathbf{k}}_\alpha. \tag{B.28}$$

In equation (B.26) we took into account that the rotation matrix $\mathbf{T}_{-\alpha}\mathbf{S}\mathbf{T}_\alpha$ does not change the norm of $\mathbf{k}$. We then use the definition of the cylindrical Fourier

given in equation (b.1.3) to obtain the following expression:

$$\tilde{\varphi}_\mu\left[R_\mu, q, n; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right] = \frac{1}{(2\pi)^{5/2}} \int_0^{2\pi} \mathrm{d}\phi'_\mu e^{in\phi'_\mu} \int_{-\infty}^{\infty} \mathrm{d}z'_\mu e^{iqz'_\mu} \times \\ \times \int \mathrm{d}\mathbf{k} e^{i\mathbf{r}'_\mu \cdot \mathbf{k}} \tilde{\varphi}_\mu\left[\mathbf{k}; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right]. \quad \text{(B.29)}$$

The above expression is very difficult (if not impossible) to calculate without approximations. In fact, the function to integrate, expressed in equation (B.26), contains complex nonlinear expressions of the vector $\mathbf{k}$. Therefore, we expand the integrand in small α and calculate an approximate expression:

$$\tilde{\varphi}_\mu\left[R_\mu, q, n; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right] \approx \tilde{\varphi}_{\mu,0}\left[R_\mu, q, n; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right] + \\ + \alpha\tilde{\varphi}_{\mu,1}\left[R_\mu, q, n; \tilde{\rho}_\nu(\mathbf{k}; j) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}; j)\right]. \quad \text{(B.30)}$$

By expanding the integrand in equation (B.29) according to equation (B.30), we may obtain the first-order images after calculating the intergral using the formulas that were already used in the derivation of B.17 (see box 6). After imposing the boundary conditions of the resulting expressions, also the first-order image charge densities may be expressed by an expansion in small α:

$$\tilde{\sigma}^{(1)}_{\mu,\nu}(q, n; j) \approx \tilde{\sigma}^{(1)}_{\nu,0}(q, n; j) + \alpha\tilde{\sigma}^{(1)}_{\nu,1}(q, n; j) \quad \text{(B.31)}$$

$$\tilde{\sigma}^{(1)}_{0,\mu,\nu}(q, n; j) = \sum_{m=-\infty}^{\infty} w_{0,n,m}(q)\tilde{\sigma}_\nu(q, m; j) \quad \text{(B.32)}$$

$$\tilde{\sigma}^{(1)}_{1,\mu,\nu}(q, n; j) = \sum_{m=-\infty}^{\infty} w_{1,n,m}(q)\tilde{\sigma}_\nu(q, m; j), \quad \text{(B.33)}$$

where the coefficients of the sums are given by

$$w^{(0)}_{n,m}(q) = (-1)^{n-m}\frac{I'_n(\kappa_q a)K_{n-m}(\kappa_q R)}{\kappa_q a I'_n(\kappa_q a)K'_n(\kappa_q a)K'_m(\kappa_q a)} \tag{B.34}$$

$$\begin{aligned} w^{(1)}_{n,m}(q) = &-(-1)^{n-m}\frac{I'_n(\kappa_q a)}{I'_n(\kappa_q a)K'_n(\kappa_q a)}\times \\ &\times\Big\{qa\,[1+t_m(q)]\,[K_{n-m+1}(\kappa_q R)I_{m-1}(\kappa_q a) - K_{n-m-1}(\kappa_q R)I_{m+1}(\kappa_q a)] + \\ &-\kappa_q I_m(\kappa_q a)\,[K_{n-m+1}(\kappa_q R) - K_{n-m-1}(\kappa_q R)]\,\frac{\partial t_m(q)}{\partial q}\Big\} \end{aligned} \tag{B.35}$$

The above equations complete the calculation of the first-order images.

B.1.6 Structure factors

We now need to sum over all the charges present in the system. Since we want to express the Fourier transform of the charge densities *in the lab frame*, we need to relate the Fourier transform of a function calculated in the local frame to the one calculated in the lab frame: to do so, we substitute equation (B.7) into the definition of the Fourier transform (equation (b.1.1)).

$$\tilde{\rho}_\nu(\mathbf{k}) = \sum_j \exp\left[i\mathbf{k}\cdot\mathbf{r}^c_\nu(s_j)\right]\tilde{\rho}_\nu(\mathbf{T}^{-1}_\alpha\mathbf{T}^{-1}_{\nu,j}\mathbf{k};j). \tag{B.36}$$

The relationship between $\rho^{(0)}_{\nu,\nu}(\mathbf{k})$ and $\rho^{(0)}_{\nu,\nu}(\mathbf{k};j)$ is the same. First, we calculate

$$\mathbf{k}\cdot\mathbf{r}^c_\nu(s_j) = Kb\cos(\phi_K-\omega_{\nu,j}) + k_z\cos\alpha s_j \approx Kb\cos(\phi_K-\omega_{\nu,j}) + k_z s_j, \tag{B.37}$$

to the first order in α. The next step is to express $\tilde{\rho}_\nu(\mathbf{T}^{-1}_\alpha\mathbf{T}^{-1}_{\nu,j}\mathbf{k};j)$. To do so, we write the components of the rotated $\mathbf{k}$ vector to the first order in α, by using the expressions for the rotation matrices given in equations (B.9) and (B.10):

$$\mathbf{T}^{-1}_\alpha\mathbf{T}^{-1}_{\nu,j}\mathbf{k} = \begin{pmatrix} K\cos(\phi_K-\omega_{\nu,j}) \\ K\cos\alpha\sin(\phi_K-\omega_{\nu,j}) + k_z\sin\alpha \\ -K\sin\alpha\sin(\phi_K-\omega_{\nu,j}) + k_z\cos\alpha \end{pmatrix} \approx \begin{pmatrix} K\cos(\phi_K-\omega_{\nu,j}) \\ K\sin(\phi_K-\omega_{\nu,j}) + k_z\alpha \\ -K\alpha\sin(\phi_K-\omega_{\nu,j}) + k_z \end{pmatrix}. \tag{B.38}$$

We may then apply the definitions of the inverse Fourier transform and the inverse cylindrical Fourier transform (see equations (b.1.2) and (b.1.4)), and take into account equation (B.38), which results in the following expression:

$$\tilde{\rho}_\nu(\mathbf{T}_\alpha^{-1}\mathbf{T}_{\nu,j}^{-1}\mathbf{k};j) = \frac{a}{(2\pi)^{3/2}}\int_0^{2\pi} \mathrm{d}\phi' \sum_{m=-\infty}^{\infty} e^{iKa\cos(\phi_K-\omega_{\nu,j}-\phi')+ik_z a\alpha\sin\phi'-im\phi'}\times \\ \times\,\tilde{\sigma}_\nu\left([k_z - K\alpha\sin\left(\phi_K-\omega_{\nu,j}\right)],m;j\right). \quad \text{(B.39)}$$

This expression may then be simplified with the aid of the Jacobi–Anger identity (see equation (b.2.12)), and by eliminating the sums by using equation (b.2.14). This leads to the following expression:

$$\tilde{\rho}_\nu(\mathbf{k}) = \frac{a}{(2\pi)^{1/2}}\sum_{n,n',n''=-\infty}^{\infty} J_n(Kb)J_{n'}(Ka)J_{n''}(k_z a\alpha)\times \\ \times\sum_j \tilde{\sigma}_\nu\left([k_z - K\alpha\sin\left(\phi_K-\omega_{\nu,j}\right)],n'+n'';j\right)e^{i(n+n'')(\omega_{\nu,j}-\phi_K+\pi/2)+ik_z s_j} \quad \text{(B.40)}$$

This expression is general, for any surface charge pattern and any precession angle $\omega_{\nu,j}$. We may now substitute the expressions for $\omega_{\nu,j}$ (equation (B.11)) and for the surface charge densities given in equations (B.19), (B.24) and (B.31):

$$\tilde{\rho}_\nu(\mathbf{k}) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}) = \frac{e_0}{(2\pi)^{3/2}}\sum_{n,n',n''=-\infty}^{\infty} J_n(Kb)J_{n'}(Ka)J_{n''}(k_z a\alpha)\times \\ \times\sum_j \left(1+t_{n'+n''}\left[k_z - K\alpha\sin\left(\phi_K - Qs_j-\psi_\nu\right)\right]\right) \\ \times q_j e^{i(n+n'')(Qs_j+\psi_\nu-\phi_K+\pi/2)+ik_z s_j} \quad \text{(B.41)}$$

This completes the calculation of the Fourier transform of the total charge density of an arbitrary pattern of surface charges.

B.1.7 Ideal double helices

We now want to specify the charge density to the case of an ideal DNA molecule (i. e. the charges are located at equally separated points in space, following the

line of an ideal helix). To do that, we start from the simpler case of a single helical line with equally spaced charges. We have for such system:

$$\phi'_{\nu,j} = gl_c j + \Phi_\nu - \omega_{\nu,j}. \tag{B.42}$$

We took into account that the charges are equally spaced (so that $s_j = jl_c$, where l_c is the axial rise between successive charges). The subtraction of $\omega_{\nu,j}$ ensures that the intrinsic twist of the chain is preserved. Substituting equations (B.42) and (B.19) into equation (B.40), we obtain:

$$\tilde{\rho}_\nu(\mathbf{k}) = \frac{N_\nu e_0}{(2\pi)^{3/2}} \sum_{n,n',n''} J_n(Kb)K_{n'}(Ka)J_{n''}(k_z a\alpha)e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\Phi_\nu} \times \\ \times \sum_j e^{[k_z+(n-n'')Q+(n'+n'')g]l_c j}. \tag{B.43}$$

The last sum can be easily calculated, if we take into account our assumption of a very long molecule. We can then approximate:

$$\sum_j e^{ixj} \approx N_\nu \delta_{x,2\pi J} \qquad J \in \mathbb{Z}. \tag{B.44}$$

Calculating the sum in equation (B.43) with the aid of formula (B.44) yields

$$\tilde{\rho}_\nu(\mathbf{k}) = \frac{N_\nu e_0}{(2\pi)^{3/2}} \sum_{n,n',n''} J_n(Kb)K_{n'}(Ka)J_{n''}(k_z a\alpha)e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\Phi_\nu} \times \\ \times \sum_{J=-\infty}^{\infty} \delta_{k_z,(n-n'')Q+(n'+n'')g-2\pi J/l_c}. \tag{B.45}$$

This calculation was originally performed by Crick (1953), and gave insight into the X-ray scattering patterns of α-helices.

For a double-stranded DNA molecule the calculation of the total structure factors is slightly more complicated, but the result shown in equation (B.45) retains its basic form. We need to specify the azimuthal orientations of the charges in a single base pair step. Following the assumptions of the KL theory (see section 1.2), we can write the following expressions for the phosphates and for the positive

counterions at the DNA surface:

$$\phi'_{\nu,4j} = gs_{4j} - \omega_{\nu,4j} + \Phi_\nu - \tilde{\phi}_s \tag{B.46}$$

$$\phi'_{\nu,4j-1} = gs_{4j} - \omega_{\nu,4j} + \Phi_\nu + \tilde{\phi}_s \tag{B.47}$$

$$\phi'_{\nu,4j-2} = gs_{4j} - \omega_{\nu,4j} + \Phi_\nu \tag{B.48}$$

$$\phi'_{\nu,4j-3} = gs_{4j} - \omega_{\nu,4j} + \Phi_\nu + \pi \tag{B.49}$$

$$s_{4j} = s_{4j-1} = s_{4j-2} = s_{4j-3} \tag{B.50}$$

These expressions correspond to the case of four helical lines of charge. Each of the four helices has the same precession frequency, which is given by $g = 2\pi/H$, where H is the DNA helical pitch. Here, Φ_ν is the azimuthal orientation of the centre of the minor groove of molecule ν, at $z = 0$. We then find that

$$\begin{aligned}\tilde{\rho}_\nu(\mathbf{k}) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}) = \frac{e_0}{(2\pi)^{3/2}} \sum_{n,n',n''=-\infty}^{\infty} J_n(Kb)J_{n'}(Ka)J_{n''}(k_z a\alpha)\zeta_{n+n''} \times \\ \times e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\psi_\nu+i(n'+n'')\Phi_\nu} \sum_j e^{i[k_z+(n-n'')Q+(n'+n'')g]s_j} \\ \times \left[1 + t_{n+n''}(k_z - K\alpha\sin(\phi_K - Qs_j - \psi_\nu))\right]\end{aligned} \tag{B.51}$$

$$\begin{aligned}\tilde{\rho}^{(1)}_{\nu,\mu}(\mathbf{k}) = \frac{e_0}{(2\pi)^{3/2}} \sum_{n,n',n''=-\infty}^{\infty} J_n(Kb)J_{n'}(Ka)J_{n''}(k_z a\alpha)\zeta_{n+n''} \times \\ \times e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\psi_\nu+i(n'+n'')\Phi_\nu} \sum_j e^{i[k_z+(n-n'')Q+(n'+n'')g]s_j} \\ \times \sum_{m=-\infty}^{\infty} \Big\{ w^{(0)}_{n'+n'',m}\left[k_z - K\alpha\sin(\phi_K - Qs_j - \psi_\nu)\right] + \\ + \alpha w^{(1)}_{n'+n'',m}\left[k_z - K\alpha\sin(\phi_K - Qs_j - \psi_\nu)\right] \Big\}.\end{aligned} \tag{B.52}$$

These expressions complete this step of the calculation. We now can substitute these structure factors into the expressions for the energy of the system (see equation (1.14)). However, before doing that, it is convenient to approximate the

expressions given in (B.51) and (B.52) to the first order in α. The result is:

$$\tilde{\rho}_\nu(\mathbf{k}) + \tilde{\rho}^{(0)}_{\nu,\nu}(\mathbf{k}) = \frac{e_0}{(2\pi)^{3/2}} \sum_{n,n',n''=-\infty}^{\infty} J_n(Kb) J_{n'}(Ka) J_{n''}(k_z a\alpha) \zeta_{n+n''} \times$$
$$\times e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\psi_\nu+i(n'+n'')\Phi_\nu} \delta_{-k_z,(n-n'')Q+(n'+n'')g}$$
$$\times \left[1 + t_{n+n''}(k_z) - K\alpha \sin(\phi_K - \psi_\nu) t'_{n+n''}(k_z)\right] \tag{B.53}$$

$$\tilde{\rho}^{(1)}_{\nu,\mu}(\mathbf{k}) = \frac{e_0}{(2\pi)^{3/2}} \sum_{n,n',n''=-\infty}^{\infty} J_n(Kb) J_{n'}(Ka) J_{n''}(k_z a\alpha) \zeta_{n+n''} \times$$
$$\times e^{i(n+n'')(\pi/2-\phi_K)+i(n-n'')\psi_\nu+i(n'+n'')\Phi_\nu} \delta_{-k_z,(n-n'')Q+(n'+n'')g}$$
$$\times \sum_{m=-\infty}^{\infty} \left\{ w^{(0)}_{n'+n'',m}(k_z) + \alpha \left[-K \sin(\phi_K - \psi_\nu) w^{(0)\prime}_{n'+n'',m}(k_z) + w^{(1)}_{n'+n'',m}(k_z) \right] \right\}. \tag{B.54}$$

In deriving equations (B.53) and (B.54), we neglected all terms in α that are higher order than first, and also performed the sum over all the charges in the same way as was shown before (see equation (B.44)).

B.1.8 Energy calculation: direct term

We are now ready to perform the calculation of the energy integrals given in equations (B.2) and (B.3). The $E_{1,2}$ term represents the *direct* electrostatic interaction term, whereas the $E_{\nu,\nu}$ terms are the *image* interaction terms. First, we describe the calculation of the direct term, then of the image term.

To calculate $E_{1,2}$, we substitute equation (B.53) into equation (B.2) and we obtain:

$$\frac{E_{1,2}(R)}{k_B T} = L\left(\Upsilon_1 + \Upsilon_2 + \Upsilon_3\right), \tag{B.55}$$

where, to the first order in α, we have

$$\Upsilon_1 = \frac{l_B}{\pi l_c^2}\int_{-\infty}^{\infty}\mathrm{d}k_z\int_0^{\infty}K\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n'\\m,m'}}^{\infty}(-1)^{m+m'}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times[1+t_{n'}(k_z)][1+t_{m'}(-k_z)]\zeta_{n'}\zeta_{m'}\delta_{(n'+m')g,-(n+m)Q}\delta(k_z+nQ+n'g)\times$$
$$\times e^{i(n+n'+m+m')(\pi/2-\phi_K)+in'\Phi_1+in\psi_1+im'\Phi_2+im\psi_2}, \tag{B.56}$$

$$\Upsilon_2 = \alpha\frac{l_B a}{2\pi l_c^2}\int_{-\infty}^{\infty}k_z\mathrm{d}k_z\int_0^{\infty}K\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n',n''\\m,m',m''}}^{\infty}(-1)^{m+m'}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times[1+t_{n'+n''}(k_z)][1+t_{m'+m''}(-k_z)]\zeta_{n'+n''}\zeta_{m'+m''}\delta(k_z+(n'+n'')g)\times$$
$$\times[\delta_{m'',0}(\delta_{n'',1}-\delta_{n'',-1})-\delta_{n'',0}(\delta_{m'',1}-\delta_{m'',-1})]\delta_{(n''-n+m''-m)Q,(n'+n''+m'+m'')g}\times$$
$$\times e^{i(n+n'+m+m')(\pi/2-\phi_K)+i(n'+n'')\Phi_1+i(n-n'')\psi_1+i(m'+m'')\Phi_2+i(m-m'')\psi_2}, \tag{B.57}$$

$$\Upsilon_3 = \alpha\frac{l_B}{\pi l_c^2}\int_{-\infty}^{\infty}\mathrm{d}k_z\int_0^{\infty}K^2\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n'\\m,m'}}^{\infty}(-1)^{m+m'}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times\left\{t'_{m'}(-k_z)[1+t_{n'}(k_z)]\sin(\phi_K-\psi_2)-t'_{n'}(k_z)[1+t_{m'}(-k_z)]\sin(\phi_K-\psi_1)\right\}\zeta_{n'}\zeta_{m'}\times$$
$$\times\delta_{(n'+m')g,-(n+m)Q}\delta(k_z+n'g)e^{i(n+n'+m+m')(\pi/2-\phi_K)+in'\Phi_1+in\psi_1+im'\Phi_2+im\psi_2}. \tag{B.58}$$

In deriving equations (B.56)-(B.58), we took into account that to the first order in α, we have

$$J_m(k_z a\alpha)\approx\delta_{m,0}+\alpha\frac{k_z a}{2}\left(\delta_{m,1}-\delta_{m,-1}\right), \tag{B.59}$$

which comes from differentiating the first-order Bessel functions, and setting α to zero (Abramowitz and Stegun, 1964). Also, we approximate

$$\frac{L}{2\pi}\delta_{-k_z,(n-n'')Q+(n'+n'')g}\approx\delta\left(k_z+[n-n'']\,Q+[n'+n'']\,g\right), \tag{B.60}$$

which is valid when $L/2\pi\gg Q^{-1},g^{-1}$. To calculate the integrals in equations (B.56)-(B.58), we use the formula already given in equation (b.2.14), and eliminate the sums with the help of the delta indices. The Kronecker delta

$$\delta_{ng,mQ} \tag{B.61}$$

has a non-zero value for $n = 0$, for $m = 0$. There is also another solution, in which $Q = g$. However, such case corresponds to a braid that has a very short pitch (equal to the DNA helical pitch), so it has a considerable tilt angle and it is not within our approximations. We discard such solution. We then need to use the following summation rule for the Bessel functions (Gradshteyn and Ryzhik, 1994):

$$\sum_{m=-\infty}^{\infty} J_m(Kb)J_{N-m}(Kb) = J_N(2Kb) \equiv J_N(KR). \tag{B.62}$$

The integrals over K are calculated using equation (b.2.18). Finally, using the standard recursion formulae for the Bessel functions, it is possible to show that

$$\left[1 - t_n\left(ng\right)\right]^2 \left[I_n\left(\kappa_n a\right)\right]^2 = \frac{1}{\left(\kappa_n a\right)^2 \left[K_n'\left(\kappa_n a\right)\right]^2}. \tag{B.63}$$

Using all the above prescriptions, after some algebra we obtain equations the direct electrostatic terms in equations (2.3) and (2.4).

B.1.9 Energy calculation: image term

The calculation of the image term follows essentially the same principles as the direct term. We substitute equations (B.53) and (B.54) into equation (B.3). This time we obtain

$$E_{\nu,\nu} \approx \frac{L}{k_B T}\left(\Lambda_1 + \Lambda_2 + \Lambda_3\right), \tag{B.64}$$

where the terms are given by

$$\Lambda_1 = \frac{l_B}{\pi l_c^2}\int_{-\infty}^{\infty}\mathrm{d}k_z\int_0^{\infty}K\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n'\\m,m',l}}^{\infty}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times w^{(0)}_{n',l}(k_z)[1+t_{m'}(-k_z)]\zeta_{n'}\zeta_{m'}\delta_{(n'+m')g,-(n+m)Q}\delta(k_z+nQ+n'g)\times$$
$$\times e^{i(n+m)\pi/2+(m-n)\phi_K)+i(n'+m')\Phi_\nu+i(n+m)\psi_\nu}, \tag{B.65}$$

$$\Lambda_2 = \alpha\frac{l_B a}{2\pi l_c^2}\int_{-\infty}^{\infty}k_z\mathrm{d}k_z\int_0^{\infty}K\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n',n''\\m,m',m'',l}}^{\infty}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times w^{(0)}_{n'+n'',l}(k_z)[1+t_{m'+m''}(-k_z)]\zeta_{n'+n''}\zeta_{m'+m''}\delta(k_z+(n'+n'')g)\times$$
$$\times[\delta_{m'',0}(\delta_{n'',1}-\delta_{n'',-1})-\delta_{n'',0}(\delta_{m'',1}-\delta_{m'',-1})]\delta_{(n''-n+m''-m)Q,(n'+n''+m'+m'')g}\times$$
$$\times e^{i(n+n''+m+m'')\pi/2+(m+m''-n-n'')\phi_K++i(n'+n''+m'+m'')\Phi_\nu+i(n-n''+m-m'')\psi_\nu}, \tag{B.66}$$

$$\Lambda_3 = \alpha\frac{l_B}{\pi l_c^2}\int_{-\infty}^{\infty}\mathrm{d}k_z\int_0^{\infty}K^2\mathrm{d}K\int_0^{2\pi}\mathrm{d}\phi_K\sum_{\substack{n,n'\\m,m',l}}^{\infty}\frac{J_n(Kb)J_{n'}(Ka)J_m(Kb)J_{m'}(Ka)}{K^2+k_z^2+\kappa^2}\times$$
$$\times\left\{w^{(0)}_{n',l}(k_z)\sin(\phi_K-\psi_\nu)t'_{m'}(k_z)+[1+t_{m'}(-k_z)][w^{(1)}_{n'}(k_z)-w^{(0)\prime}_{n'}\sin(\phi_K-\psi_\nu)\right\}\zeta_{n'}\zeta_{m'}\times$$
$$\times\delta_{(n'+m')g,-(n+m)Q}\delta(k_z+n'g)e^{i(n+n'+m+m')(\pi/2-\phi_K)+i(n'+m')\Phi_\nu+i(n+m)\psi_\nu}. \tag{B.67}$$

Then, to obtain the image component in equations (2.3) and (2.4), we use the same formulae as those used previously. This completes the calculation.

B.2 Non ideal DNA: torsional flexibility and adaptation

In the previous section we derived the formulas for the electrostatic interactions within a braid formed by a pair of ideal, rigid DNA molecules. As explained in the main text (see section 1.5), DNA is not ideal and it is torsionally flexible. It was shown in previous works that the effect of such non-ideality and torsional flexibility always needs to be taken into account (Cherstvy et al., 2004; Wynveen et al., 2008; Lee et al., 2010). We therefore here extend the theory of electrostatic interactions in DNA braids to include these effects.

There are several types of contributions to the free energy that one could take into account. These include undulations of the molecular axis, and fluctuations in twist and stretch of the base pair structure of DNA. In this work, we did not include the effect of undulations of the molecular axis. It was previously noticed (Lee et al., 2010) that the undulations do not affect the alignment of the opposedly charged groups, and their effect becomes significant only at larger interaxial separations. Here, at the equilibrium interaxial distances predicted by the theory (see section 2.2), we expect that the undulations will be suppressed and their contribution to the total free energy is small. In this work we neglected their contribution.

In this section we will first illustrate the description of non-ideal DNA helices (section B.2.1), then we will write an energy functional that describes the system (section B.2.2) and then solve for the free energy by applying a variational approximation in section B.2.3.

B.2.1 Non-ideal double helices

To describe non-ideal DNA, we utilize the same theoretical approach as was used in previous work (see in particular the Supplementary Material to (Lee et al., 2010)). The main difference between ideal and non-ideal helices is that in the latter case the azimuthal angle of the centre of the minor groove does not precess in a regular way. We write it as a function of the axial coordinate s:

$$\Phi_\nu(s) \approx \Phi_\nu(0) + \int_0^s \mathrm{d}\tau \frac{\Omega_\nu(\tau) - \bar{g}h_\nu(\tau)}{\bar{h}}, \tag{B.68}$$

where Ω_ν and h_ν are the twist and rise per base pair at the height s, respectively, $\bar{g}$ is the average DNA wavelength, and $\bar{h}$ is the average height per base pair. As before, the angle $\Delta\Phi$ is defined as the difference, at a given height, between the azimuthal angles of the two molecules. Equation (B.68) takes into account that both torsional and stretching deformations affect the value of the azimuthal orientation at a particular height.

Now, the approximation we make for the electrostatic energy is to consider

that it may be expressed in the following way:

$$E_{es} \approx \int_0^L \mathrm{d}s \sum_{n=-\infty}^{\infty} \left[(-1)^n \cos\left[n\Delta\Phi(s)\right] u_n + u_n^{im}\right]. \tag{B.69}$$

B.2.2 Energy functional

Using our ansatz for the helical phase (equation (B.68)) the energy of the elastic deformations in the braid becomes:

$$\frac{E_{ER}}{k_B T} = L l_p^b \frac{4\alpha^4}{R^2} + \frac{l_p^h}{2} \sum_{\nu=1}^{2} \int_0^L \mathrm{d}s \left[\Delta\varpi(s)\right]^2 + \\ + \frac{l_p^h}{2} \sum_{\nu=1}^{2} \int_0^L \mathrm{d}s \left[\frac{\mathrm{d}\Phi_\nu(s)}{\mathrm{d}s} - \frac{\Omega_\nu^0(s) - \bar{g} h_\nu^0(s)}{\bar{h}}\right]^2. \tag{B.70}$$

Here, we grouped the terms that represent an independent twist-stretch fluctuation into the term

$$\varpi(s) = \left(\frac{\bar{g}^2 C_t}{C_s}\right)^2 \frac{\Omega_\nu(s) - \Omega_\nu^0(s)}{\bar{h}} + \bar{g} \left(\frac{C_s}{\bar{g}^2 C_t}\right)^{1/2} \frac{h_\nu(s) - h_\nu^0(s)}{\bar{h}}. \tag{B.71}$$

Such term represents a constant, and may be discarded. In equation (B.70) the definition of the **helical persistence length** was given:

$$l_p^h = \frac{C_s C_t}{(C_s + \bar{g}^2 C_t) k_B T}. \tag{B.72}$$

Using the values reported in box 8, one obtains $l_p^h \approx 350$ Å.

Summing the electrostatic energy approximation as reported in equation (B.69), we arrive at this form for the energy functional of a non-ideal braid:

$$\frac{E_b}{k_B T} = L l_p^b \frac{4\alpha^4}{R^2} + \frac{l_p^h}{4} \int_0^L \mathrm{d}s \left(\frac{\mathrm{d}\Delta\Phi(s)}{\mathrm{d}s} - \frac{\Delta\Omega^0(s) - \bar{g}\Delta h^0(s)}{\bar{h}}\right)^2 + \\ + \int_0^L \mathrm{d}s \sum_{n=0}^{\infty} \left[(-1)^n \cos(n\Delta\Phi(s)) u_n + u_n^{im}\right] \tag{B.73}$$

Box 8: The elastic rod model

Consider a DNA molecule with a centreline described by a curve $\mathbf{r}(s)$, and with twist and rise per base pair given respectively by $\Omega(s)$ and $h(s)$. The energy function of the bending, torsional, and stretching deformations of the molecule, as described by the elastic rod model, is given by (Yakushevich, 2005):

$$E_{ER} = \frac{1}{2}\int \mathrm{d}s \left[B\left(\frac{\mathrm{d}^2\mathbf{r}(s)}{\mathrm{d}s^2}\right)^2 + C_s\left(\frac{h-h_0}{<h>}\right)^2 + C_t\left(\frac{\Omega-\Omega_0}{<h>}\right)^2\right]. \qquad \text{(b.2.19)}$$

Here, $<h>$ is the average rise per base pair ($\approx$ 3.4 Å), and $h_0 \equiv h_0(s)$ and $\Omega_0 \equiv \Omega_0(s)$ are the intrinsic axial rise and intrinsic of the DNA molecule at zero stress. The first term in equation (b.2.19) represents the bending energy cost, and it is assumed here that the relaxed form of DNA is straight (we assume that the sequences are completely random, so that there are no effects on the intrinsic curvature due to, e.g., phased A-tracts (Porschke et al., 1986)). The values of the elastic modulii are reported in the table below.

Name	Symbol	Value	Reference
Bending persistence length	$l_p^b = B/k_BT$	50 nm	(Hagerman, 1988)
Torsional persistence length	$l_p^t = C_t/k_BT$	85 nm	(Strick et al., 1999)
Stretching modulus	C_s	1200 pN	(Wenner et al., 2002)

Table 2: names and values of the DNA elastic constants.

B.2.3 Variational approximation

To calculate the free energy of a braid, first we separate the helical phase $\Delta\Phi(s)$ into two components: the intrinsic, sequence-dependent phase $\Delta\Phi^{(0)}(s)$, and the thermally fluctuating part, $\delta\Phi(s)$:

$$\Delta\Phi(s) = \Delta\Phi^{(0)}(s) + \delta\Phi(s) \qquad \text{(B.74)}$$

To calculate the contributions to the free energy of these two components, different strategies are used, as explained extensively in the Supplementary Material of (Lee et al., 2010). Here, we give a brief description of the procedure.

To account for thermal fluctuations, we write an approximate energy functional,

using a variational approximation:

$$\tilde{E}_{eff} = \frac{k_B T}{2} \int_0^L \mathrm{d}s \left[\frac{l_p^h}{2} \left(\frac{\mathrm{d}\delta\Phi(s)}{\mathrm{d}s} \right)^2 + \beta \left[\delta\Phi(s) \right]^2 \right]. \tag{B.75}$$

Here, β is a variational parameter that needs to be optimized. The free energy associated with this functional is given by the "standard" partition function, expressed as a path integral, plus the average of the difference between the real and effective energy functional:

$$F_b \approx -k_B T \ln \left[\int \exp\left(-\frac{\tilde{E}_{eff}}{k_B T} \right) \mathscr{D}\left(\delta\Phi\right) \right] + \left\langle\!\left\langle \tilde{E}_{eff} - E_b \right\rangle\!\right\rangle \tag{B.76}$$

where the path integration given by $\mathscr{D}\delta\Phi$ is performed over all possible "trajectories", and the double brackets $\ll\gg$ indicate averaging over both thermal and sequence-related fluctuations. The averaging over sequence-dependent structures should be intended as an average over all possible realizations of $\Omega_\nu^0(s) - \bar{g}h_\nu^0(s)$.

The optimization of the free energy of braid formation should be performed also over the realizations of the intrinsic phase angle $\Delta\Phi^{(0)}(s)$. To take into account the experimentally known fact that the correlation functions decay in a random walk-like way, we search for the optimum among a particular class of functions, given by

$$\Delta\Phi^{(0)} = \overline{\Delta\Phi} + \frac{1}{2\bar{h}} \int_0^L \mathrm{d}s' \left[\Delta\Omega^0(s') - \bar{g}\Delta h^0(s') \right] \frac{s - s'}{|s - s'|} e^{-|s-s'|/\lambda_h}, \tag{B.77}$$

where λ_h is now the second variational parameter to be optimized. The energy optimization is performed as prescribed in the Supplementary Material of (Lee et al., 2010). Once that is performed, the free energy function given in equation (2.13) is obtained.

Bibliography

Abramowitz M and Stegun I 1964 *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* Vol. 55 Courier Dover Publications.

Adrian M, ten Heggeler-Bordier B, Wahli W, Stasiak A, Stasiak A and Dubochet J 1990 *The EMBO journal* **9**(13), 4551.

Allemand J, Bensimon D, Lavery R and Croquette V 1998 *Proceedings of the National Academy of Sciences* **95**(24), 14152.

Allison S, Herr J and Schurr J 1981 *Biopolymers* **20**(3), 469–488.

Anderson P and Bauer W 1978 *Biochemistry* **17**(4), 594–601.

Arscott P, Li A and Bloomfield V 1990 *Biopolymers* **30**(5-6), 619–630.

Arscott P, Ma C, Wenner J and Bloomfield V 1995 *Biopolymers* **36**(3), 345–364.

Bai Y, Das R, Millett I, Herschlag D and Doniach S 2005 *Proceedings of the National Academy of Sciences of the United States of America* **102**(4), 1035.

Baldwin G, Brooks N, Robson R, Wynveen A, Goldar A, Leikin S, Seddon J and Kornyshev A 2008 *J. Phys. Chem. B* **112**(4), 1060–1064.

Barzel A and Kupiec M 2008 *Nature Reviews Genetics* **9**(1), 27–37.

Battle C, Van Den Broek B, Noom M, Van Mameren J, Wuite G and MacKintosh F 2009 *Physical Review E* **80**(3), 031917.

Bauer W and Benham C 1993 *Journal of molecular biology* **234**(4), 1184–1196.

Bauer W and Vinograd J 1970 *Journal of molecular biology* **47**(3), 419.

Baumann C, Bloomfield V, Smith S, Bustamante C, Wang M and Block S 2000 *Biophysical journal* **78**(4), 1965–1978.

Bednar J, Furrer P, Stasiak A, Dubochet J, Egelman E and Bates A 1994 *Journal of molecular biology* **235**(3), 825–847.

Benham C 1977 *Proceedings of the National Academy of Sciences of the United States of America* **74**(6), 2397.

Benham C and Mielke S 2005 *Annual Reviews of Biomedical Engineering* .

Besteman K, Hage S, Dekker N and Lemay S 2007 *Physical review letters* **98**(5), 58103.

Besteman K, Van Eijk K and Lemay S 2007 *Nature Physics* **3**(9), 641–644.

Bloomfield V 1996 *Current opinion in structural biology* **6**(3), 334–341.

Bloomfield V 1997 *Biopolymers* **44**(3), 269–282.

Boles T, White J and Cozzarelli N 1990 *Journal of molecular biology* **213**(4), 931–951.

Bopp P, Kornyshev A and Sutmann G 1996 *Physical review letters* **76**(8), 1280–1283.

Böttcher C, Endisch C, Fuhrhop J, Catterall C and Eaton M 1998 *Journal of the American Chemical Society* **120**(1), 12–17.

Bouchiat C and Mezard M 1998 *Physical review letters* **80**(7), 1556–1559.

Brady G and Fein D 1976 *Nature* **264**, 231–234.

Brady G, Satkowski M, Foos D and Benham C 1987 *Journal of molecular biology* **195**(1), 185–191.

Brent R 1971 *The Computer Journal* **14**(4), 422–425.

Bryant Z, Stone M, Gore J, Smith S, Cozzarelli N and Bustamante C 2003 *Nature* **424**(6946), 338–341.

Bustamante C, Marko J, Siggia E and Smith S 1994 *Science* **265**(5178), 1599.

Călugăreanu G 1959 *Rev. Math. pures appl* **4**(5), 5–20.

Camerini-Otero R and Felsenfeld G 1978 *Proceedings of the National Academy of Sciences of the United States of America* **75**(4), 1708.

Champoux J 2001 *Annual review of biochemistry* **70**, 369.

Charvin G, Bensimon D and Croquette V 2003 *Proceedings of the National Academy of Sciences of the United States of America* **100**(17), 9820.

Charvin G, Vologodskii A, Bensimon D and Croquette V 2005 *Biophysical journal* **88**(6), 4124–4136.

Chattoraj D, Gosule L and Schellman J 1978 *Journal of molecular biology* **121**(3), 327–337.

Cherny D and Jovin T 2001 *Journal of Molecular Biology* **313**(2), 295–307.

Cherstvy A, Kornyshev A and Leikin S 2004 *J. Phys. Chem. B* **108**(20), 6508–6518.

Clauvelin N, Audoly B and Neukirch S 2008 *Macromolecules* **41**(12), 4479–4483.

Clauvelin N, Audoly B and Neukirch S 2009 *Biophysical journal* **96**(9), 3716–3723.

Conwell C, Vilfan I and Hud N 2003 *Proceedings of the National Academy of Sciences of the United States of America* **100**(16), 9296.

Cormack A 1957 *Acta Crystallographica* **10**(5), 354–358.

Cortini R, Kornyshev A, Lee D and Leikin S 2011 *Biophysical Journal* **101**(4), 875–884.

Cortini R, Lee D and Kornyshev A 2012 *Journal of Physics: Condensed Matter* **24**, 162203.

Cozzarelli N and Wang J 1990 *DNA topology and its biological effects* Cold Spring Harbor Lab. Press Cold Spring Harbor (NY).

Craigie R and Mizuuchi K 1986 *Cell* **45**(6), 793–800.

Crick F 1953 *Acta Crystallographica* **6**(8-9), 685–689.

Danilowicz C, Lee C, Kim K, Hatch K, Coljee V, Kleckner N and Prentiss M 2009 *Proceedings of the National Academy of Sciences of the United States of America* **106**(47), 19824.

DeRouchey J, Parsegian V and Rau D 2010 *Biophysical Journal* **99**(8), 2608–2615.

Dickerson R and Drew H 1981 *Journal of Molecular Biology* **149**(4), 761–786.

Drlica K 1992 *Molecular microbiology* **6**(4), 425–433.

Duguet M 1993 *Nucleic acids research* **21**(3), 463.

Duguid J, Bloomfield V, Benevides J and Thomas Jr G 1993 *Biophysical journal* **65**(5), 1916–1928.

Dulbecco R and Vogt M 1963 *Proceedings of the National Academy of Sciences of the United States of America* **50**(2), 236.

Erdélyi A and Bateman H 1954 *Tables of integral transforms* New York.

Feuerstein B, Pattabiraman N and Marton L 1986 *Proceedings of the National Academy of Sciences* **83**(16), 5948.

Feuerstein B, Pattabiraman N and Marton L 1990 *Nucleic acids research* **18**(5), 1271–1282.

Fogg J, Catanese D, Randall G, Swick M and Zechiedrich L 2009 *Mathematics of DNA Structure, Function and Interactions* pp. 73–121.

Forterre P 2002 *Trends in Genetics* **18**(5), 236–237.

Forterre P, Mirambeau G, Jaxel C, Nadal M and Duguet M 1985 *The EMBO Journal* **4**(8), 2123.

Forth S, Deufel C, Sheinin M, Daniels B, Sethna J and Wang M 2008 *Physical review letters* **100**(14), 148301.

Franklin R and Gosling R 1953 *Nature* **171**, 740–741.

Fu W, Wang X, Zhang X, Ran S, Yan J and Li M 2006 *Journal of the American Chemical Society* **128**(47), 15040–15041.

Fuller F 1971 *Proceedings of the National Academy of Sciences of the United States of America* **68**(4), 815.

Galassi M, Gough B, Jungman G, Theiler J, Davies J, Booth M and Rossi F 2009 *GNU Scientific Library Reference Manual - Third Edition* Network Theory Ltd. **URL:** *http://www.gnu.org/software/gsl*

Gebe J, Delrow J, Heath P, Fujimoto B, Stewart D and Schurr J 1996 *Journal of molecular biology* **262**(2), 105–128.

Gore J, Bryant Z, Nöllmann M, Le M, Cozzarelli N and Bustamante C 2006 *Nature* **442**(7104), 836–839.

Gosule L and Schellman J 1976 *Nature* **259**, 333–335.

Gosule L and Schellman J 1978 *Journal of molecular biology* **121**(3), 311–326.

Gradshteyn I and Ryzhik I 1994 *New York: Academic Press,— c1994, 5th ed. completely reset, edited by Jeffrey, Alan* **1**.

Grosberg A, Nguyen T and Shklovskii B 2002 *Reviews of modern physics* **74**(2), 329.

Hagerman P 1988 *Annual review of biophysics and biophysical chemistry* **17**(1), 265–286.

Hammermann M, Brun N, Klenin K, May R, Toth K and Langowski J 1998 *Biophysical journal* **75**(6), 3057–3063.

Hearst J and Hunt N 1991 *The Journal of Chemical Physics* **95**(12), 9322.

Htun H and Dahlberg J 1988 *Science* **241**(4874), 1791–1796.

Hud N and Downing K 2001 *Proceedings of the National Academy of Sciences of the United States of America* **98**(26), 14925.

Hud N V 2009 *Nucleic acid-metal ion interactions* Royal Society of Chemistry.

Iizuka E 1977 *Polymer Journal* **9**(2), 173–190.

Kirkwood J 1934 *The Journal of Chemical Physics* **2**, 351.

Kłysik J, Stirdivant S M, Larson J E, Hart P A and Wells R D 1981 *Nature* **290**, 672–677.

Kornyshev A 2010 *Physical Chemistry Chemical Physics* **12**(39), 12352–12378.

Kornyshev A, Lee D, Leikin S and Wynveen A 2007 *Reviews of Modern Physics* **79**(3), 943–996.

Kornyshev A, Lee D, Wynveen A and Leikin S 2011 *Nucleic acids research* **39**(16), 7289–7299.

Kornyshev A and Leikin S 1997 *The Journal of Chemical Physics* **107**, 3656.

Kornyshev A and Leikin S 1999 *Physical Review Letters* **82**(20), 4138–4141.

Kornyshev A and Leikin S 2000 *Physical Review E* **62**(2), 2576–2596.

Kornyshev A and Leikin S 2001 *Physical Review Letters* **86**(16), 3666–3669.

Kornyshev A, Leikin S and Malinin S 2002 *The European Physical Journal E: Soft Matter and Biological Physics* **7**(1), 83–93.

Kornyshev A and Wynveen A 2009 *Proceedings of the National Academy of Sciences of the United States of America* **106**(12), 4683.

Krasnow M and Cozzarelli N 1982 *J. Biol. Chem* **257**, 2687–2693.

La Porta A and Wang M 2004 *Physical review letters* **92**(19), 190801.

Langowski J 1987 *Biophysical chemistry* **27**(3), 263–271.

Lee C, Mizusawa H and Kakefuda T 1981 *Proceedings of the National Academy of Sciences of the United States of America* **78**(5), 2838.

Lee D, Wynveen A, Kornyshev A and Leikin S 2010 *The Journal of Physical Chemistry B* **114**(35), 11668–11680.

Leforestier A and Livolant F 2009 *Proceedings of the National Academy of Sciences* **106**(23), 9157.

Lerman L 1971 *Proceedings of the National Academy of Sciences* **68**(8), 1886.

Leslie A, Arnott S, Chandrasekaran R and Ratliff R 1980 *Journal of molecular biology* **143**(1), 49.

Levin D 1973 *International Journal of Computer Mathematics* **3**(1-4), 371–388.

Liu L and Wang J 1987 *Proceedings of the National Academy of Sciences* **84**(20), 7024–7027.

Livolant F, Levelut A, Doucet J and Benoit J 1989 *Nature* **339**, 724–726.

Lyubchenko Y and Shlyakhtenko L 1997 *Proceedings of the National Academy of Sciences of the United States of America* **94**(2), 496.

Ma C and Bloomfield V 1994 *Biophysical journal* **67**(4), 1678–1681.

Maffeo C, Luan B and Aksimentiev A 2012 *Nucleic Acids Research* **40**(9), 3812–3821.

Maffeo C, Schöpflin R, Brutzer H, Stehr R, Aksimentiev A, Wedemann G and Seidel R 2010 *Physical Review Letters* **105**(15), 158101.

Manning G 1969 *The Journal of Chemical Physics* **51**, 924.

Marčelja S and Radić N 1976 *Chemical Physics Letters* **42**(1), 129–130.

Marenduzzo D, Orlandini E, Stasiak A, Tubiana L, Micheletti C et al. 2009 *Proceedings of the National Academy of Sciences* **106**(52), 22269–22274.

Marguet E and Forterre P 1994 *Nucleic acids research* **22**(9), 1681.

Marko J 1997 *Physical Review E* **55**(2), 1758–1772.

Marko J 2007 *Physical Review E* **76**(2), 021926.

Marko J and Neukirch S 2012 *Physical Review E* **85**(1), 011908.

Marko J and Siggia E 1994 *Science* **265**(5171), 506.

Marko J and Siggia E 1995 *Macromolecules* **28**(26), 8759–8770.

Martin A, Davies M, Rackstraw B, Roberts C, Stolnik S, Tendler S and Williams P 2000 *FEBS letters* **480**(2-3), 106–112.

Marx K and Reynolds T 1982 *Proceedings of the National Academy of Sciences* **79**(21), 6484.

Marx K and Ruben G 1983 *Nucleic acids research* **11**(6), 1839–1854.

Marx K and Ruben G 1984 *Journal of biomolecular structure & dynamics* **1**(5), 1109.

Mazur A 2012 *Physical Review E* **86**(1), 011914.

Mocci F and Laaksonen A 2012 *Soft Matter* **8**(36), 7289–7299.

Moroz J and Nelson P 1997 *Proceedings of the National Academy of Sciences* **94**(26), 14418–14422.

Mosconi F, Allemand J, Bensimon D and Croquette V 2009 *Physical review letters* **102**(7), 78301.

Murayama Y, Sakamaki Y and Sano M 2003 *Physical Review Letters* **90**(1), 018102.

Nakata M, Zanchetta G, Chapman B, Jones C, Cross J, Pindak R, Bellini T and Clark N 2007 *Science* **318**(5854), 1276–1279.

Nelder J and Mead R 1965 *The computer journal* **7**(4), 308–313.

Neukirch S and Marko J 2011 *Physical Review Letters* **106**(13), 138104.

Neuman K, Charvin G, Bensimon D and Croquette V 2009 *Proceedings of the National Academy of Sciences of the United States of America* **106**(17), 6986.

Nguyen T, Rouzina I and Shklovskii B 2000 *The Journal of Chemical Physics* **112**, 2562.

Noom M, van den Broek B, van Mameren J and Wuite G 2007 *Nature Methods* **4**(12), 1031–1036.

Ouameur A and Tajmir-Riahi H 2004 *Journal of Biological Chemistry* **279**(40), 42041–42054.

Parker C and Halford S 1991 *Cell* **66**(4), 781–791.

Pelta J, Durand D, Doucet J and Livolant F 1996 *Biophysical journal* **71**(1), 48–63.

Pelta J, Livolant F and Sikorav J 1996 *Journal of Biological Chemistry* **271**(10), 5656.

Peng H and Jackson V 2000 *Journal of Biological Chemistry* **275**(1), 657–668.

Plum G, Arscott P and Bloomfield V 1990 *Biopolymers* **30**(5-6), 631–643.

Podgornik R, Rau D and Parsegian V 1989 *Macromolecules* **22**(4), 1780–1786.

Podgornik R, Rau D and Parsegian V 1994 *Biophysical journal* **66**(4), 962–971.

Porschke D et al. 1986 *Journal of biomolecular structure & dynamics* **4**(3), 373.

Post C and Zimm B 1982 *Biopolymers* **21**(11), 2139–2160.

Pulleyblank D, Shure M, Tang D, Vinograd J and Vosberg H 1975 *Proceedings of the National Academy of Sciences of the United States of America* **72**(11), 4280.

Qiu X, Andresen K, Kwok L, Lamb J, Park H and Pollack L 2007 *Physical review letters* **99**(3), 38104.

Qiu X, Kwok L, Park H, Lamb J, Andresen K and Pollack L 2006 *Physical review letters* **96**(13), 138101.

Randall G, Pettitt B, Buck G and Zechiedrich E 2006 *Journal of Physics: Condensed Matter* **18**, S173.

Rau D, Lee B and Parsegian V 1984 *Proceedings of the National Academy of Sciences of the United States of America* **81**(9), 2621.

Rau D and Parsegian V 1992*a* *Biophysical journal* **61**(1), 260–271.

Rau D and Parsegian V 1992*b* *Biophysical journal* **61**(1), 246–259.

Ray J and Manning G 1994 *Langmuir* **10**(7), 2450–2461.

Rhoades M, Thomas C et al. 1968 *Journal of Molecular Biology* **37**(1), 41–56.

Ritort F, Mihardja S, Smith S and Bustamante C 2006 *Physical review letters* **96**(11), 118301.

Robinson C 1961 *Tetrahedron* **13**(1–3), 219–234.

Rouzina I and Bloomfield V 1996 *J. Phys. Chem* **100**(23), 9977–9989.

Ruiz-Chica J, Medina M, Sánchez-Jiménez F and Ramirez F 2001 *Biophysical journal* **80**(1), 443–454.

Rybenkov V, Cozzarelli N and Vologodskii A 1993 *Proceedings of the National Academy of Sciences of the United States of America* **90**(11), 5307.

Rybenkov V, Vologodskii A and Cozzarelli N 1997 *Nucleic Acids Research* **25**(7), 1412–1418.

Samori B, Siligardi G, Quagliariello C, Weisenhorn A, Vesenka J and Bustamante C 1993 *Proceedings of the National Academy of Sciences of the United States of America* **90**(8), 3598.

Schellman J and Parthasarathy N 1984 *Journal of Molecular Biology* **175**(3), 313–329.

Seidl A and Hinz H 1984 *Proceedings of the National Academy of Sciences of the United States of America* **81**(5), 1312.

Shao Q, Goyal S, Finzi L and Dunlap D 2012 *Macromolecules* **45**(7), 3188–3196.

Sharp K and Honig B 1990 *Annual review of biophysics and biophysical chemistry* **19**(1), 301–332.

Shaw S and Wang J 1993 *Science* **260**(5107), 533.

Sheinin M, Forth S, Marko J and Wang M 2011 *Physical Review Letters* **107**(10), 108102.

Shen M, Downing K, Balhorn R and Hud N 2000 *Journal of the Americal Chemical Society* **122**(19), 4833–4834.

Shklovskii B 1999*a* *Physical Review E* **60**(5), 5802–5811.

Shklovskii B 1999*b* *Physical Review Letters* **82**(16), 3268–3271.

Shlyakhtenko L, Miloseska L, Potaman V, Sinden R and Lyubchenko Y 2003 *Ultramicroscopy* **97**(1-4), 263–270.

Singleton C, Klysik J, Stirdivant S and Wells R 1982 *Nature* **299**, 312–316.

Smith S, Finzi L and Bustamante C 1992 *Science* **258**(5085), 1122–1126.

Spengler S, Stasiak A and Cozzarelli N 1985 *Cell* **42**(1), 325–334.

Sperrazza J, Register III J and Griffith J 1984 *Gene* **31**(1-3), 17–22.

Stanley C, Hong H and Strey H 2005 *Biophysical journal* **89**(4), 2552–2557.

Strick T, Allemand J, Bensimon D, Bensimon A and Croquette V 1996 *Science* **271**(5257), 1835.

Strick T, Bensimon D and Croquette V 1999 *Genetica* **106**(1), 57–62.

Su T, Theofanidou E, Arlt J, Dryden D and Crain J 2004 *Journal of Fluorescence* **14**(1), 65–69.

Tanaka F and Takahashi H 1985 *The Journal of chemical physics* **83**, 6017.

Tanigawa M and Okada T 1998 *Analytica Chimica Acta* **365**(1-3), 19–25.

Teif V and Bohinc K 2011 *Progress in biophysics and molecular biology* **105**(3), 208–222.

Timsit Y and Moras D 1991 *Journal of molecular biology* **221**(3), 919–940.

Timsit Y and Moras D 1994 *The EMBO Journal* **13**(12), 2737.

Timsit Y, Várnai P and Wölfl S 2010 *PloS one* **5**(2), 430–440.

Timsit Y, Westhof E, Fuchs R and Moras D 1989 *Nature* **341**(6241), 459–462.

Todd B, Adrian Parsegian V, Shirahata A, Thomas T and Rau D 2008 *Biophysical journal* **94**(12), 4775–4782.

Todd B and Rau D 2008 *Nucleic acids research* **36**(2), 501–510.

Tombolato F and Ferrarini A 2005 *The Journal of chemical physics* **122**(5), 054908–1.

Torbet J and DiCapua E 1989 *The EMBO Journal* **8**(13), 4351.

Travers A and Muskhelishvili G 2007 *EMBO reports* **8**(2), 147–151.

Tsuru H and Wadati M 1986 *Biopolymers* **25**(11), 2083–2096.

Ubbink J and Odijk T 1999 *Biophysical Journal* **76**(5), 2502–2519.

Valenti A, Perugino G, Rossi M and Ciaramella M 2011 *Biochemical Society transactions* **39**(1), 58.

Várnai P and Timsit Y 2010 *Nucleic Acids Research* **38**(12), 4163.

Vetcher A, McEwen A, Abujarour R, Hanke A and Levene S 2010 *Biophysical chemistry* **148**(1-3), 104–111.

Vinograd J, Lebowitz J, Radloff R, Watson R and Laipis P 1965 *Proceedings of the National Academy of Sciences of the United States of America* **53**(5), 1104.

Vologodskii A 1994 *Macromolecules* **27**(20), 5623–5625.

Vologodskii A, Levene S, Klenin K, Frank-Kamenetskii M and Cozzarelli N 1992 *Journal of molecular biology* **227**(4), 1224–1243.

Vologodskii A V and Cozzarelli N R 1994 *Annual Review of Biophysics and Biomolecular Structure* **23**(1), 609–643.

Wang J 2002 *Nature Reviews Molecular Cell Biology* **3**(6), 430–440.

Wang X, Zhang X, Mao C and Seeman N 2010 *Proceedings of the National Academy of Sciences* **107**(28), 12547.

Watson J and Crick F 1953 *Nature* **171**, 737–738.

Wemmer D, Srivenugopal K, Reid B and Morris D 1985 *Journal of molecular biology* **185**(2), 457–459.

Wenner J, Williams M, Rouzina I and Bloomfield V 2002 *Biophysical Journal* **82**(6), 3160.

White J 1969 *American Journal of Mathematics* **91**, 693–728.

Widom J and Baldwin R 1983 *Biopolymers* **22**(6), 1595–1620.

Wilson R and Bloomfield V 1979 *Biochemistry* **18**(11), 2192–2196.

Witz G and Stasiak A 2010 *Nucleic Acids Research* **38**(7), 2119.

Wynveen A, Lee D, Kornyshev A and Leikin S 2008 *Nucleic acids research* **36**(17), 5540.

Xu X, Zhi X and Leng F 2012 *Biochimie* **94**(12), 2665–2672.

Xu Y and Bremer H 1997 *Nucleic acids research* **25**(20), 4067.

Yakushevich L V 2005 *Nonlinear Physics of DNA* Wiley-VCH.

Yang J and Rau D 2005 *Biophysical journal* **89**(3), 1932–1940.

Yevdokimov Y, Skuridin S and Salyanov V 1988 *Liquid Crystals* **3**(11), 1443–1459.

Yoshikawa K, Takahashi M, Vasilevskaya V and Khokhlov A 1996 *Physical review letters* **76**(16), 3029–3031.

Zanchetta G, Giavazzi F, Nakata M, Buscaglia M, Cerbino R, Clark N and Bellini T 2010 *Proceedings of the National Academy of Sciences of the United States of America* **7**(1), 83–93.

Zhang R and Shklovskii B 2005 *Physica A: Statistical Mechanics and its Applications* **349**(3), 563–570.

Zickler D and Kleckner N 1999 *Annual Review of Genetics* **33**(1), 603–754.

Acknowledgements

First and foremost, I wish to thank my supervisor, Professor Alexei A. Kornsyhev, for his support, his extraordinary enthusiasm and curiosity, and for giving me the unique opportunity to work in his group, and to live in London for three years.

The work presented in this thesis would not have been possible, however, without the support and the technical abilities of Dr. Dominic J. Lee. His incredible skills and his passion for science were a true driving force of this project.

The third, but not less important, person I want to thank is Dr. Sergey Leikin, at the National Institutes of Health in Bethesda. In the first year of my PhD, and in the preparation of our first research article, I learned from him invaluable lessons of scientific clarity, rigour, method, and integrity. The "nested" style of presentation of this thesis is due to him.

I am in great debt to Dr. Svyatoslav Kondrat for his immense patience with my difficulties in writing the code that was used in this work. Without his advice, I would have never learned to use the software which today I cannot live without. I wish to thank him also for his incredible work in building and maintaining the group cluster.

Thanks to Jack Paget for his suggestions on the manuscript, and for dedicating so much time and energy to administrative tasks in the research group.

I wish to thank our project partners at the Mathematics Department of University College London, Professor Gert van der Heijden, Dr. Eugene Starostin, and Dr. Anthony Korte, for their patient proofreading of our technical manuscripts, and their precious contribution to the project.

I am very grateful to Professor Gijs Wuite and Dr. Graeme King at Vrije Universiteit Amsterdam, for their effort in designing and performing the challeging four-bead experiments, and for giving me the opportunity to visit their group in Amsterdam.

My sincere gratitude goes to all our experimental collaborators. It was a pleasure and an honour to collaborate with Dr. Tim Albrecht, at Imperial College London, and all the members of his group: Azadeh Bahrami, Deanpeng Japrung, Agnieska Rutkowska, William Pitchford, Philippa Nuttall. I wish to thank also Dr. Nick Brooks for his experiments on small-angle X-ray scattering. At Imperial,

I also wish to thank Dr. Geoff Baldwin and Timothy Wilson for their experimental efforts and for stimulating discussions. Thanks also to Professor Lynn Zechiedrich, at Baylor College of Medicine, for her magnificent experimental work.

I am immensely grateful to Professor Jeremy Smith, at the Oak Ridge National Laboratories, for giving me the possibility to work within his group for two months, and for his kind support and motivation. Within his group, thanks to Dr. Xiaolin Cheng, I had the opportunity to learn and use the most advanced methods for simulating DNA molecules.

My debt of gratitude goes also to my PhD mentor at Imperial College London, Professor John Seddon, for his kind support, and his unique ability to listen and to advise.

Many thanks go to Professor Sarah Harris, at Leeds University, for inviting me to the Newton Institute workshop on DNA topology, and for her great enthusiasm for our theory, and willingness to test it in computer simulations in her group.

I also wish to thank many other scientists that in these three years gave support to our work with invaluable discussions: Professor Wilma Olson, Dr. Tony Maxwell, Dr. Andy Bates, Professor Giovanni Dietler, Professor Andrzej Stasiak, Dr. Christian Micheletti.

Special thanks go to Professor Serge Lemay, at Twente University, for the discussions his experiments on single molecule supercoiling, which stimulated our interest in formulating the theory as it was.

This experience at Imperial College London would not have been possible without the support of my Master mentor, Professor Antonio Coniglio, whom I wish to thank, once again, for his never-ending support.

Thanks also to the organizers of the "Frontiers in Statistical Physics" conference in Catania, Professor Lucilla de Arcangelis, Dr. Annalisa Fierro, Dr. Hans Herrmann, Professor Francesco Mallamace, and Dr. Andrea Rapisarda for inviting me to give a presentation there.

Thanks to the Engineering and Physical Sciences Resarch Council (EPSRC), that sponsored this work with their grant number EP/H004319/1.