POLITECNICO DI TORINO

International Master course in Physics of Complex Systems

Supercoiled knotted DNA: theory and simulations

Supervisor:
Cristian Micheletti
Scuola Internazionale Superiore
di Studi Avanzati SISSA

Internal Supervisor:
Andrea Pagnani

Candidate:
Pietro Luigi Muzzeddu

Academic year 2019 – 2020

Abstract

Only recently have experimentalists been able to detect and measure the incidence of various types of knots in eukaryotic DNA, and specifically in yeast minichromosomes. Both the abundance and the complexity of these knots have a precise dependence on intrinsic properties of the DNA (such as its length) as well as on extrinsic ones, such as the degree of supercoiling that can be introduced by tighty-regulated molecular machines, or enzymes. Interestingly, it has also been shown that the same supercoiling level affects to varying degrees the electrophoretic migration velocity of two chiral enantiomers of the same knot species. The aim of this thesis is to use coarse-grained structural models and simulations to advance our understanding of how knotting and supercoiling affect the properties of yeast minichromosomes. In particular, we developed a minimalistic coarse-grained model of circular DNA filament as an elastic rod, and set up Monte Carlo simulations to sample the configurational space populated at canonical equilibrium. We considered both knotted and unknotted filaments, with and without supercoiling, or nicked. For each combination of topological (knotting) or geometrical (supercoiling, nicking) entanglement, we systematically profiled various observables, such as the gyration radius, asphericity, anisotropy, twist and writhe and used them to clarify the extent to which minimalistic physical models can account for the recent experimental measurements, and illuminate their interpretation.

Contents

Abstract **i**

Introduction **1**

1 A brief introduction to Knot Theory **3**
1.1 Definition of Knot . . . 3
1.2 Knot diagram and Reidemeister moves . . . 4
1.3 Crossing number and unknotting number . . . 6
1.4 Average crossing number . . . 7
1.5 Chirality . . . 8
1.6 Knot composition . . . 9
1.7 Main knot families . . . 10
1.8 Dowker notation . . . 11
1.9 Polynomials . . . 12
1.9.1 Alexander polynomial . . . 13

2 An overview of DNA biology **15**
2.1 DNA structure . . . 15
2.2 Supercoiling . . . 19
2.3 Topoisomerases . . . 21
2.3.1 Type I topoisomerases . . . 21
2.3.2 Type II topoisomerases . . . 22
2.4 Biological role . . . 22
2.4.1 Replication . . . 23
2.4.2 Transcription . . . 24
2.4.3 Translation . . . 25

3 DNA geometry and topology **28**
3.1 Geometrical descriptors . . . 28
3.1.1 Twist . . . 28
3.1.2 Writhe . . . 30
3.2 Topological descriptor: Linking number . . . 31
3.2.1 White-Călugăreanu-Fuller relation . . . 31
3.2.2 Superhelical density . . . 33
3.3 Supercoiled knotted configurations . . . 34

3.3.1 Ideal knots . . . 34
3.3.2 Supercoiling evaluation . . . 35
3.4 Topoisomerases action . . . 36

4 Supercoiled knotted DNA: recent experimental findings 37
4.1 Experimental assessment of DNA knots presence . . . 37
4.2 Knots complexity and migration velocity . . . 37
4.3 Knots occurence in eukaryotic chromatin . . . 39
4.4 DNA knots chirality . . . 40

5 Methodology and Statistical Physics model 42
5.1 DNA Coarse Grained Model . . . 42
5.1.1 Stretching . . . 43
5.1.2 Bending . . . 43
5.1.3 Excluded volume . . . 44
5.1.4 Torsion . . . 44
5.2 Geometric observables . . . 45
5.2.1 Gyration radius and tensor . . . 46
5.2.2 Anisotropy and asphericity . . . 46
5.3 Monte Carlo Sampling . . . 47
5.3.1 Type of moves . . . 48
5.3.2 Autocorrelation time . . . 50
5.3.3 Tuning the moves amplitude . . . 52
5.4 Validation of the model . . . 52

6 Results and open challenges 56
6.1 Purely topological effects . . . 60
6.2 Supercoiling effects . . . 61
6.2.1 Unknot topology . . . 61
6.2.2 Trefoil topology . . . 62
6.2.3 Comparing supercoiled 0_1 and 3_1 . . . 64
6.3 Conclusions . . . 65

Appendix A 66

Appendix B 68

List of figures 72

Bibliography 77

Other sources 80

Acknowledgements 81

Introduction

DNA is an extremely important biological macromolecule (bio-polymer) that represents the hereditary material and codes the genetic information of every living being. In eukaryotes, DNA is contained inside the nucleus, where it reaches a high packing ratio thanks to an articulate hierarchical spatial arrangement, referred to as eukaryotic chromatin. In particular, the DNA double helix is wrapped around histones to form DNA-protein complexes called nucleosomes. Such nucleosomal fibers further condense in more packed structures, and the architectural complexity increases all the way up to chromosome scale. Even though the genomic DNA is linear, some extrachromosomal circular DNA (eccDNA) are found in all eukaryotes (e.g. in yeast cells).
Differently, the prokaryotic genetic material freely moves in the so-called nucleoid region of the cytoplasm, its architecture is much less complex and it is generally naked (not associated with proteins). Prokaryotic organisms, e.g. bacteria, contain also small circular extrachromosomal DNA called plasmids.

The spatial arrangement and conformational complexity of closed DNA molecules can be better understood by also paying attention to their topological and geometrical constraints. In particular, closed DNA duplexes can assume knotted conformations of various topological complexity, from the trivial knot (unknot) or the simple trefoil knot, up to more entangled configurations. Strand passages are the only way to vary the topology of a closed DNA molecule. At the same time, torsional strains due to either over-twisting or under-twisting of the double helix can affect the molecular compaction. DNA molecules which are not torsionally relaxed are said to be supercoiled. Both knotting/unknotting activity and supercoiling level regulation *in vivo* are performed by specialized enzymes called topoisomerases (type I and type II). If knots occurrence in bacterial plasmids has been documented since few decades [23, 12], only recently experimentalists have been able to assess the incidence of various types of knots in eukaryotic DNA (e.g. yeast minichromosomes) [28, 27].

In this work, we implemented a coarse-grained model of a closed DNA filament that takes into account both bending and torsional rigidity, and we set up Monte Carlo simulations to sample its configurational space at canonical equilibrium. We characterized both knotted and unknotted DNA closed chains, with and without supercoiling, and we estimated the molecular compaction and shape by analyzing some relevant geometrical observables (as gyration radius, asphericity, etc). The aim of this project is to give a direct insight into how knotting and supercoiling affects the geometrical conformation and spatial arrangement of closed DNA molecules, in the attempt to advance our understanding of the recent experimental evidences resulting from gel electrophoresis measurements, and illuminate their interpretation.

In particular, the work is organized in the following chapters:

Chapter 1: This chapter is intended to briefly introduce the main concepts of Knot Theory. Starting from the mathematical definition of knot and how to diagrammatically represent it, we then provided a description of some of the most common quantities used to characterize knot species. Then, we explained what it means to compose two knots and we listed the main knot families. To conclude, we mentioned an important class of knot invariants known as knot polynomials.

Chapter 2: This chapter is devoted to the DNA description from a biological point of view. The DNA chemical composition, structural features and elastic properties are presented, together with a rather simplified characterization of the eukaryotic chromatin architecture. Then, we focused on supercoiling and its effects on molecules' spatial arrangement. We proceeded by describing some enzymes called topoisomerases, whose activity can alter the supercoiling level and/or the molecules topology. Finally, we summarized some of the biological roles of the DNA.

Chapter 3: Some of the concepts already introduced in the previous chapters are mathematically formalized. We defined twist Tw and writhe Wr, and we described the White-Călugăreanu-Fuller relation that relates them to the conserved topological descriptor called linking number Lk. Then we explained how to take advantage of the aforementioned quantities to quantify the supercoiling level and the topoisomerase action.

Chapter 4: Here, we reported some of the more recent experimental evidences regarding the migration velocities of molecules with various supercoiling levels and topologies during gel electrophoresis. In particular, we recalled how the knot complexity and the interplay between chirality and supercoiling have been shown to affect the gel migration velocity.

Chapter 5: This chapter contains the information about the model and the methods that we used to carried out our simulations and try to reproduce what has been obtained experimentally. We also described the geometrical observables that we profiled to characterize compaction and shape of both knotted and unknotted molecules, with and without supercoiling. We finally explained how we tuned the parameters and how we validated our model.

Chapter 6: Here, we presented our results and we made a comparison with the experimental findings described in chapter 4.

Chapter 1

A brief introduction to Knot Theory

1.1 Definition of Knot

Everyday life leads us to consider as knots open strings that have been previously tied in a way similar to shoelaces. However, this intuition of the knot concept is slightly different to the one suggested by Knot Theory, which expects the two ends of the string to be glued together. So, as a first elementary attempt to define a knot, one may resort to an operational approach: imagine to take a piece of rope, tying a knot in it and finally joining the two free extremities (see Figure 1.1). This requisite of dealing with a closed string paves the way for a more rigorous description and classification based on a topological approach. Already at this point it is immediate to see that the simplest knot (also called trivial knot or unknot) is nothing else than a ring.

Figure 1.1: Illustration of the knot operational definition. Image from [1].

By reproducing the series of steps described in the Figure 1.1 one ends up with a knotted configuration which exhibits a well defined topology. Such a configuration cannot be disentangled (recast into the unknot) unless one cuts the rope, allows for a passage of a rope portion through the broken part, and finally reseals it. Any other continuous deformation (which does not involve any breaking) of the knotted rope will leave its topology unaltered. This statement can be made mathematically more precise by introducing the concept of *homeomorphism*.
An homeomorphism is a function $h : S_1 \to S_2$ between two topological spaces S_1, S_2 which satisfies the following properties:

1. h is bijective
2. h is continuous
3. the inverse h^{-1} is continuous

If such a transformation exists, S_1 and S_2 are said to be homeomorphic.
Differently from the knotted rope used in the operational definition, which has a finite thickness, the mathematical knot is defined as a simple closed curve $K : [0, 1] \to \mathbb{R}^3$. The function K is continuous and its only "non-injectivity" is $K(0) = K(1)$ (closure condition).
Two knots K_1, K_2 are said to be equivalent if there exists an orientation-preserving homeomorphism $h : \mathbb{R}^3 \to \mathbb{R}^3$ such that $h(K_1) = K_2$. This equivalence relation based on homeomorphic transformations suggests to extend the knot definition to the *equivalence class* containing all the homemorphic realizations with a certain topology. Note that the homeomorphism is defined on the whole space $\mathbb{R}^3$ in which the curve is embedded, and any attempt to define the equivalence relation by deforming only a subset of the space may lead to erroneous equivalences (e.g. a knotted configuration could be transformed into the unknot by pulling the curve tighter and tighter until the knot shrinks to a point and disappears, as represented in the Figure 1.2).

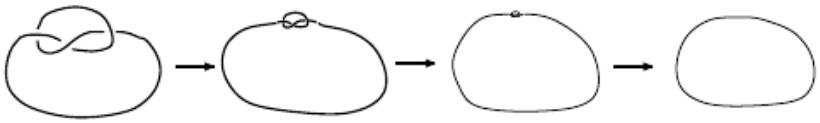

Figure 1.2: Knot disappearance with a continuous transformation which does not involve the ambient space. Image from [16].

In particular if two knots K_1 and K_2 are equivalent, it is natural to require also the 3-dimensional manifolds defined by their complements in $\mathbb{R}^3$ to be homeomorphic.
Another transformation suitable for defining knots while preventing the occurrence of such singularities is the *ambient isotopy*, which is a set of homeomorphisms

$$\{h_t : \mathbb{R}^3 \to \mathbb{R}^3,\, t \in [0, 1]\}$$

parametrized by t, such that at the end of the transformation K_1 is mapped into K_2. According to the equivalence relation based on the ambient isotopy one can say that K_1 is equivalent to K_2 if there exists a continuous $H : \mathbb{R}^3 \times [0, 1] \to \mathbb{R}^3$ such that:

1. $\forall t \in [0, 1]$, H_t is an homeomorphism of $\mathbb{R}^3$ onto itself.
2. $\forall \boldsymbol{x} \in \mathbb{R}^3,\ H(\boldsymbol{x}, 0) = \boldsymbol{x}$
3. $H(K_1, 1) = K_2$

1.2 Knot diagram and Reidemeister moves

Thus, different knot types belong to different equivalence classes, built on the basis of the ambient isotopy. But how can one practically distinguish them? A useful representation is obtained by looking at the knot from a certain point of view and depicting the closed curve in two dimensions specifying under- and over-passages whenever a crossing is encountered. In particular, discontinuous segments represent under-passing strands in a crossing, whereas continuous segments are used for

the over-passing ones. The representation obtained by following this procedure is the so called *knot diagram*. Clearly, very different knot diagrams may result from different viewing perspectives and the $2D$ drawings can be rather complicated. Nonetheless, since both diagrams are obtained from the same knot, there should exists a series of continuous manipulations that leads from a diagram to the other. Analogously one could think to perform an arbitrary ambient isotopy while keeping fixed the point of view, and look how the knot diagram varies during the transformation. It has been shown by the mathematician Reidemeister that all the possible diagram changes resulting from the application of an ambient isotopy can be encoded and reproduced by means of a series of planar isotopies and a set of three basic moves (reported in Figures 1.3, 1.4, 1.5), called Reidemeister moves.

Figure 1.3: Type I Reidemeister move

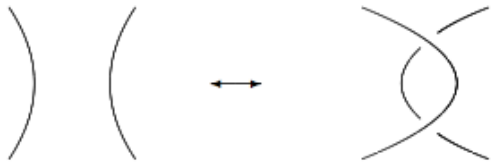

Figure 1.4: Type II Reidemeister move

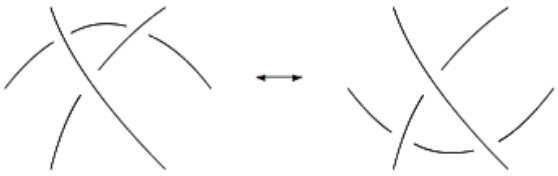

Figure 1.5: Type III Reidemeister move

The type I Reidemeister move either adds or removes a crossing from a region of the diagram, while keeping unchanged all the rest. The type II Reidemeister move can increase/decrease the number of crossings by 2, and the type III Reidemeister move let one strand slide from one side of a crossing to the other, without changing the overall number of crossings. Since every continuous deformation can be decomposed in a sequence of planar isotopies and Reidemeister moves, these elementary transformations can in principle be used to simplify the $2D$ projection by getting rid of some redundant crossings.

1.3 Crossing number and unknotting number

As highlighted in the previous section, the number of projected crossings of a knotted curve depends in general on the particular point of view chosen to perform the projection. Among all possible diagrams, there is a subset containing those which exhibit the minimum number of crossings compatible with the knot topology. This quantity is called *crossing number* and for a knot K it will be denoted with $c(K)$. All the diagrams presenting $c(K)$ crossings are called minimal knot diagrams.

A crucial aspect in Knot Theory is to find some quantities or mathematical descriptors through which knots can be properly classified. Such descriptors are called *knot invariants* because their value or expression should not depend on the particular knot realization and diagram. Some of them are obtained by simply taking the minimum value of an integer function defined on the space of knot diagrams: the crossing number is an instance of this kind of quantities, and is the first knot invariant encountered so far. The importance of the crossing number $c(K)$ lies also in the

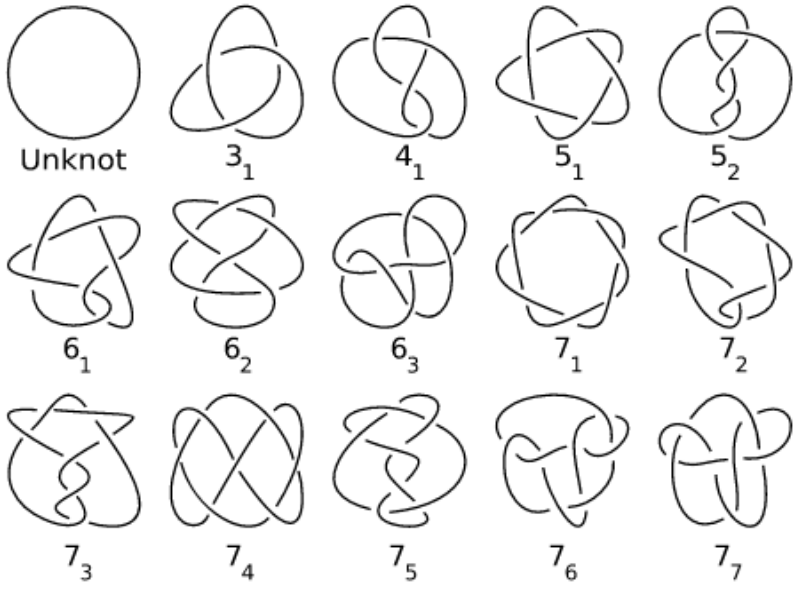

Figure 1.6: Table of knots up to crossing number $c(K) = 7$.

fact that it appears in the knot nomenclature, as represented in Figure 1.6. The label c_i is assigned to the i^{th} knot with crossing number $c(K) = c$. As the crossing number and the topological complexity increase, there will be more and more knot species for a fixed $c(K)$. The unknot (also indicated as 0_1) has the lowest possible crossing number $c(0_1) = 0$, followed by the so-called trefoil knot $c(3_1) = 3$, and so on. There are 21 knot species with crossing number $c(K) = 8$, 165 with $c(K) = 10$ and 552 with $c(K) = 11$, suggesting a rather steep increasing trend. The difficulty to recognize a knot species from a complex non-minimal diagram encouraged the research of more complete and efficient invariants, like the knot polynomials which will be briefly treated in the Section 1.9.

A similar descriptor (invariant) is given by the *unknotting number* , i.e. the minimum number of crossing changes required to turn K into the unknot, and it will be denoted with $u(K)$. More explicitly, having unknotting number $u(K)$ means that there exists at least a knot diagram which can be unknotted with $u(K)$ cross inversions, and there are no other diagrams requiring less crossing changes to become unknotted. Despite their similar definition, the unknotting number $u(K)$ and the crossing number $c(K)$ are not identical, and some knot species with a fair amount of crossings are characterized by a low value of $u(K)$ (see Figure 1.7).

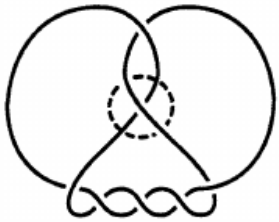

Figure 1.7: The knot 7_2 can be unknotted performing a single strand passage where denoted by the dashed circle. Image from [1].

To determine the unknotting number of an arbitrary knot species is generally not trivial: it has been found for all the knots with $c(K) \leq 9$, but it is still unknown for more complex cases like the 10_{11}. For some knot families (as twist knots and torus knots) which will be presented in the Section 1.7, the computation of the unknotting number $u(k)$ is extremely facilitated.

1.4 Average crossing number

A physically relevant quantity related to the crossing number $c(K)$ is the so called *average crossing number* (ACN). The way to compute it is to collect the number of crossings emerging from the knot diagrams obtained by looking at the knot from all the possible viewing perspectives, and finally averaging over them. A bit more formally, imagine to call $\boldsymbol{v}$ the unit vector defining the direction along which the knot is looked to produce the knot diagram. Such a diagram has a number of crossings equal to $n(\boldsymbol{v})$. The average crossing number ACN can be found by solving the following surface integral

$$ACN(K) = \frac{1}{4\pi} \int_{\Omega} n(\boldsymbol{v})\, dS$$

where the domain of integration is the unit sphere

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^3 \,|\, x_1^2 + x_2^2 + x_3^2 = 1\}$$

Note that not all the directions produce a meaningful knot diagram, but the pathological ones are a set of zero measure with respect to the integration domain, hence they don't affect the integral. Moreover, the average crossing number value is strictly correlated to the knot geometrical realization, and for that reason it cannot be considered a knot invariant.

1.5 Chirality

The crossing number alone is not sufficient to completely characterize knots, and often their nomenclature requires an information regarding their handedness. Indeed, the minimal diagram of the vast majority of knots is not equivalent to its mirror image, in the sense that there does not exist an ambient isotopy which brings one into the other. All the knots with this property are called *chiral*, and their left- and right-handed realizations *enantiomers*. Conversely, whenever a knot can be superimposed to its mirror image, it is said to be *achiral* (or *amphichiral*). The first and simplest chiral knot is the trefoil knot 3_1, whose enantiomers are reported in Figure 1.8: the simplicity of this case make intuitively clear the lack of an ambient isotopy mapping one enantiomer into the other. For the sake of clarity, the notation of the two enantiomers will be equipped with a superscript describing the knot handedness (e.g. 3_1^L and 3_1^R).

Figure 1.8: Left- and right-handed version of the trefoil knot

For relatively simple knots ($c(K) \leq 8$) the handedness can be computed in a rather direct way: the minimal diagrammatic representation of the knot needs to be endowed with an orientation, and each crossing labeled with $+1$ or -1 according to the right-hand rule (as sketched in the Figure 1.9). Once that the labels have been assigned to the crossings, one performs their algebraic sum, whose sign will define the knot handedness (e.g. for the two enantiomers of the trefoil knot the sums will be -3 for the left-handed and $+3$ for the right-handed one).

Figure 1.9: Right-hand rule to establish the sign of a crossing

Note that the amphichirality concept described so far completely disregards whether the knot orientation is preserved or not at the end of the ambient isotopy.

The knot chirality classification can be improved by introducing the concept of invertibility. In particular, an oriented-knot is said to be invertible if there exists an ambient isotopy which maps the curve image to itself, but reverses its orientation. In other words an oriented knot is invertible if it is equivalent to its inverse. If a knot is both chiral and invertible, it is called *reversible*. A chiral

knot which is not invertible is called *fully chiral*. Note that the fully chiral knots are quite rare for species with low crossing number, the first instance being the 9_{32}. What about the amphicaral knots? If a knot is equivalent to both its mirror image and its inverse it is called *fully amphichiral*. Otherwise, if an amphichiral knot is not invertible, either it preserves its orientation in the mapping to its mirror image (*positive amphichirality*) or it reverses its orientation (*negative amphichirality*, which corresponds to the equivalence with the inverse mirror image).

1.6 Knot composition

Imagine to have the diagrammatic representation of two knots K_1 and K_2, to remove an arc from both of them and finally bridge the free endpoints by means of two new arcs, in such a way to end up with a single knot K_3 (see Figure 1.10). Note that the added arcs must not introduce any other crossings by overlapping with each other or with a pre-existing arc. What can be said about the resulting knot type?

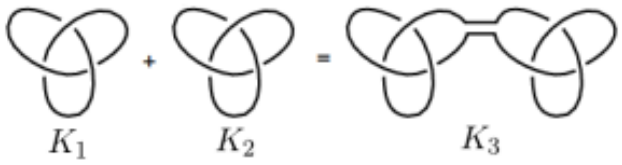

Figure 1.10: Connected sum of two left-handed trefoils. The resulting knot is called granny knot. See Other sources.

This procedure can be formalized through the notion of connected sum (indicated by #). The knot $K_3 = K_1 \# K_2$ is called *composite knot* and K_1, K_2 are its *factors*. Already from this operative construction, it's evident that composing an arbitrary K with the unknot gives K itself ($K \# 0_1 = K$): in other words, the trivial knot 0_1 is the neutral element of the connected sum operation.

Whenever a knot cannot be written as the connected sum of two non trivial knots, it is said to be a *prime knot*. The analogy with the natural number, as suggested also by the nomenclature, is straightforward: as they can be factorized in a product of prime numbers, in the same way every knot admits a decomposition into prime knots (*knot factorization*).
If it is clear that 1 cannot be expressed as a multiplication of two positive integers greater than 1, it is much less intuitive that the unknot 0_1 cannot result from the composition of two non trivial knots. The non existence of the "anti-knot" can be formally proven introducing the Seifert surfaces and other geometrical techniques. Since the number of prime knots exponentially increases with the crossing number $c(k)$, they have been completely tabulated only up to 16 crossings by Hoste et al. [10].

Another interesting aspect is how the handedness of a composite knot is influenced by that of its factors. Let us focus on the simple examples in Figures 1.10 and 1.11, where the factors'

handedness can be determined with the method of the Section 1.5 since their crossing number is 3. In this case the chirality of the composite knot can be evinced by the sign of the crossings algebraic sum: so $3_1^L \# 3_1^L$ and $3_1^R \# 3_1^R$ are respectively the left- and right- handed enantiomers of the granny knot, whereas the square knot $3_1^L \# 3_1^R$ is amphichiral being the sum of its signed crossing equal to 0.

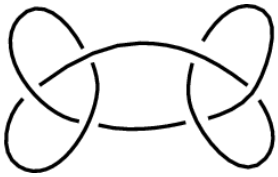

Figure 1.11: Square knot, obtained by doing the connected sum of two trefoils with different handedness.

1.7 Main knot families

It has been noted that among all the possible knots, there are some subsets whose elements share certain characteristics or topological properties. For this reasons knots have been classified in some main families, some of which are introduced in the following.

Alternating knots are an important category whose prime elements have been enumerated up to 22 crossings. A knot is said alternating if there exists a minimal diagrammatic representation where the crossing signs alternate along the path. However, even if a knot is alternating, there could be several diagrams in which the previous requisite is not satisfied. The alternating knots are common when the crossing number value is $\lesssim 10$ (the first instance of prime non-alternating knot being the 8_{19}), but there's a conjecture stating that as the crossing number increases, the fraction of alternating knots decreases exponentially to 0.

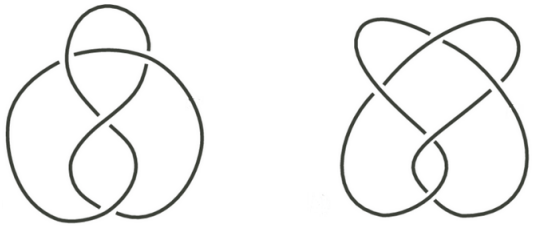

Figure 1.12: Examples of alternating knots: 4_1 (figure-8-knot) on the left and 5_2^R on the right

The unique knot that can be obtained by drawing a simple closed curve on the surface of a sphere without self-intersections is the unknot 0_1. Even if all the other non trivial knots don't have this property, it turns out that some of them can lie on the surface of an unknotted torus. For this reason the are called *torus knots*. To better describe a torus knot, let's introduce two more concepts: if a circle wraps the torus the short way (passing through the hole) it is called meridian,

otherwise if it does it the long way it is called longitude. Each torus knot can be characterized by a pair of integer numbers p, q, which inform of how many times the knot is wound around the torus respectively longitudinally and meridionally. In particular, to determine p and q, one can draw the knot together with a meridian and a longitude on the torus surface, and count how many times it crosses both of them. For a torus knot p and q are always coprime, and a torus knot is not trivial (different from the unknot) if and only if neither p nor q are ± 1, the sign standing for the winding direction. Some other properties are summarized in the following:

1. (p, q) is equivalent to (q, p).
2. (p, q) and $(-p, -q)$ have the same image but reversed orientation.
3. (p, q) and $(p, -q)$ are mirror images, meaning that a torus knot is always chiral.
4. for a torus knot $K = (p, q)$ with both $p, q > 0$, the crossing number is equal to $c(K) = \min\{q(p-1), p(q-1)\}$.
5. for a torus knot $K = (p, q)$ with both $p, q > 0$, the unknotting number is equal to $u(K) = \frac{1}{2}(p-1)(q-1)$.

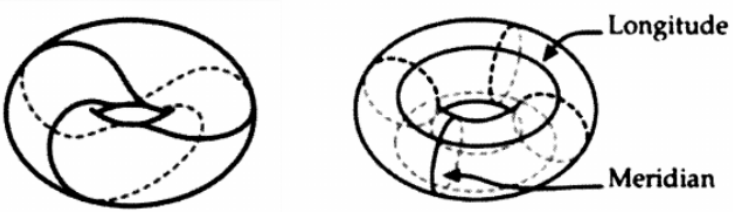

Figure 1.13: On the left: the trefoil 3_1 is the easiest torus knot. On the right: sketch of meridians and longitudes. Image from [1].

Another important family is that of *twist knots*: they can be obtained starting from a loop, adding some twist and finally clasping together the arcs at the two extremities. This construction makes immediately clear that the unknotting number $u(K)$ of a twist knot is always 1. Moreover, if a twist knot has n half twists, its crossing number is simply $c(K) = n + 2$. A part from the figure-eight-knot 4_2 and the 0_1, all the others twist knots are chiral.

1.8 Dowker notation

This section is not strictly necessary to understand the rest of the thesis

The Dowker notation is a sequence of even integer numbers which can be assigned to a knot diagram according to the following steps:

1. Move along the curve and add an increasing natural number to the sequence (starting from 1) each time that a crossing is encountered. Whenever an even label $2n$ is added while traveling on a crossing understrand, it should be replaced by $-2n$. At the end of this procedure each crossing will have both an even and odd label.

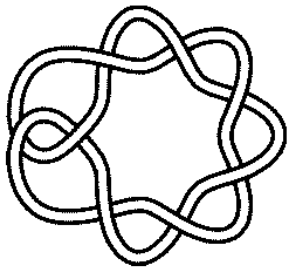

Figure 1.14: Example of a twist knot with 6 half twists (8_1). See Other sources.

2. Arrange this label pairs in increasing order of their odd component and finally retain only the even one. This will give a sequence of signed even number representing the Dawker code.

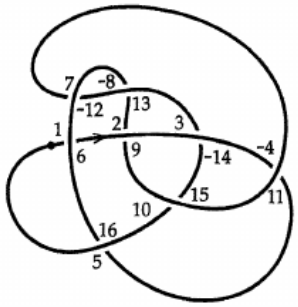

Figure 1.15: Example of crossings numbering for the Dawker code computation (in this case 6, −14, 16, −12, 2, −4, −8, 10). Image from [1].

The Dowker code, which is supposed to be like a fingerprint for a knot species, is actually inappropriate in some circumstances: indeed, in general, composite knots cannot be unambiguously determined by this notation, as well as the two topological enantiomers of a chiral knot. For that reason, more efficient knot invariants are presented in the following.

1.9 Polynomials

This section is not strictly necessary to understand the rest of the thesis

Another approach used for knot classification is that of computing from the diagrammatic representation a polynomial whose coefficients encode the information about the knot. Several polynomials have been developed during the last century as they turned out to be really powerful knot invariants. The key idea is to find a mathematical descriptors which does not vary if one performs a topology-preserving geometrical deformation. Even though more precise invariants are known, the knot polynomials strength lies in the fact that they are more suited for numerical computations.

1.9.1 Alexander polynomial

Historically the first to have been introduced, the *Alexander polynomial* Δ is a single variable t polynomial which can be computed from the knot diagram according to the following sequence of steps:

1. An orientation should be assigned to the diagram as well as a sign to each crossing, according to the right-hand rule.
2. Following the curve orientation, crossings and arcs are numbered from 1 to n.
3. A square matrix $M^{n\times n}$ is defined: its rows are relative to the crossings, while its columns to the arcs. For each crossing, the only non negative matrix elements are located in columns corresponding to the arcs involved in the crossing. Let's call i_c, j_c, k_c respectively the arc above the crossing c and the ones below, so that the only non negative entries of the row c will be $M(c, i_c)$, $M(c, j_c)$ and $M(c, k_c)$. In particular:
   (a) if the crossing c is positive $M(c, i_c) = 1 - t$, $M(c, j_c) = -1$ and $M(c, k_c) = t$.
   (b) if the crossing c is negative $M(c, i_c) = 1 - t$, $M(c, j_c) = t$ and $M(c, k_c) = -1$.
4. The Alexander matrix is obtained by removing arbitrarily a row and a column from M, and the Alexander polynomial by computing the determinant of the Alexander matrix.

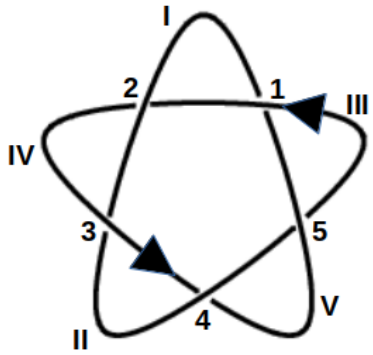

Figure 1.16: Crossings and arcs numbering for the 5_1^L Alexander matrix computation.

The Alexander polynomial computation for the knot 5_1^L is reported in the following as an example of the described procedure (the knot diagram is sketched in the Figure 1.16). Notice that all the crossings in the diagram are negative.

$$\begin{vmatrix} -1 & 0 & 1-t & 0 & t \\ 1-t & 0 & t & -1 & 0 \\ t & -1 & 0 & 1-t & 0 \\ 0 & 1-t & 0 & t & -1 \\ 0 & t & -1 & 0 & 1-t \end{vmatrix} \longrightarrow \begin{vmatrix} 0 & t & -1 & 0 \\ -1 & 0 & 1-t & 0 \\ 1-t & 0 & t & -1 \\ t & -1 & 0 & 1-t \end{vmatrix} =$$

$$= \begin{vmatrix} 0 & t & -1 \\ -1 & 0 & 1-t \\ t & -1 & 0 \end{vmatrix} + (1-t) \begin{vmatrix} 0 & t & -1 \\ -1 & 0 & 1-t \\ 1-t & 0 & t \end{vmatrix} = -1 + t - t^2 + t^3 - t^4$$

But what happens if one considers an arbitrary diagrammatic representation in place of the minimal one? As soon as the number of crossings increases, one would expect to have a bigger matrix size and so an higher polynomial degree. It seems that this behavior violates the polynomial requisite to be unaffected by geometric manipulations which preserve the knot topology. However, it can be shown that the Alexander polynomials resulting from different knot diagrams can only differ by a factor $\pm t^k$ with $k \in \mathbb{Z}$. By removing such a factor one ends up with the so called irreducible Alexander polynomial, which does not depend on the particular diagrammatic representation. Some other properties of the Alexander polynomial are listed in the following:

1. For an arbitrary knot K, $\Delta_K(t = 1) = \pm 1$.
2. The Alexander polynomials $\Delta_K(t)$ and $\Delta_K(t^{-1})$ are equals up to a factor t^k with $k \in \mathbb{Z}$.
3. The Alexander polynomial of a composite knot is equal to the product of the Alexander polynomials of its factor knots, i.e. $\Delta_{K_1 \# K_2}(t) = \Delta_{K_1}(t)\Delta_{K_2}(t)$.

However, there are some pair of different knots which cannot be distinguished by means of the Alexander polynomial (e.g. the 8_{20} and the granny knot have the same polynomial, and there are some species with $\Delta(t) = 1$ as the trivial knot). Another drawback of the Alexander polynomial is its incapability to discern the handedness of two enantiomers. Some other polynomials, such as the HOMFLY or the Jones polynomial, have been devised at a later time to overcome this ambiguities: although they can distinguish the knot chirality, there are still some knot species which cannot be discriminated.
Since the algorithms to extract one of the previous polynomials from the knot diagram can rapidly increase their computational cost with the number of crossings, it is recommended to simplify as much as possible the knot geometry before to run them (notice however that such a simplification is already not trivial).

Chapter 2

An overview of DNA biology

2.1 DNA structure

DNA is a large polymeric macromolecule with a double helical shape (experimentally discovered by Watson and Crick in 1953 [29] with the help of X-ray diffraction data furnished by Gosling and Franklin), whose backbones consist of an alternation of deoxyribose and phospate residues and whose internal part is occupied by four types of nitrogenous bases: *adenine, thymine, cytosine, guanine*. Such bases belong to two different classes of chemical species: indeed the adenine (A) and the guanine (G) are *purines*, while thymine (T) and cytosine (C) are *pyramidines*. The nitrogenous ring of both purines and pyramidines is linked to the DNA filaments by means of a covalent bond with the carbon-1 of the pentose sugar (deoxyribose). The nitrogenous bases together with the sugar are called *nucleosides*, that once linked to a phosphate group become *nucleotides*, the building blocks of DNA.

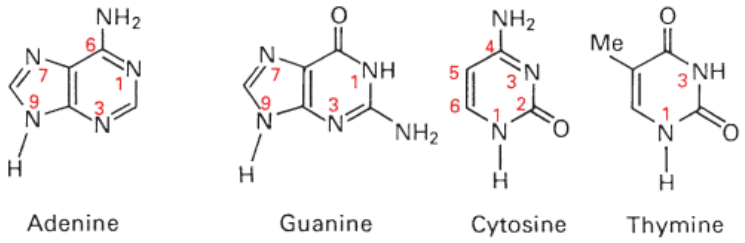

Figure 2.1: Nitrogenous bases structure, with IUPAC numbering for purines and pyramidines. Image from [3].

In addition to their biological importance, the nitrogenous bases have also the structural role of holding the two backbones together by means of hydrogen bonds. In doing that, they are characterized by a precise complementarity: in fact the adenine can establish a bond only with the thymine and the cytosine only with the guanine (no other pairings are allowed). Moreover, the strength of the chemical interactions AT and CG is slightly different in that they involve respectively 2 and 3 hydrogen bonds, a difference which may lead to sequence dependent effects in some physical phenomena. Still from the molecular shape point of view, the nitrogenous bases planes are found

to be approximately perpendicular to the helical axis direction. Notice that the hydrogen bonds help also the molecule stabilization in water, by constraining the hydrophobic bases to stay in the internal region of the helix and exposing the polar residues along the backbones to the solvent.

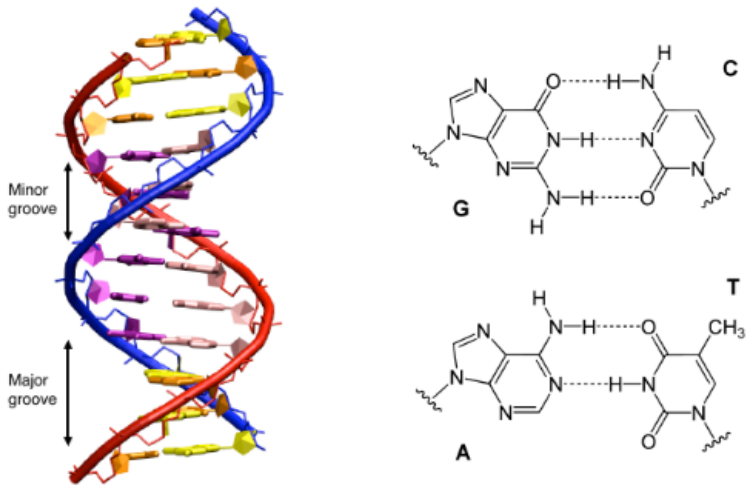

Figure 2.2: On the left: B-form DNA representation. The minor and major grooves are indicated on the left and the perpendicularity between nitrogenous bases and helical axis can be appreciated from this point of view. Image from [32] .On the right: chemical structure of AT and CG pairings, with hydrogen bonds indicated by the dashed lines.

Depending on the values and interplay of some parameters like humidity, salt concentration, etc, the DNA double helix can take on different arrangements. However, we will not enter into the details of the DNA structural polymorphism, and we will mainly focus on the so called B-form. The double helix has a width of 23.7Å, the distance between two nitrogenous bases along the backbones is 3.4Å and 10.4 base pairs are required for the two filaments to complete a full turn around the double helix axis (pitch of 3.53 nm). Minor and major grooves (explicitly pointed out in the Figure 2.2) alternate along the DNA duplex and their difference is biologically relevant: in fact, because of the backbones architectural disposition, the major groove side can more easily accommodate some DNA binding proteins which interact with the nitrogenous bases.

The DNA thickness given above is only due to the chemical structure of the macromolecule (bonds length) and does not take into account the fact that DNA is a polyelectrolyte. Indeed, the phospates along the backbones are negatively charged in solution because of the hydroxyl groups deprotonations. This leads to the development of an electric potential felt by the charged particles in the neighborhood. A lot of DNA binding proteins with a positively charged residues exploit such potential to bind the double helix. On the other hand, every negatively charged object tends to be repelled and feels the DNA presence as if it was effectively thicker. This effective radius consequence can be tuned by varying the salt concentration: indeed the counterions can screen the long range potential producing an exponential decay of the latter.

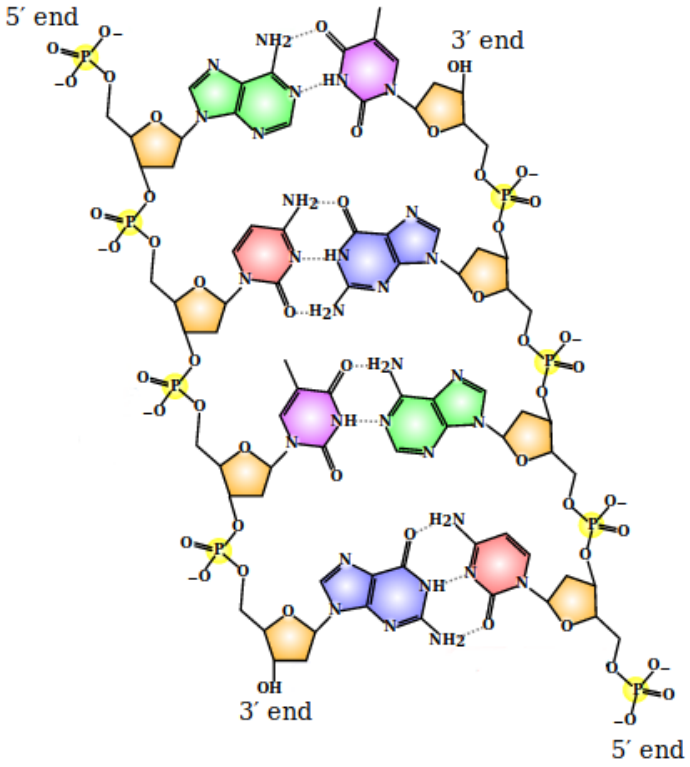

Figure 2.3: DNA chemical structure. The phosphates groups with their negatively charged residues are highlighted with yellow circles. The deoxyriboses are colored in orange, the adenines in green, the thymines in purple, the guanines in blue and the cytosines in red. See Other sources.

Antiparallelism As highlighted in Figure 2.3, the two ends of each strand are characterized by an asymmetry due to the fact that either the 3^{rd} or 5^{th} carbon of the deoxyribose rings on the two chain extremities miss the phosphodiester bond which would have linked them to the subsequent sugar residue. That's the reason why the two extremities are called $3'$ and $5'$ ends. The $5'$ extremity usually ends with a phosphate linked to the deoxyribose 5-carbon, while the $3'$ one ends with the -OH group of the deoxyribose 3-carbon. In the DNA double helix the two complementary strands always anneal in such a way to have opposite direction, and for that reason they are said to be antiparallel. A portion of genetic code or a DNA strand related structure developed toward the $5'$ end is called *upstream* (conversely, it is said *downstream* if it's toward the $3'$ end).

Elastic properties Both the chemical architecture and salt concentration in the solution are responsible for the elastic properties of the DNA. An idea of the molecule resistance to bending can be given by the *bending persistence length* l_p, which expresses the polymer orientation memory as one travels along the chain and which will be mathematically described in the Chapter 5. Whenever the salt concentration is low, the screening effect is less intense and the chain self avoidance due to the negative charges is more pronounced. In this case the chain is more rigid and its persistence length l_p is higher. Conversely, as the salt concentration increases the chain flexibility gets larger

and the l_p is lower. A typical value of l_p for a salt concentration of the order of ~ 100 mM which disregards the particular bases sequence is $l_p \approx 50$ nm.

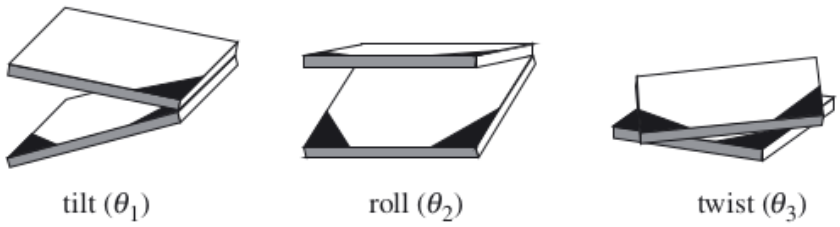

Figure 2.4: Some configurations of neighboring base-pairs: the tilt and roll angles θ_1 and θ_2 describe the local molecule bending, while θ_3 refers to the local torsion. Image from [18].

However, in some circumstances, sequence dependent effects can be quite important since they affect the elastic DNA properties. In fact, a part from the hydrogen bonds strength difference between AT and CG, it has been shown that sequence differences lead to a base stacking variation. As already said, the nitrogenous bases are hydrophobic, so that they try to minimize their water exposure to lower the energy by getting closer to each other. Usually this can be achieved by rotating of $\approx 30°$ about the helical axis the plane of two paired bases with respect to that of the underlying ones, in such a way to reinforce their hydrophobic and Van der Waals interactions. In some cases the particular sequence can influence the DNA mechanical properties by favoring tilted or rolled neighboring base-pairs as described in the Figure 2.4, and so it can affect the molecule bending persistence length.

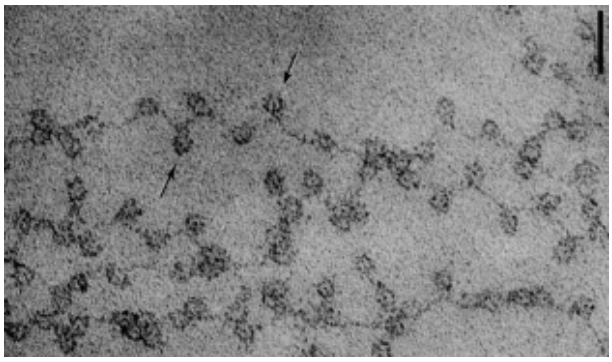

Figure 2.5: Electron micrograph of nucleosomal fibers and their "beads on a string" appearance. The nucleosomes are indicated by the black arrows. The size marker on the top-right correspond to 30 nm. Image taken from [17].

Chromatin architecture From an architectural point of view, it's quite interesting how long DNA macromolecules can fit into the nucleus of eukaryotic cells. In the case of human genome, which is made by $\sim 7,8 \cdot 10^9$ bps, the genetic material total length is about 2 m: this suggests that a high level of spatial and structural organization is reached in eukaryotic cells. Indeed the DNA

duplex is extremely well arranged to attain a suitable compaction, and in order for the packing ratio to become fairly high a hierarchical organization is realized. Such architecture goes under the name of *chromatin*. As a first step, the genetic material appears as a series of subunits, the *nucleosomes*, composed by DNA duplex wrapped around a bunch of 8 *histone* proteins in a left-handed superhelical fashion, for a total of approximately 1.7 turns. Two subsequent nucleosomes are joined by means of linker DNA, a spacer portion of DNA duplex whose length varies according to the species ($30 \div 40$ bps on average). An electron micrograph of the "beads on a string" nucleosomes disposition from [17] is reported in Figure 2.5. Although already at this stage the fibers length is reduced by a factor of ~ 7, the chromatin needs a further processing to be properly accommodated into the nucleus (whose diameter is usually of the order of 10 μm). When the salt concentration *in vitro* is high enough (order of 100 mM), the nucleosomal fibers have been observed to condense in *solenoids*, whose packing ratio can reach 40 (it is still unclear if they form also *in vivo*). The structural complexity keeps increasing, up to reach the chromosomes formation. During the cell

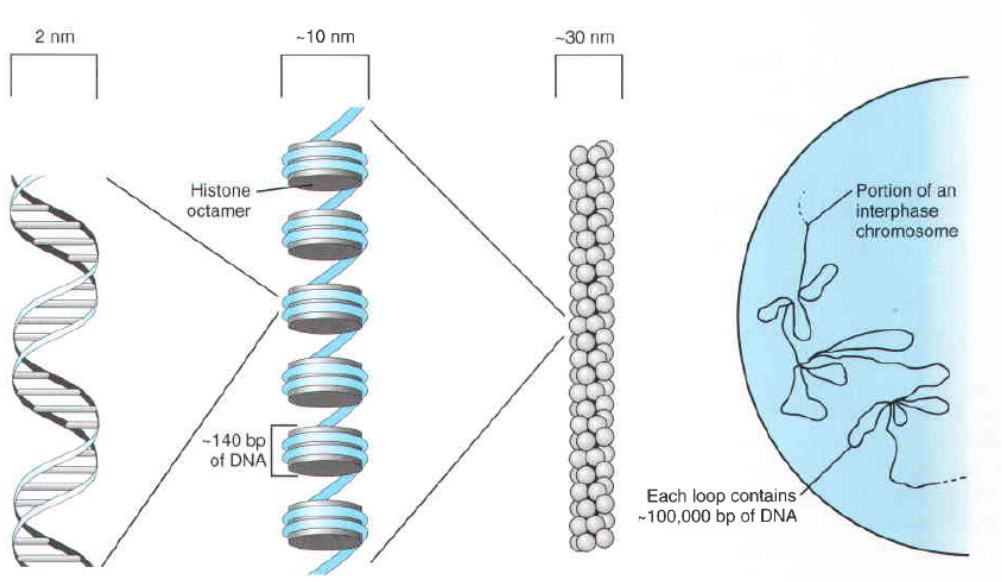

Figure 2.6: Representation of the chromatin hierarchical organization. From left to right: a simple DNA duplex, a nucleosomal fiber, a solenoidal structure, some supercoiled loops of a decondensed interphase chromosome. Image from [19].

cycle, the genetic material experiences different degrees of condensation. In particular, when the mytotic phase begins the chromosomes can be more easily microscopically visualized. The described organization is crucial to allow the regular implementation of the DNA biological roles (e.g. the DNA replication) and to prevent dangerous entanglements.

2.2 Supercoiling

As already mentioned, the DNA has a pitch of 10.4 bps in its torsionally relaxed state. However it has been shown that bacterial plasmids are dynamically kept, by means of the action of some

enzymes, slightly underwound *in vivo*. This condition is said to be of *negative supercoiling*. Conversely, whenever the DNA is twisted more tightly, it is *positively supercoiled*. To get a glimpse of the supercoiling implications on the molecule geometrical conformation, imagine to start with an energetically relaxed linear fragment of DNA. In order to minimize the torsional energy, the number of helical turns should be given by $\frac{N_{bps}}{10.4}$, where N_{bps} is the number of base pairs of the linear DNA fragment. The bending energy is clearly minimized being the molecule in a straight configuration. Suppose that the number of helical turns is an integer, and imagine to circularize the chain without introducing any further twisting. The molecule will be stable in the configuration of planar ring, since the torsional level didn't change during the closure procedure and the circular shape is the one which minimizes the bending energy among all the closed configurations. Now, suppose to start again from the same linear configuration as before, and to add an extra integer number of helical turns by keeping fixed one end and twisting the other. If one of the two extremities is freed, the molecule will untwist to relax the torsional strain and will come back to the initial configuration. If instead the molecule is circularized with the extra twist and then freed, it won't be able

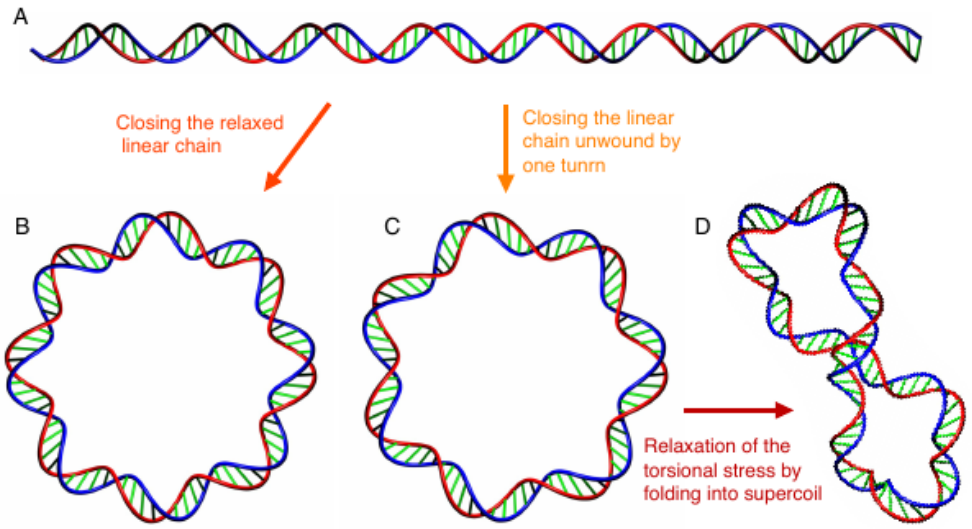

Figure 2.7: (A): energetically relaxed linear DNA fragment, with 8 helical turns. (B): circular chain obtained by closing the linear one without introducing any twist (still torsionally relaxed). (C): circular chain obtained by under-winding of one turn the linear fragment and then circularizing the molecule. This configuration is not torsionally relaxed and it will evolve in (D), where the balance between bending and twisting stiffness results in a figure-eight contorted conformation. Image from [32].

to completely get rid of the torsional stress and it will arrange into a more compact shape called supercoil. The particular geometry assumed by the macromolecule in a supercoiled state, and so the compaction itself, is strongly affected by the balance between bending and twisting stiffness. The supercoiling fingerprint is the emergence of helically wrapped structures called plectonemes (predominant in prokaryotic organisms with small circular chromosomes) or solenoidal supercoils particularly suitable for packaging the eukaryotic chromatin. A part from the genome packaging, the supercoiling is involved in several other biological processes such as DNA synthesis, transcrip-

tion, replication, etc. A more quantitative description of supercoiling and superhelical density will be given in Chapter 3.

2.3 Topoisomerases

During biological processes such as transcription or replication, which will be more precisely treated in Section 2.4, the DNA can locally accumulate some torsional stress. If these strain excesses are not promptly removed, they could alter the normal course of the processes (e.g. they can block DNA or RNA polymerases moving along a strand). Topoisomerases are enzymes able to underwind/overwind the DNA thanks to their capacity to break the phospate backbones and reseal them, and they have the task of tuning the supercoiling level wherever is needed. Moreover, they are responsible to tune the entanglement degree of the genetic material, which can be regulated thanks to topological operations as catenation/decatenation or knotting/unknotting.
Depending on the number of strands they work on, topoisomerases are divided in two main families: type I and type II topoisomerases.

2.3.1 Type I topoisomerases

This family of topoisomerases has the ability to change the supercoiling level of the chain through a single strand breaking and re-ligation. The cleavage mechanism involves the formation of a phosphodiester bond between a strand extremity and the enzyme. Once that the cleavage has occurred, the unwinding process happens either by rotating the duplex around its intact strand or by letting the intact strand pass through the temporary break. This mechanisms are respectively called *swiveling* and *strand passage*. A further classification requires to pay more attention to the

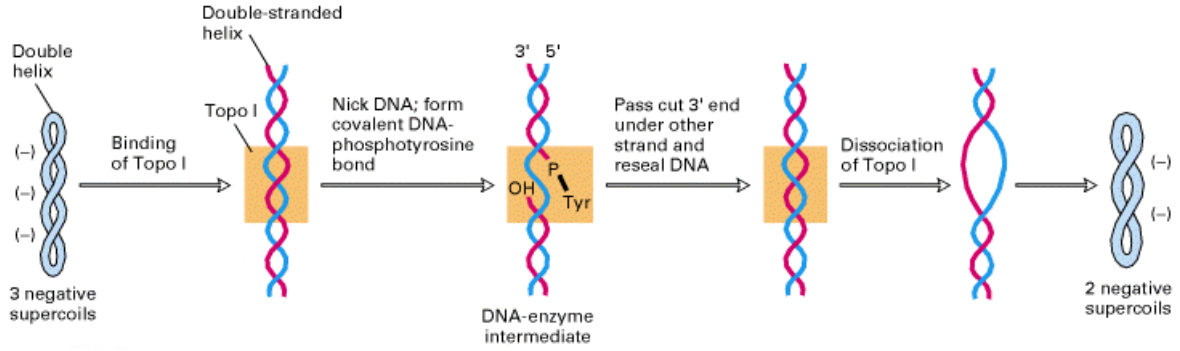

Figure 2.8: From left to right: a closed DNA duplex chain is characterized by 3 negative supercoils. A type IA topoisomerase cleave to one strand of the duplex and it forms a phopshotyrosine bond with the $5'$ end of the red filament. The blue backbone passes through the break, decreasing the number of negative supercoils by 1. See Other sources.

exact position of the cleavage: the subfamily IA enzymes form a bond with a $5'$ phosphate, whereas those which link to the $3'$ phosphate belong to the IB subfamily. Moreover, type IA topoisomerases adopt the strand passage mechanism to accomplish their task, while those of type IB under-wind

the chain through the swiveling. A difference between these two mechanisms is that with the former the number of helical turns can only change by one unit, whereas the latter allows for a variation of helical turns also greater than one.

2.3.2 Type II topoisomerases

At difference with the type I family, the torsional energy is dissipated by the type II topoisomerases with a double strand break. Once the backbones are cut, another duplex portion passes through the opening and then the original duplex is resealed. This mechanism is called *cross inversion* and it allows the type II topoisomerases to change the chain topology. At difference from the type I enzymes, which are monomers, the type II topoisomerases have a dimeric structure since they cleave both the duplex backbones. In this case, the cleavage involves the formation of two phosphotyrosine bonds (one for each dimer sub-unit) with the $5'$ DNA ends. Depending on their structure and biochemistry, type II topoisomerases are specialized in two main sub-classes: type IIA, which includes bacterial gyrase, topo IV and eukaryotic topo II, and type IIB including topo VI. Thanks to their ability to disentangle intertwined duplexes or chromosomes by means of both inter- and intra- molecular double strand passages, type II topoisomerases regulate the genetic material topological complexity. Furthermore, they can help also the supercoiling level tuning by changing the supercoils number by 2 units. Again, a more rigorous mathematical treatment of topoisomerases action on the supercoiling level will be developed in Chapther 3.

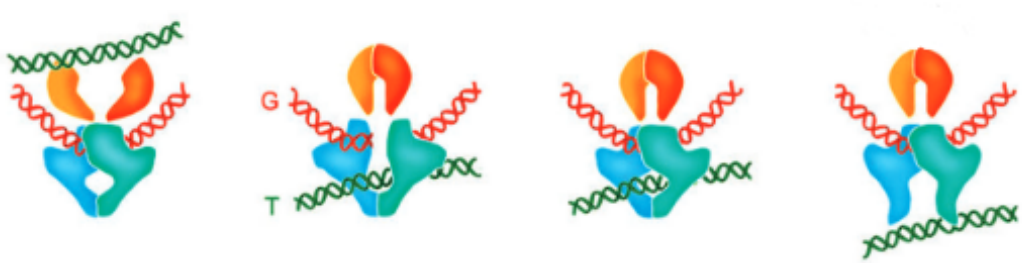

Figure 2.9: Cross inversion mechanism of the type IIA topoisomerase. The enzyme presents three gates, called from top to bottom N-gate, DNA-gate, C-gate. They are mechanically coupled in such a way that a gate can be open only if the other two are closed. The double strand which is going to be gated (broken) is called G-segment, while the intact duplex which passes through the gate is called T-segment. Image from [20].

2.4 Biological role

From a biological point of view, DNA carries the genetic information needed to all organisms in order to properly carry out their vital functions, from gene expression to growth and reproduction. The genes are the stretches of DNA where such information is stored. In the following, a brief zoom on the principal processes which involve DNA is performed.

2.4.1 Replication

DNA replication is the process through which new DNA macromolecules are synthesized starting from a pre-existing one. The key aspect which allows a DNA molecule to form an exact duplicate of itself is base pairing. The *helicase* enzyme opens the double helix into a replication fork in such a way that each strand may serve as a template for a new filament. In doing that, the helicase provokes a supercoiling accumulation in the parent duplex, which needs to be relieved by topoisomerases in order for the replication fork to properly proceed. At this point, the new strands synthesis occurs thanks to the enzyme DNA polymerase, which is able to elongate the new filament by adding nucleotides through the formation of phosphodiester bonds, respecting the base complementarity. However, the DNA polymerase is not able to initiate the synthesis process, and it requires the presence of an already paired small strand portion to start its work. For that reason, the enzyme *primase* (which belongs to the family of RNA polymerases) cleaves to the template filament $3'$ end and synthesizes a very short RNA strand called *primer*. At this stage, the DNA polymerase reach the primer $3'$ end and begins the DNA duplex elongation. Notice that, at the end of this process, the product will be an RNA-DNA duplex, and the primer portion will be removed by the action of a specialized enzyme called *RNase H*.

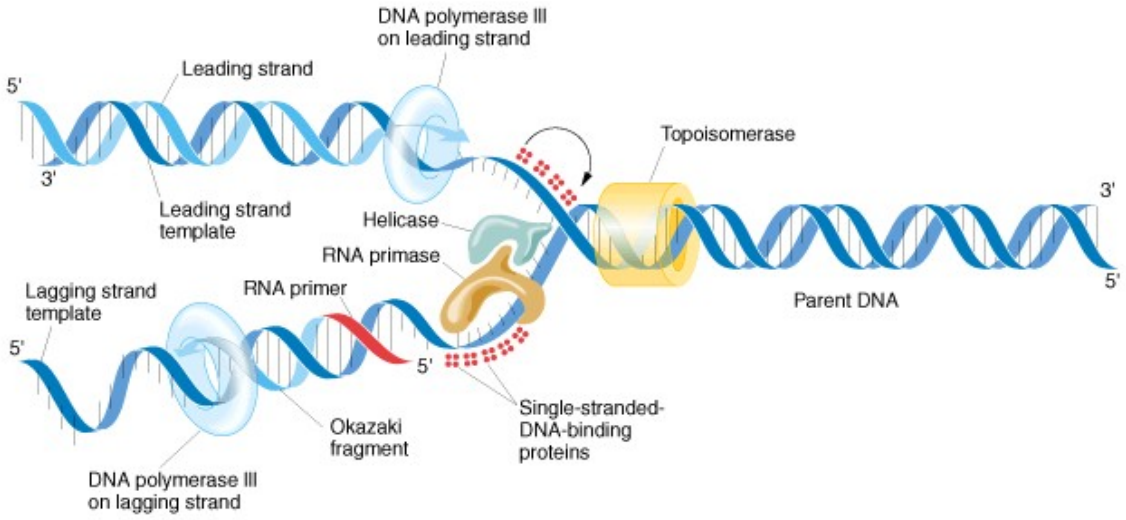

Figure 2.10: Sketch of replication process and main enzymes which come into play. Topoisomerases relieves the supercoiling excess generated by the replication fork progression. DNA polymerases elongate the duplex both in the leading and lagging strands. RNA primase and primer are present every time that an Okazaki fragment is being synthesized on the lagging strand (and also at the beginning of the leading strand). See Other sources.

Another DNA polymerase peculiarity is the fact that it can carry out its work only by reading the template filament in $3' \to 5'$ direction. However, since the two template strands are antiparallel, only one of them can be readily copied, and it goes under the name of *leading strand*. Notice that the complementary filament of the leading strand will be synthesized in $5' \to 3'$ direction.

For what concerns the $5' \to 3'$ template filament, called *lagging strand*, it cannot be read in the same direction of the replication fork progression because of its opposite orientation with respect to the one required by DNA polymerase to proceed. To overcome this obstacle, the lagging strand complementary filament will be synthesized from the replication fork toward the lagging strand free extremity, in small series of $\sim$ 1000 nucleotides called *Okasaki fragments*. Similarly to the leading strand complementary filament initiation, every Okasaki fragment starts with an RNA primer. Once synthesized, the Okasaki fragments are joined up by the DNA polymerase I, which is able to substitute the RNA primers with DNA, and by DNA ligase which adds the missing phosphodiester bonds to complete the backbone. Since only the leading strand complementary filament is synthesized in a continuous fashion, the replication process is called *semi-discontinuous*. Moreover, in addition to being exact replicas of the parent, each daughter molecule contains one of its strands, and the process is said to be *semi-conservative*. The replication ends whenever another replication fork is encountered or when a telomere (the chromosome final part) is reached.

2.4.2 Transcription

The mechanism through which the genes can be expressed is so important that its explanation goes under the name of *central dogma of molecular biology*. Transcription represents the first step of this process, where the genetic information encoded in the DNA is copied into RNA thanks to the action of the enzyme RNA polymerase. Even though both replication and transcription are copying processes and use the same DNA as a template, they are at the same time quite different. If in the replication the whole chromosomes have to be duplicated, only the regions containing a gene needs to be transcribed during transcription. The transcription product can be mRNA if the gene contains a sequence which encodes a protein, or it can be non-coding RNA (such as rRNA or tRNA) otherwise: all of them are essential in order for the protein synthesis to properly occur. Each cell transcribe different DNA segments, also called transcription units, on the basis of its needs. Notice that only a small percentage of the genetic material is coding DNA, since in the genes themselves a lot of non-coding sequence are present and the remaining part of the genome contains a lot of nonfunctional repetitive segments.

Usually only one strand plays the role of template during transcription, and the newly synthesized filament together with the template strand make a DNA-RNA hybrid duplex. Analogously to DNA polymerase, also the RNA polymerase has the $3' \to 5'$ as preferential direction to read the template. The RNA strand built in accordance to the base complementarity rule is called *primary transcript*. Since the transcript has direction $5' \to 3'$, the DNA duplex nontemplate strand is often called *sense strand*, while the template one is called *antisense strand*.
RNA polymerase is not able to start its activity by itself. For this reason the genes are usually preceded by short sequences which can be recognized by *transcription factors*. Their role is to anchor the DNA in the region occupied by the promoter, usually located upstream of the coding sequence, and activate the RNA polymerase. Once activated, the RNA polymerase-promoter complex starts to open into a *transcription bubble* of about 14 nucleotides. In this region the DNA is practically single stranded and the copying process can start. However, the local DNA unwinding necessary to allow the transcription bubble formation produce an excess of positive supercoiling ahead as well as of negative supercoiling behind. Analogously to the replication, the resolution of this side effect is attributed to the topoisomerases.

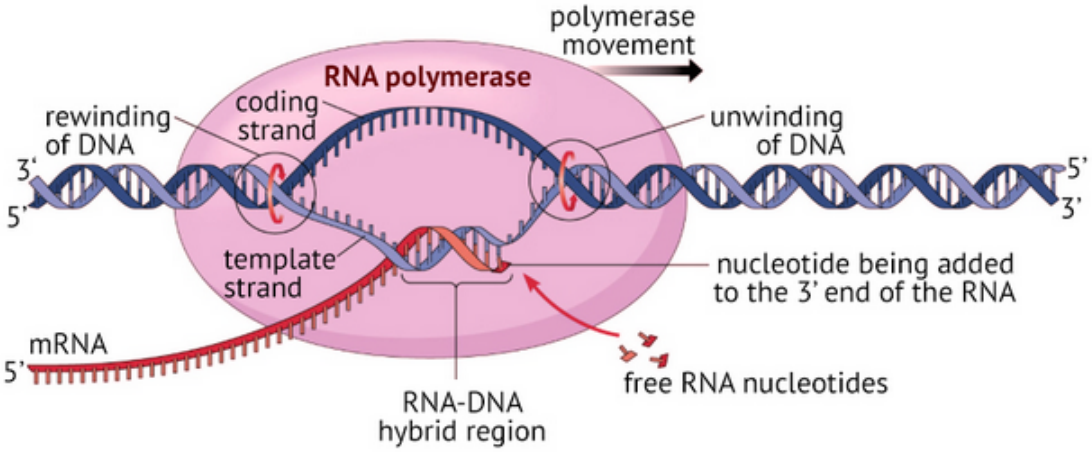

Figure 2.11: Transcription process representation. The RNA polymerase reads the $3' \to 5'$ template strand (the lighter one) and progressively adds the RNA nucleotides according to the base complementarity rule. Remember that in RNA the thymine is substituted by the uracil. As soon as the mRNA strand is synthesized, it detaches from the template strand. The DNA unwinding and rewinding highlighted on the two sides of the transcription bubble are responsible for the supercoiling accumulation in the rest of the duplex. See Other sources.

Once it has been synthesized, the RNA usually needs to be processed. Indeed the majority of genes, both those encoding a protein and those which don't, have only a small percentage of meaningful sequences which require to be read. In particular, the primary transcript obtained by copying the entire gene may be formed by both *exons*, which carry the genetic information, and *introns*, which are usually discarded. The series of operations through which intronic parts are removed and exons are joined up into a shorter RNA strand is called *RNA splicing*. Once that the RNA product has been cleaned up by the intronic regions and the RNA splicing has been carried out, one says that the precursor messenger RNA (pre-mRNA) has been turned into a mature messenger RNA (mRNA).

2.4.3 Translation

In eukaryotic cells the mRNA resulting from the transcription process can leave the nucleus, where it has been synthesized, passing through the nuclear pores. Once in the cytoplasm it reaches some organelles called *ribosomes*, whose RNA processing is crucial for the proper genes expression to happen. Indeed they are the agents where the protein "manufacturing" takes place. Differently, in prokaryotic organisms the mRNA can couple to the ribosomes also when the transcription process is still not completed, because of the absence of a nucleus-like structure. In this situation, the translation process starts from the free mRNA $5'$ end, which has already been synthesized, while the $3'$ is paired to the template DNA strand and it is still being processed (see the Figure 2.11).

Ribosomes consist of two main subunits (50S the bigger and 30S the smaller for prokaryotes, 60S and 40S for eukaryotes) composed by rRNA and ribosomal proteins. Notice that whenever

they are not directly involved in the translation process, the two subunits are separated and free to move in the cytoplasm (in alternative, few of them can be anchored to the endoplasmic reticulum). Only when a newly transcribed mRNA is ready to be read, they join together to form a functional RNA-protein complex. The ribosome catalytic function is due to the rRNA, whereas the proteins, which mainly lies on the subunits surface, help to the organelle stabilization. The genetic information stored in the mRNA is read in triplets called *codons*, whose sequences are related to a particular amino acid. Notice that, since each nucleotide can be chosen among 4, we have a total of 64 possible codons, but only 20 types of amino acids. This means that the genetic code is degenerate, and each amino acid corresponds on average to 3 codons.

The translation process begins when the smaller subunit together with a tRNA linked to methionine amino acid, encounter the starting codon AUG and recruit the bigger subunit. At this stage the ribosome is activated and it is ready for protein synthesis, performed by moving along the mRNA, which serves as a template, reading the codon triplets one by one and adding the corresponding amino acid to the growing polypeptide chain with tRNA help. The tRNA molecules are structured as follows: on one side they can bind to an amino acid (aminoacyl-tRNA) or to a polypeptide chain (peptidyl-tRNA) and on the other they exhibit a 3-nucleotide sequence called *anticodon*, complementary to the codon and crucial for the molecule recognition. The activated ribosome present 3 sites between their subunits: the A-site, which host an aminoacyl-tRNA, the P-site containing the peptidyl-tRNA with the polypeptide chain, and the exit E-site. Since the mRNA reading is performed by the ribosomes in $5' \rightarrow 3'$, the sites order is E-P-A $5'$ to $3'$. As

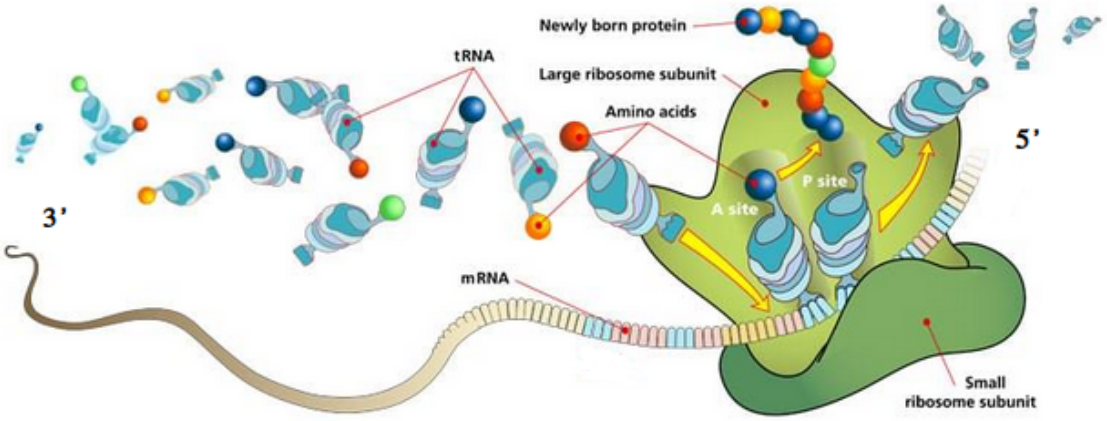

Figure 2.12: Translation process representation, with an active ribosome traversing the mRNA in $5' \rightarrow 3'$ direction. The exit E-site in correspondence of the right ribosome side is omitted, and only empty leaving tRNA molecules are depicted. The amino acid on the A-site is forming a peptide bond with the polypeptide chain, which is going to be transferred from P- to A-site. Other aminoacyl-tRNA molecules are waiting to reach their corresponding codon. See Other sources.

long as the ribosome proceed along the mRNA, a new aminoacyl-tRNA lands on the A-site and its

amino acid forms a peptide bond with the growing protein anchored to the P-site, so that the chain is transferred to the A-site. Now empty, the P-site tRNA is ready to be expelled, and the whole machinery is translocated toward mRNA $3'$ end. The translation process ends when the A-site moves in correspondence of a stop codon: in place of a charged tRNA molecule, such a codon binds a release factor and the polypeptide chain is dissociated from the ribosome.

Chapter 3

DNA geometry and topology

In order to characterize DNA geometrical and topological aspects, its strands can be modeled as a pair of helical curves $\boldsymbol{\phi}$, $\boldsymbol{\varphi} : \mathbb{R} \to \mathbb{R}^3$ wound in a right handed manner around a cylindrical surface with axis defined by the curve $\boldsymbol{\alpha} : \mathbb{R} \to \mathbb{R}^3$. Both prokaryotic and eukaryotic organisms present examples of DNA molecules constituted by circular closed chains (e.g. bacterial plasmids and yeast minichromosomes), which require the imposition of proper boundary conditions to be modeled, in such a way to match the curve extremities. A deeper understanding of the DNA geometrical and topological quantities and a mathematical formalization of the relationships to which they undergo, might shed new light on some biological phenomena explanation and elucidate their mechanisms.

3.1 Geometrical descriptors

A rigorous study of supercoiling is encouraged by its wide presence in many DNA biological processes and the possibility to give a further insight into problems inherent to transcription, replication, DNA packaging and compaction, criteria behind some enzymes action and more. For that reason, an appropriate geometrical treatment of DNA requires the introduction of *twist* and *writhe* as the main actors involved in the supercoiling quantification.

3.1.1 Twist

DNA twist, Tw, describes the winding of a backbone strand, e.g. $\boldsymbol{\phi}$, around the helical axis $\boldsymbol{\alpha}$ (or equivalently around the other filament $\boldsymbol{\varphi}$). Although for closed chains it may be tempting to believe Tw to be an integer, it is actually a real number as shown by configurations whose helical axis $\boldsymbol{\alpha}$ is not planar but curved in $\mathbb{R}^3$.

To define Tw, imagine to consider a cross section of the cylindrical surface representing the DNA duplex: this will determine a point $\boldsymbol{\alpha}(t)$ on the axis, and a point $\boldsymbol{\phi}(t)$ on the backbone strand $\boldsymbol{\phi}$. Let us denote the vector lying on the cross section which goes from the axis to the backbone with $\boldsymbol{z}_{\alpha\to\phi}(t) = \boldsymbol{\phi}(t) - \boldsymbol{\alpha}(t)$ and its versor with $\boldsymbol{v}_{\alpha\to\phi}(t) = \frac{\boldsymbol{z}_{\alpha\to\phi}(t)}{|\boldsymbol{z}_{\alpha\to\phi}(t)|}$ (see Figure 3.1). Intuitively, Tw corresponds to the number of full turns that $\boldsymbol{v}_{\alpha\to\phi}(t)$ makes rotating about the helical axis $\boldsymbol{\alpha}(t)$ as the curve parameter t increases. Let's define the unit vector tangent to the helical axis at the point $\boldsymbol{\alpha}(t)$ as $\boldsymbol{u}(t)$. Since $\boldsymbol{u}(t)$ is perpendicular to the whole cross-section individuated by the

parameter t, it must also be perpendicular to $\boldsymbol{v}_{\alpha\to\phi}(t)$. The local contribution to the rotation of

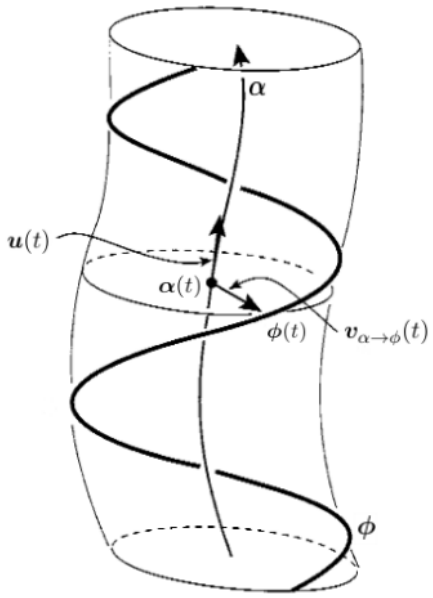

Figure 3.1: Sketch of the $\boldsymbol{\phi}$ backbone around the curved cylinder and its helical axis $\boldsymbol{\alpha}$. At the cross-section level parametrized by t, the unit vector tangent to the helical axis $\boldsymbol{u}(t)$ and $\boldsymbol{v}_{\alpha\to\phi}(t)$ are represented. Image adapted from [9].

$\boldsymbol{v}_{\alpha\to\phi}(t)$ around the helical axis is obtained by projecting the variation $d\boldsymbol{v}_{\alpha\to\phi}(t)$ along the direction perpendicular to both $\boldsymbol{v}_{\alpha\to\phi}(t)$ and $\boldsymbol{u}(t)$. So,

$$dTw(t) = d\boldsymbol{v}_{\alpha\to\phi}(t) \cdot [\boldsymbol{u}(t) \wedge \boldsymbol{v}_{\alpha\to\phi}(t)]$$

The total twist Tw can be obtained by integrating the previous expression along the curve $\boldsymbol{\alpha}$ (the normalization $\frac{1}{2\pi}$ converts the unit of measurement from radians to number of turns):

$$Tw = \frac{1}{2\pi}\int_{\boldsymbol{\alpha}} dTw(t) = \frac{1}{2\pi}\int_{\boldsymbol{\alpha}} [\boldsymbol{u}(t) \wedge \boldsymbol{v}_{\alpha\to\phi}(t)] \cdot d\boldsymbol{v}_{\alpha\to\phi}(t)$$

Example: open helix with straight axis Consider a fragment of linear DNA with n turns around its straight axis of length L (which for simplicity corresponds to the z-axis). The curve $\boldsymbol{\phi}$ describing one strand is parametrized by the curvilinear variable t in the following way:

$$\boldsymbol{\phi}(t) = (\rho\cos\Big(\frac{2\pi t}{p}\Big), \rho\sin\Big(\frac{2\pi t}{p}\Big), t)$$

with $t \in [0, L]$, ρ the helical radius and $p = \frac{L}{n}$ the helical pitch. Since the cross-section lies in the xy-plane for all value of t, the vector $\boldsymbol{v}_{\alpha\to\phi}(t)$ can be easily computed:

$$\boldsymbol{v}_{\alpha\to\phi}(t) = (\cos\Big(\frac{2\pi t}{p}\Big), \sin\Big(\frac{2\pi t}{p}\Big), 0)$$

In order to compute the integral one needs the vector product

$$\boldsymbol{u}(t) \wedge \boldsymbol{v}_{\alpha\to\phi}(t) = \begin{vmatrix} \boldsymbol{x} & \boldsymbol{y} & \boldsymbol{z} \\ 0 & 0 & 1 \\ \cos\left(\frac{2\pi t}{p}\right) & \sin\left(\frac{2\pi t}{p}\right) & 0 \end{vmatrix} = (-\sin\left(\frac{2\pi t}{p}\right), \cos\left(\frac{2\pi t}{p}\right), 0)$$

and the derivative with respect to the curvilinear variable of $\boldsymbol{v}_{\alpha\to\phi}(t)$:

$$\frac{\partial \boldsymbol{v}_{\alpha\to\phi}(t)}{\partial t} = (-\frac{2\pi}{p}\sin\left(\frac{2\pi t}{p}\right), \frac{2\pi}{p}\cos\left(\frac{2\pi t}{p}\right), 0)$$

Finally, substituting the previous expressions in the twist integral:

$$\begin{aligned} Tw &= \frac{1}{2\pi}\int_{\boldsymbol{\alpha}} [\boldsymbol{u}(t) \wedge \boldsymbol{v}_{\alpha\to\phi}(t)] \cdot d\boldsymbol{v}_{\alpha\to\phi}(t) = \frac{1}{2\pi}\int_0^L [\boldsymbol{u}(t) \wedge \boldsymbol{v}_{\alpha\to\phi}(t)] \cdot \frac{\partial \boldsymbol{v}_{\alpha\to\phi}(t)}{\partial t} dt \\ &= \frac{1}{2\pi}\int_0^L \frac{2\pi}{p}[\sin^2\left(\frac{2\pi t}{p}\right) + \cos^2\left(\frac{2\pi t}{p}\right)]dt = n \end{aligned}$$

Thus, the complete calculation yields the expected result $Tw = n$. Similar calculations for more articulated helical axis configurations are reported in full detail in a paper by White and Bauer [31].

3.1.2 Writhe

As proved by Călugăreanu in 1961 [8], the writhe Wr is an observable that can be defined for a single curve, which in our case is the DNA duplex helical axis $\boldsymbol{\alpha}$. Qualitatively, the writhe captures to what extent a curve is coiled in the 3-dimensional space. Thus, it can be a good candidate to characterize the DNA supercoiling level. Analogously to what has been done for knot diagrams in Chapter 1, a $2D$ diagram can be extracted from an oriented curve by projecting it onto a plane while preserving the information regarding over- and under-passes by drawing in discontinuous way the under-strand in correspondence of each crossing. Moreover, all the crossings should be labeled with a $+$ or $-$ sign according to the right hand rule. This $2D$ diagram could be thought of as an almost planar realization of a curve which exploits the third dimension to a perturbative extent and only in correspondence of a strands juxtaposition. Its writhe, which can also be interpreted as the *projected writhe* of the original curve with the appropriate observation direction, is equal to the algebraic sum of the signed crossings. In general, to compute the writhe Wr of an arbitrary curve living in $\mathbb{R}^3$, one needs to average over the projected writhing numbers obtained from all the possible observation points. If one defines with $Wr_{proj}(\boldsymbol{v})$ the projected writhe relative to the observation direction $\boldsymbol{v}$, the Wr will be given by:

$$Wr = \frac{1}{4\pi}\int_\Omega Wr_{proj}(\boldsymbol{v})\, dS$$

with integration domain

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^3 \,|\, x_1^2 + x_2^2 + x_3^2 = 1\}$$

Notice that although for certain curves there may be some unitary vectors $\boldsymbol{v}$ for which $Wr_{proj}(\boldsymbol{v})$ is not properly defined, they are a zero measure set with respect to Ω. Intuitively, if two different

DNA duplex are almost juxtaposed to each other, they will contribute $\approx \pm 1$ to the writhe Wr, since a crossing will result from almost all the projections. Similarly, every planar configuration will have $Wr = 0$ since all non singular projections do not originate apparent crossings.
An equivalent method to compute the writhe is resorting to the following Gauss integral:

$$Wr = \frac{1}{4\pi} \int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\alpha}} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1)$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are two points on the integration curve $\boldsymbol{\alpha}$. However, the numerical evaluation of this kind of integral can be pretty expensive from a computational point of view, and approximated expressions have been devised to accelerate the calculations. In particular, some of them take advantage of the fact that in many numerical simulations coarse grained models are exploited to deal with DNA, which is often represented by a series of straight segments (polygonal closed chain). Some methods to compute the Wr of supercoiled DNA are discussed in full detail in a paper of 2000 by Klenin and Langowski [14]. Since one of them has been used in the Monte Carlo simulations presented in Chapter 5, its description is reported in the Appendix A.

3.2 Topological descriptor: Linking number

The *linking number* Lk is a topological quantity which can be defined for a pair of closed curves (e.g. the two edges of a ribbon or $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ representing the two strands in the case of DNA). It assumes only integer values if the curves live in $\mathbb{R}^3$ and it can be intuitively thought as the number of times that a curve is wound around the other. Analogously to the writhe Wr, the linking number Lk of two curves $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ admits a description in terms of a Gauss integral (also called linking integral):

$$Lk = \frac{1}{4\pi} \int_{\boldsymbol{\phi}} \int_{\boldsymbol{\varphi}} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1)$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are points respectively on $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$.
Surprisingly, as long as the two curves are deformed in a way which does not involve any break, their linking number does not change. In principle, in the case of DNA, each strand could be continuously deformed into the helical axis $\boldsymbol{\alpha}$. This means that the same Lk value is found if one substitutes a backbone with the helical axis in the Lk definition given above:

$$\begin{aligned} Lk &= \frac{1}{4\pi} \int_{\boldsymbol{\phi}} \int_{\boldsymbol{\varphi}} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) = \frac{1}{4\pi} \int_{\boldsymbol{\alpha}} \int_{\boldsymbol{\varphi}} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) \\ &= \frac{1}{4\pi} \int_{\boldsymbol{\phi}} \int_{\boldsymbol{\alpha}} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) \end{aligned}$$

3.2.1 White-Călugăreanu-Fuller relation

One of the main results of the so called ribbon theory, is the White-Călugăreanu-Fuller equation

$$Lk = Tw + Wr$$

which establishes a relation between the two geometrical descriptors previously introduced and the linking number [11, 7, 30]. The crucial observation is that every DNA duplex deformation which

disallows strands breaks, affects Tw and Wr in such a way that their sum remains unaltered. For this reason, the linking number Lk is said to be a *topological invariant*.

For a more intuitive explanation, let us consider some examples. Imagine to have a torsionally relaxed DNA linear fragment (with n helical turns) and to circularize it into a planar ring. Since the Wr of a planar configuration is always 0, one has that in this case the linking number coincides with the twist, $Lk = Tw = n$.
Suppose now that some extra turns are added/removed from the linear DNA fragment. The planar circular configuration immediately after the closure will have again $Lk = Tw = n + n_{extra}$. However, as soon as the DNA duplex is freed, it starts to evolve in order to distribute its mechanical stress between bending and twisting. In doing that, it acquires a 3 dimensional shape that makes its writhe $Wr \neq 0$. In particular, Wr will increase in case of an initial over-twisting, and viceversa it will decrease in case of under-twisting. However, the DNA twist variation will compensate the writhe one, thus keeping the linking number constant and equal to $Lk = n + n_{extra}$.

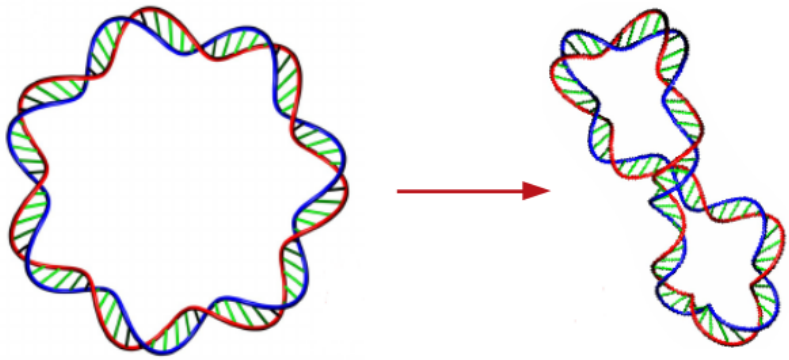

Figure 3.2: On the left: an under-wound DNA duplex which has just been circularized ($n_{extra} = -1$). The linking number Lk initially corresponds to the twist Tw ($Lk = Tw = n - 1$). During its evolution, the DNA duplex will release a bit of torsional strain by increasing its bending energy. In particular, on the right, it reaches a "figure-of-eight" arrangement. Since the two double strand portions are extremely close in correspondence of the crossing, one may estimate the following values for the final configuration: $Wr \approx -1$, $Tw \approx n$ such that Lk is exactly $n - 1$. Image adapted from [32].

Another way to compute the linking number Lk is the modified projection method. The procedure to obtain the link diagram is similar to the one described in Section 3.1.2 for the writhe, except for the following differences:

1. the link diagram contains 2 distinct curves;
2. only the crossings which involve two different curves require a sign. The crossings resulting from the intersection of a curve with itself must not be taken into account for the Lk determination.

The algebraic sum of the signed crossings will give twice the linking number value. An interesting point is that the Lk value obtained from this procedure does not depend on the particular plane

chosen for the projection. Thanks to this peculiarity, the modified projection method is a powerful tool to numerically compute the Lk.

The link diagram representation makes immediately understandable some more linking number properties:

1. two chiral enantiomers have opposite Lk;
2. every time that a curve orientation is reversed, the linking number changes its sign.
3. Whenever two curves are not entangled, their Lk is 0. However, this is not a necessary condition: there exist also intertwined curves whose $Lk = 0$.

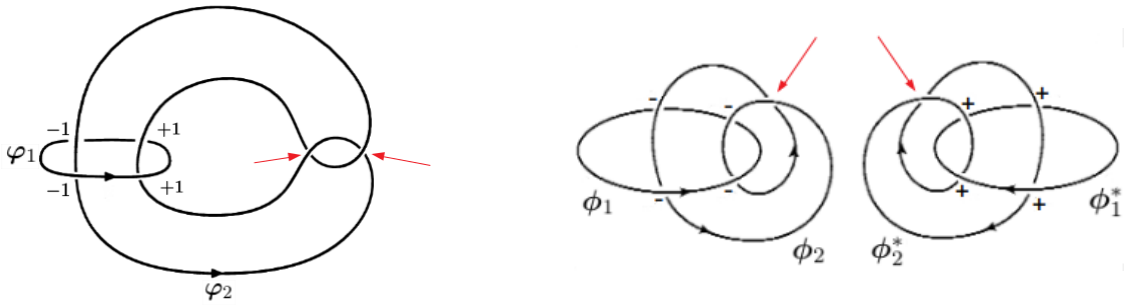

Figure 3.3: On the left: the Whitehead link is an example of intertwined curves whose $Lk = 0$. The crossings signs has been assigned according to the usual right hand rule. On the right: two chiral enantiomers with opposite linking number ± 2. In both images the crossings of a curve with itself (irrelevant for Lk calculation) are indicated with red arrows. Images adapted from [4, 9].

3.2.2 Superhelical density

The descriptors previously introduced can be used to quantify the supercoiling level of DNA. As already mentioned in Chapter 2, the DNA double helix makes a full turn around its own axis every 10.4 base pairs when it is torsionally relaxed. Moreover, for closed chains, the configuration which minimizes also the bending energy is the circular planar one. Such energetically relaxed configuration could be taken as reference for the supercoiling characterization, and its linking number is denoted as Lk_0 (if the molecule length is an integer multiple of 10.4). Given another arbitrary unknotted configuration with linking number Lk, we define the quantity

$$\Delta Lk = Lk - Lk_0 = Lk - Tw_0$$

and the so-called superhelical density

$$\sigma = \frac{\Delta Lk}{Lk_0}$$

For DNA modeling, one usually adopts the convention for which the two strands are oriented in a parallel fashion. In this way, the linking number Lk is always positive, and the supercoiling sign

corresponds to that of ΔLk. In particular, coming back to the example of the previous subsection, under-twisting ($n_{extra} < 0$) will produce negative supercoiling $\sigma = -\frac{|n_{extra}|}{n}$ (as in Figure 3.2, where $\sigma = -\frac{1}{n}$). Viceversa, positive supercoiling $\sigma = \frac{n_{extra}}{n}$ will result from over-twisting ($n_{extra} > 0$).

3.3 Supercoiled knotted configurations

The examples presented so far involved DNA duplex configurations with an helical axis topologically equivalent to the unknot 0_1, i.e. that can be flattened into a planar ring. In this case, the reference Lk_0 refers to a planar configuration with $Wr_0 = 0$, so that $Lk_0 = Tw_0$. The supercoiling level evaluation for knotted conformations requires more care. Before dealing with the proper supercoiling quantification, the ideal knot concept should be introduced.

3.3.1 Ideal knots

The ideal knot is a particular geometrical realization of a knot species which can be obtained by replacing the knot curve with a cylindrical tube, and increasing its thickness while keeping the axial length constant. The axial shape will be deformed in order to allow the maximal uniform inflation of the cylindrical tube without making the surface singular, or self-intersecting. In this way, one gets the knotted tube configuration which maximizes the volume-to-surface ratio. Moreover, the ideal shape does not depend on the spatial scale, and ideal knots of different axial length share the same axial length to tube diameter ratio.

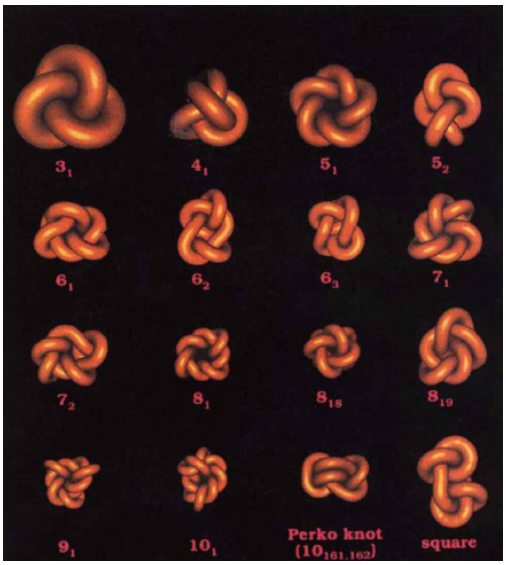

TABLE Characteristics of ideal geometrical representations of knots

Knot type	Axis length/ diameter	Writhe	Average crossing number
link 2_1	6.3	0.00	0.0
3_1	16.4	3.41	4.4
4_1	21.2	0.00	6.6
5_1	24.2	6.26	7.9
5_2	24.9	4.54	8.4
6_1	29.3	1.23	10.3
6_2	29.2	2.70	10.2
6_3	30.5	0.16	10.7
7_1	30.9	9.15	11.5
7_2	33.2	5.82	12.1
8_1	37.0	2.33	14.0
8_{18}	38.3	0.02	15.4
8_{19}	31.0	8.64	11.2
8_{20}	32.7	2.00	12.0
8_{21}	33.9	4.62	12.8
9_1	38.3	12.07	15.2
9_2	40.0	6.84	16.2
10_1	44.8	3.47	18.4
11_1	47.0	14.85	19.1
$3_1\#3_1$ (granny)	29.1	6.81	10.3
$3_1\#3_1^*$ (square)	28.9	0.01	9.6
Perko knot ($10_{161,162}$)	37.6	9.45	15.3

Figure 3.4: On the left: representation of an ideal knots selection. All the knots are characterized by the same axial length and it is immediate to notice that their compaction decreases with the knot complexity. These ideal configurations have been obtained by means of Monte Carlo simulations [13]. On the right: a table summarizing some of the ideal conformations main properties, like axis length to tube diameter ratio, writhe and ACN. Image and table adapted from [13].

The writhe of an ideal configuration, sometimes called intrinsic writhe of the knot, is a crucial quantity to be taken into account for the supercoiling quantification of a knotted DNA duplex, as better explained in the next subsection. Notice from Figure 3.4 that almost all knot species have a non zero intrinsic writhe due to the presence of topologically intrinsic crossings in all the projections.

3.3.2 Supercoiling evaluation

A way to assign the Lk excess due to supercoiling to an arbitrary closed double stranded DNA configuration is explained in what follows. This quantity does not change during the dynamical evolution: indeed, due to self-avoidance, the molecule topology is preserved.
Consider the case of a torsionally relaxed chain, which can be modeled with a WLC (worm-like-chain) Hamiltonian with excluded volume. The torsionally relaxed ensemble is characterized by an average writhe equal to that of the ideal configuration with the corresponding topology. For this reason, to evaluate the supercoiling level of an arbitrary knot species (included the unknot) it is reasonable to consider the torsionally relaxed chain with ideal geometry as a reference configuration. As already mentioned, being the 0_1 ideal writhe $Wr_0 = 0$, for arbitrary unknotted geometries one has

$$\Delta Lk = Lk - Tw_0 = Lk - \frac{L}{p} \tag{3.1}$$

with L the DNA length and p the DNA pitch. Notice that for a planar ring whose length is not multiple of 10.4 base pairs, it is necessary to add a deficit/excess of twist to close the molecule, since the Lk must be an integer. For non trivial knots, the intrinsic writhe (see Figure 3.4) comes into play as described in the following example.

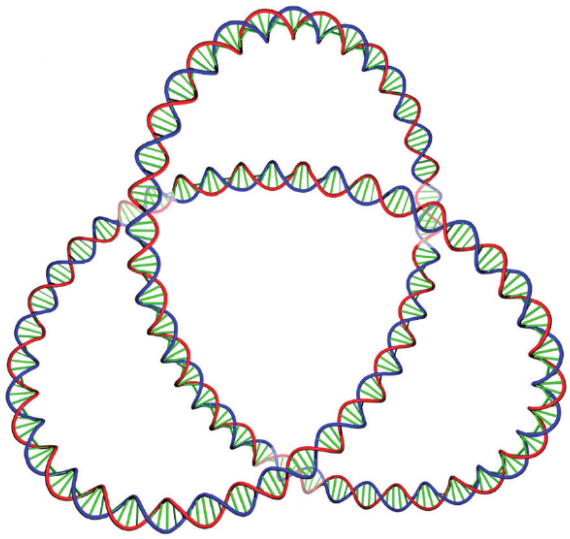

Figure 3.5: A DNA duplex whose helical axis geometry is the one of an ideal left-handed trefoil 3_1^L. Image adapted from [33].

Consider a DNA double strand whose helical axis assumes an arbitrary geometry topologically

equivalent to 3_1^L trefoil knot and whose axial length is n times the DNA pitch. Its supercoiling level can be quantified with the effective ΔLk (ΔLk_e) [5], which is the difference between the actual molecule Lk and $(Tw_0 + Wr_0)$ of the reference configuration:

$$\Delta Lk_e = Lk - Tw_0 - Wr_0 = \Delta Lk - Wr_0$$

The corresponding superhelical density will be:

$$\sigma = \frac{\Delta Lk_e}{Tw_0}$$

Notice that we did not call $Tw_0 + Wr_0$ of the reference configuration a proper linking number since it may be a non integer value (in particular, in our example $Tw_0 + Wr_0 = n - 3.41$). This means that a DNA duplex of length $L = np$, with an ideal left-handed trefoil geometry and minimum torsional energy cannot be exactly torsionally relaxed, and a deficit/excess of supercoiling is necessary to close the molecule. However, one may find in principle an appropriate length such that the ideal knotted configuration with the minimum twisting energy is perfectly torsionally relaxed. Such a length varies with the knot species because of their different intrinsic writhe Wr_0. So, the possibility to have an ideal torsionally relaxed DNA duplex depends on the interplay between intrinsic writhe of the correspondent topology and axial length.

3.4 Topoisomerases action

With the formalism described in the previous sections the topoisomerases action can be better quantified. The swiveling mechanism induced by type IB topoisomerases corresponds to take a point $\boldsymbol{\phi}(t_0)$ on the curve representing one backbone, imagine to cut the curve in that point and let it rotate around the axis $\boldsymbol{\alpha}$ before to reseal it in the same point $\boldsymbol{\phi}(t_0)$. This process results in a twist variation ΔTw equal to the number of turns occurred, and it leaves the writhe Wr unaltered since the helical axis $\boldsymbol{\alpha}$ does not undergo any appreciable deformation, so $\Delta Lk = \Delta Tw$.
The typeIA topoisomerases action, which releases torsional strain thanks to the strand passage mechanism, results in an increasing/decreasing of the helical turns number by 1 unit. Again, the molecule writhe is practically unchanged at the end of this process, and $\Delta Lk = \Delta Tw = \pm 1$.

Conversely, the cross inversion mechanism typical of type II topoisomerases changes the writhe by approximately 2 units ($\Delta Wr \approx \pm 2$), whereas the twist stays almost still ($\Delta Tw \approx 0$). This happens because after a double strand passage a positive crossing is replaced by a negative one or viceversa. The linking number variation is given by $\Delta Lk = \pm 2$.

Chapter 4

Supercoiled knotted DNA: recent experimental findings

4.1 Experimental assessment of DNA knots presence

As already said in the previous chapters, the activity of type II topoisomerases is not only able to modify the molecule supercoiling level, but also the topology of closed DNA chains. In particular, whenever a strand-passage involves two different molecules, they can be catenated (or decatenated, as often happens for newly replicated DNA chains). Instead, in case of intra-molecular DNA passages, a closed DNA molecule can be knotted/unknotted.

Only recently this knotting/unknotting activity of type-2A topoisomerases has been studied *in vivo* in eukaryotic organisms, particularly in yeast circular minichromosomes [28]. It turned out that a steady state fraction of DNA molecules is kept in knotted configurations. The role of topoisomerase IIA and its regular activity is thought to be a crucial element in the preservation of such a steady state fraction. This belief is corroborated by some previous experiments with bacterial plasmids from *Escherichia coli* strains carrying gyrase gene mutations, which showed an increased production of knotted chains [23, 12].

Electrophoretic characterization A common experimental method to separate DNA molecules on the basis of their size is the gel electrophoresis. DNA samples are loaded in some wells placed at one extremity of a gel (usually agarose) and an electric field is turned on. Since DNA molecules are negatively charged, they start to migrate along the gel (passing through its pores) with a velocity correlated to their size: in particular, smaller DNA molecules will move faster. At the end of this procedure, the DNA chains will be divided in bands. The addition of a DNA-binding dye allows the visualization of such bands when they are exposed to UV radiation.

4.2 Knots complexity and migration velocity

The electrophoresis technique can be used to assess the presence of knots in circular DNA molecules. Indeed, knots are expected to make the DNA chains more compact, favoring their passage through the agarose gel pores and so increasing their migration velocity. Electrophoretic characterizations

of knotted circular DNA (prepared *in vitro*) have been done for bacterial plasmids of length 4.4 kbp after their reaction with an excess of topoisomerase II [28]. However, in order to get a merely topological information, one has to get rid of the effects introduced by the supercoiling: for that reason all the samples have been also reacted with nicking endonuclease. In this way the molecules can dissipate all the torsional energy, thanks to a rotation of the broken strand around the intact one.

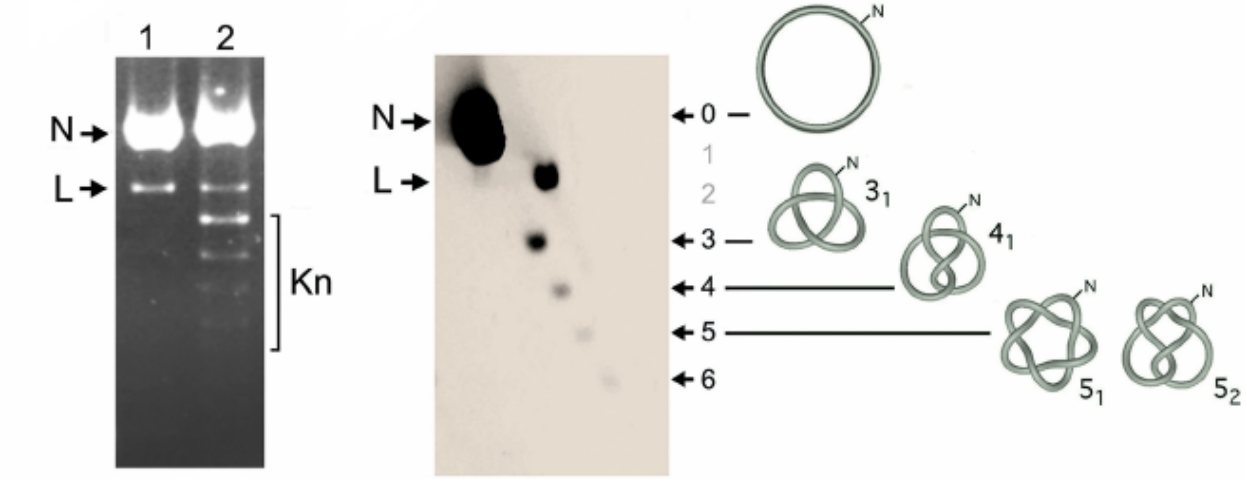

Figure 4.1: On the left: one-dimensional gel electrophoresis with nicked bacterial plasmids of 4.4 kbp. Before the reaction with a molar excess of topo II, only unknotted (N) and linear (L) molecules are present (lane 1). After the reaction (lane 2), a knots ladder (Kn) appears. As can be evinced from the bands thickness and brightness, the population of a certain knot type decreases with its complexity. Moreover, since a low voltage is applied, the distance in the gel is proportional to the knot complexity and compaction. On the right: high resolution two-dimensional electrophoresis on the same samples. A higher voltage is applied in the left-to-right direction in order to retard the knot species compared to the linear fragments. The crossing number of each knot type is reported on the right. Image adapted from [28].

The application of a low voltage in the top-to-bottom direction causes the desired separation, and a knots ladder appears in the gel (see the left part of Figure 4.1). This suggests that the migration velocity of a knot species depends on its topological complexity. In particular, as the average crossing number of a knot species increases, it will move faster along the gel. This behavior has been further confirmed by numerical Monte Carlo simulations that show a linear relation between ACN and migration velocity [24].

However, with a simple one dimensional electrophoresis, the knot identification might be a bit difficult for samples made by DNA mixtures containing also linear DNA fragments. In this case, the band corresponding to the linear molecules may overlap with those of the various knot species. To solve this ambiguity, a higher voltage can be applied in the horizontal direction to help the knots bands identification (see the right part of Figure 4.1 representing the two-dimensional

electrophoresis result). This strategy is based on the fact that when knotted molecules undergo a higher electric field, they can migrate more slowly than the linear ones, despite their greater complexity and compaction [26] (see Figure 4.2). Probably, this is due to the linear fragments higher flexibility and capability to adapt their shape to pass through the gel pores, which favor reptation effects and consequently the molecules migration.

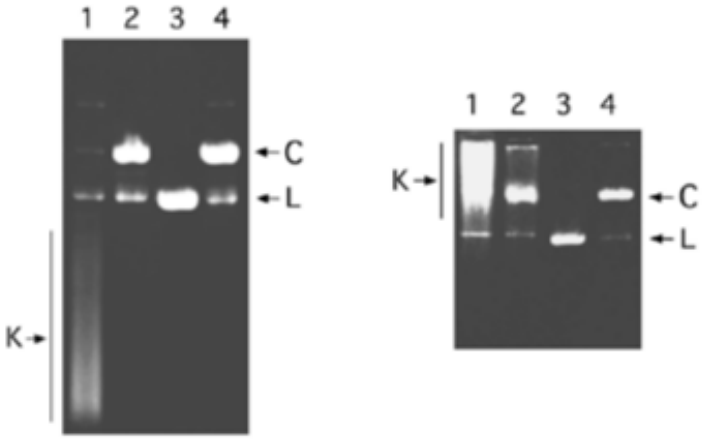

Figure 4.2: Effect of the applied electric field on the electrophoretic migration of DNA samples obtained by bacteriophage P4 capsid. Lane 1: sample with no additional treatment after the extraction. Lane 2: sample heated to 75°C. Lane 3: sample treated with nicking endonuclease. Lane 4: sample treated with topoisomerases II. The one-dimensional electrophoresis on the left has been performed at low voltage (25V for 20h), whereas the one on the right at high voltage (100V for 1h). From a comparison of the two cases it emerges that, despite their greater compaction, knotted configurations can migrate more slowly than linear fragments when the applied voltage is high. Image adapted from [26].

4.3 Knots occurence in eukaryotic chromatin

Various circular minichromosomes of different length present in yeast cells have been analyzed to evaluate their knotting probability (e.g. the 4.4 kbp YRp4) [28]. In order to avoid topological variation during their extraction and manipulations, the minichromosomes topology has been fixed *in vivo* before the harvest by inhibiting the topoisomerases action. A further treatment with nicking endonuclease allowed the DNA molecules to relax their supercoiling before to be loaded in the gel. A high resolution two-dimensional electrophoresis with low-high voltage showed a bands separation similar to that of the bacterial plasmids characterization described in the previous subsection. To confirm that the bands ladder was populated by knots of increasing complexity, the same sample has been treated separately with topo I and topo II and the same experiment has been conducted a second time (see Figure 4.3). In the first case, no appreciable differences emerged. Conversely, the reaction with topo II simplified the knot topology and most of the original bands disappeared from the gel.

An interesting aspect that has been examined is the knotting probability scaling with the chromatin length. Numerical simulations and knots occurrence in bacterial plasmids due to the topo II activity *in vitro* suggested that the knotting probability should increase with the DNA molecule length. On the other hand, the curve obtained from the electrophoretic experiments with yeast circular minichromosomes shows a saturation of the knotting probability to a value between 0.02 and 0.03 (see Figure 4.3). Tentative explanations for this phenomenon are related to variations in topo II unknotting efficiency and/or in chromatin packaging as the nucleosomal fibers length increases.

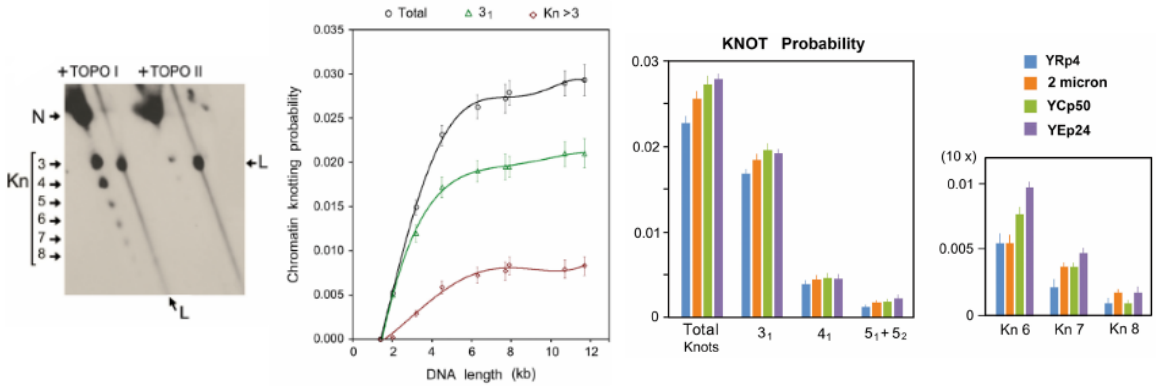

Figure 4.3: On the left: high resolution two-dimensional electrophoresis at low-high voltage with yeast minichromosomes YRp4 (4.4 kbp) after their reaction with topo I (lane 1) and topo II (lane 2). The bands disappearance in the lane 2 confirmed that the ladder was populated by knot species (labeled with their crossing number). Central plot: knotting probability vs DNA length profiles relative to trefoils (green), more complex knots with crossing number > 3 (red) and their sum (black). As the chromatin length increases, the curves saturate to a stationary value. On the right: knot probability for various knot species. A colors legend is reported to distinguish the minichromosomes types. Image adapted from [28].

4.4 DNA knots chirality

As described in the previous section, nicked DNA molecules with different topology can be separated during gel electrophoresis experiments since their compaction is strictly related to their topological complexity. However, the same method is not suitable to distinguish chiral enantiomers of the same knot type. Indeed, once nicked, they become mirror images and thus they exhibit the same compaction.

Chiral enantiomers can be experimentally recognized by means of two main techniques. The first one uses electron microscopy to visualize the crossings sign, a delicate procedure which is often not viable because it requires the sample purification and coating with proteins [15]. Alternatively, Shaw and Wang [22] found that, for trefoil knots, a coupling between knot handedness and supercoiling can result in a slightly different gel migration velocity for the positive- and negative-noded forms. Also this analysis needs the nicked molecules to be purified before the supercoiling addi-

tion and resealing, making the processing of samples with a reduced knotted fraction more difficult.

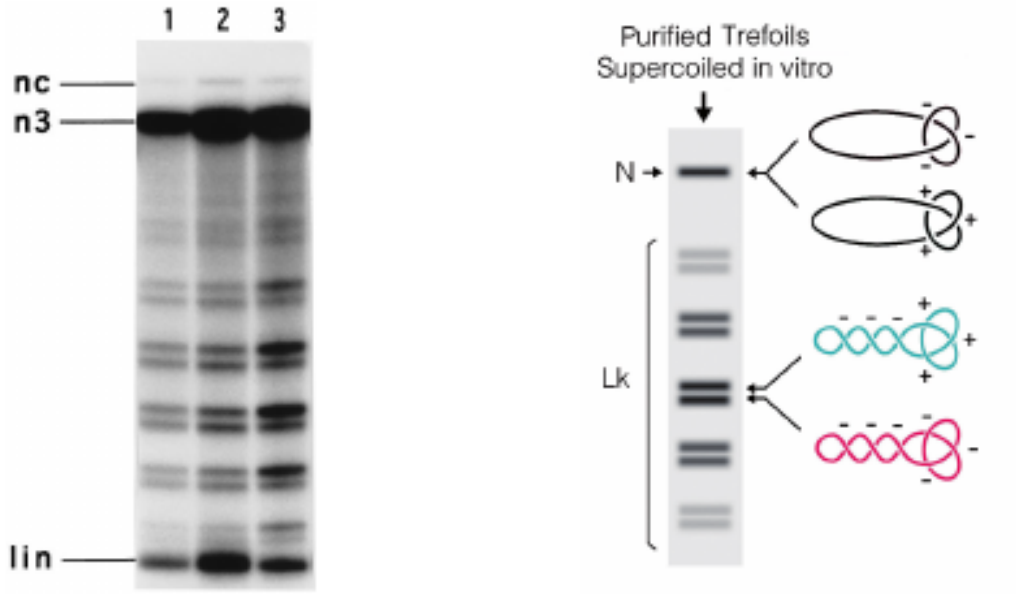

Figure 4.4: Trefoils doublets separation with one-dimensional gel electrophoresis. On the left: knotted supercoiled configuration of a 5.6 kbp DNA plasmid obtained by cyclization in diverse reaction conditions (lanes 1 and 2) or after a treatment with molar excess of topo II (lane 3). Image from [22]. On the right: trefoil enantiomers separation obtained with YRp4 plasmid (4.4 kbp). Image from [27].

More recently, Roca et al. [27] improved the method proposed by Shaw and Wang performing high resolution two-dimensional electrophoresis on supercoiled molecules. The top to bottom direction is used to separate the various species according to their supercoiling level. Although at this stage the two enantiomers are already resolved into doublets, they are masked by other DNA molecules presents in the mixture (e.g. supercoiled unknotted circles). To tackle this problem, an appropriate concentration of chloroquine can been added in the left to right direction. The intercalator positioning between base pairs induces a double helix unwinding (twist reduction) that is compensated by a writhe increasing. In this way, the molecular compaction effect induced by negative supercoiling can be reduced, but not that due to the knotting. Thus, after the chloroquine addition, unknots and trefoils change their migration velocity to different extents, and the enantiomers doublets can be properly isolated.

Chapter 5

Methodology and Statistical Physics model

5.1 DNA Coarse Grained Model

The purpose of this section is to present the coarse grained model we implemented to study the systems involved in the phenomenology presented in Chapter 4. In particular, we aimed to describe 4.4 kbp circular dsDNA molecules (for a better comparison with the experimental results in [28, 27]). The DNA filament has been treated as an elastic rod, modeled with a chain of $N = 150$ beads connected by cylindrical segments. Beads positions will be denoted with the vectors $\{\boldsymbol{r}_i\}_{i=1...N}$ and the segments between them with $\{\boldsymbol{b}_i = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i\}_{i=1...N}$. Notice that in order to fulfill the closed chain boundary conditions, we need to enforce $\boldsymbol{r}_{N+1} \equiv \boldsymbol{r}_1$. With the coarse graining level we chose, each segment is supposed to represent the equivalent of ≈ 30 nucleotides. Moreover, since no information regarding the nitrogenous basis has been added, the model is not able to capture any sequence dependent effect.

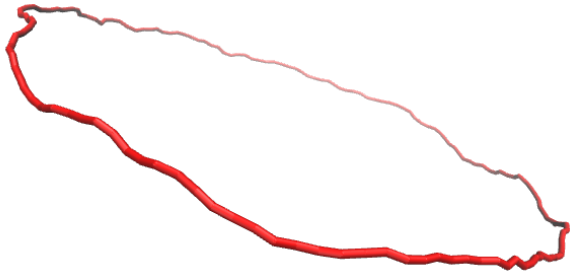

Figure 5.1: Closed DNA molecule modeled as an elastic rod made by $N = 150$ cylindrical segments.

To simulate the system mechanical properties, the following contributions to the Hamiltonian have been taken into account: stretching energy H_s, bending energy H_b, excluded volume interactions H_{ev}, torsional energy H_t. Thus, the total Hamiltonian of our model reads:

$$H_{tot} = H_s + H_b + H_{ev} + H_t \tag{5.1}$$

The expression (as reduced Hamiltonian βH that directly appears in the Boltzmann weight) and description of each term are given in the following subsections.

5.1.1 Stretching

With the stretching term H_s we regulated the extent to which each segment was able to change its length when subject to thermal fluctuations. Being the distance between consecutive base pairs 0.34 nm, the segment length at rest is given by $0.34 \cdot \frac{4400}{150} \approx 10$ nm. We denote this length with b_0. In order to prevent too pronounced contour length and segments variations, a FENE potential with both quadratic and quartic terms has been used. If we indicate with

$$\rho_i = \frac{||\boldsymbol{r}_{i+1} - \boldsymbol{r}_i|| - b_0}{\Delta b_0} = \frac{||\boldsymbol{b}_i|| - b_0}{\Delta b_0} \tag{5.2}$$

the relative bond fluctuation, our stretching energy H_s is described by the following expression:

$$\beta H_s(\{\boldsymbol{r}_i\}_{i=1\ldots N}) = \kappa_s \sum_{i=1}^{N} (\rho_i^2 + \rho_i^4) \tag{5.3}$$

with κ_s the stretching stiffness, $\beta = \frac{1}{K_b T}$ the inverse temperature and Δb_0 the bond fluctuation parameter. We looked at [6] to set these phenomenological parameters, and we finally assigned the values $\kappa_s = 10$ and $\Delta b_0 = 0.1 b_0$.

5.1.2 Bending

A bending potential has been introduced to take into account the semi-flexible behavior of DNA. For each couple of consecutive segments i and $i+1$, the bending energy depends on the angle θ_i between $\boldsymbol{b}_i$ and $\boldsymbol{b}_{i+1}$. If we denote with $\boldsymbol{u}_i$ (called *bond* in the following) the normalized versor along $\boldsymbol{b}_i$, the bending energy reads

$$\beta H_b(\{\boldsymbol{r}_i\}_{i=1\ldots N}) = -\kappa_b \sum_{i=1}^{N} \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} = -\kappa_b \sum_{i=1}^{N} \cos(\theta_i) \tag{5.4}$$

with κ_b the bending stiffness. For semi-flexible chains the angles θ_i are usually quite small, and expanding $\cos(\theta_i)$ to second order is a good approximation. In that case, one obtains the harmonic potential

$$\beta H_b(\{\boldsymbol{r}_i\}_{i=1\ldots N}) = -\kappa_b \sum_{i=1}^{N} (1 - \frac{\theta_i^2}{2}) = -\kappa_b N + \frac{\kappa_b}{2} \sum_{i=1}^{N} \theta_i^2 \tag{5.5}$$

with the constant term usually omitted. The bending stiffness parameter has been set to $\kappa_b = 5$ in such away to reproduce the DNA bending persistence length (which is estimated to be around ≈ 50 nm at 0.1 M salt concentration). The relation between bending persistence length and stiffness parameter is given in Appendix B.

5.1.3 Excluded volume

To avoid configurations in which two portions of the chain overlap with each other, we added an excluded volume contribution H_{ev} to the Hamiltonian, assigning infinite energy to all the forbidden configurations and zero energy to the admitted ones. Explicitly,

$$\beta H_{ev}(\{\boldsymbol{r}_i\}_{i=1\dots N}) = \begin{cases} 0 & \text{if } \sum_{j=3}^{N-1} \mathbb{I}[d(\boldsymbol{b}_1, \boldsymbol{b}_j) > \sigma] + \sum_{i=2}^{N-2} \sum_{j=i+2}^{N} \mathbb{I}[d(\boldsymbol{b}_i, \boldsymbol{b}_j) > \sigma] = \frac{(N-3)N}{2} \\ +\infty & \text{if } \sum_{j=3}^{N-1} \mathbb{I}[d(\boldsymbol{b}_1, \boldsymbol{b}_j) > \sigma] + \sum_{i=2}^{N-2} \sum_{j=i+2}^{N} \mathbb{I}[d(\boldsymbol{b}_i, \boldsymbol{b}_j) > \sigma] < \frac{(N-3)N}{2} \end{cases} \tag{5.6}$$

where $d(\boldsymbol{b}_i, \boldsymbol{b}_j)$ is the distance between the segments i and j, σ is the DNA duplex thickness and $\mathbb{I}[\cdot]$ the indicator function that returns 1 if the condition is true and 0 otherwise.

5.1.4 Torsion

Taking into account the torsion is a bit more delicate and requires the introduction of new degrees of freedom. As already done by Allison et al.[2], Chirico and Langowski [6], we associated to each bead i a body fixed coordinate system (bfc) $\Sigma_i = \{\boldsymbol{f}_i, \boldsymbol{v}_i, \boldsymbol{u}_i\}$. The vectors $\{\boldsymbol{u}_i\}_{i=1\dots N}$ are the bonds previously described, whereas $\{\boldsymbol{f}_i\}_{i=1\dots N}$ and $\{\boldsymbol{v}_i\}_{i=1\dots N}$ are perpendicular unit vectors called respectively *normals* and *binormals*. In this way, a molecular configuration will involve $9N$ variables: $3N$ are given by $\{\boldsymbol{r}_i\}_{i=1\dots N}$ and $6N$ by $\{\boldsymbol{f}_i, \boldsymbol{v}_i, \boldsymbol{u}_i\}_{i=1\dots N}$ since they are all unit vectors. At the same time, they must satisfy the following constraint equations

$$\begin{cases} \boldsymbol{u}_i = \frac{\boldsymbol{r}_{i+1} - \boldsymbol{r}_i}{||\boldsymbol{r}_{i+1} - \boldsymbol{r}_i||} & 2N \text{ constraints} \\ \boldsymbol{v}_i = \boldsymbol{u}_i \wedge \boldsymbol{f}_i & 2N \text{ constraints} \\ \boldsymbol{f}_i \cdot \boldsymbol{u}_i = 0 & N \text{ constraints} \end{cases} \tag{5.7}$$

so that the number of degrees of freedom reduces to $9N - 5N = 4N$. The new N d.o.f. are related to the twisting of each body fixed coordinate system around its bond $\boldsymbol{u}_i$, and will be described in the following by the set of angles $\{\phi_i\}_{i=1\dots N}$.

To analyze the different orientation of two consecutive coordinate systems we considered the Euler transformations defined by the three angles $(\alpha_i, \beta_i, \gamma_i)$ (see Figure 5.2). In particular, given the coordinate system Σ_i written as a matrix whose columns are the vectors $\{\boldsymbol{f}_i, \boldsymbol{v}_i, \boldsymbol{u}_i\}$ represented in the canonical basis $\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$ of the laboratory reference frame, we have

$$\Sigma_{i+1} = \Sigma_i \cdot \boldsymbol{E}(\alpha_i, \beta_i, \gamma_i) \tag{5.8}$$

with $\boldsymbol{E}(\alpha_i, \beta_i, \gamma_i)$ the matrix representation of the Euler transformation

$$\boldsymbol{E}(\alpha_i, \beta_i, \gamma_i) = \begin{vmatrix} c(\alpha_i)c(\beta_i)c(\gamma_i) - s(\alpha_i)s(\gamma_i) & -c(\alpha_i)c(\beta_i)s(\gamma_i) - s(\alpha_i)c(\gamma_i) & c(\alpha_i)s(\beta_i) \\ s(\alpha_i)c(\beta_i)c(\gamma_i) + c(\alpha_i)s(\gamma_i) & -s(\alpha_i)c(\beta_i)s(\gamma_i) + c(\alpha_i)c(\gamma_i) & s(\alpha_i)s(\beta_i) \\ -s(\beta_i)c(\gamma_i) & s(\gamma_i)s(\beta_i) & c(\beta_i) \end{vmatrix} \tag{5.9}$$

where $c(\cdot)$ and $s(\cdot)$ are just a shortcut notation for $\cos(\cdot)$ and $\sin(\cdot)$ functions. By looking at the Figure 5.2 one can readily see that the Euler angle β_i is nothing else than the bending angle that we called θ_i in the subsection 5.1.2. On the other hand, the angle $\alpha_i + \gamma_i$ quantifies how much

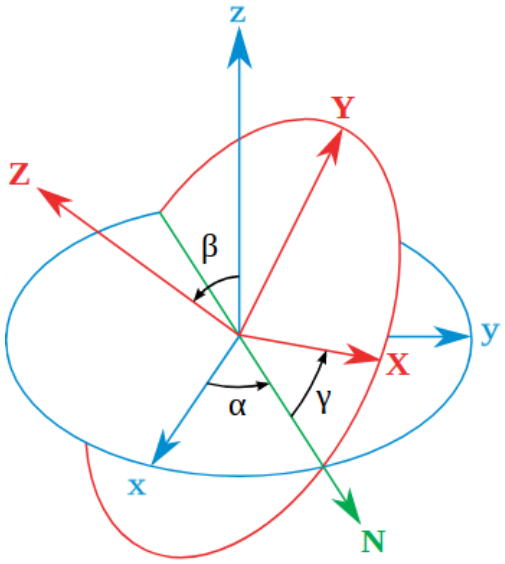

Figure 5.2: Euler transformation bringing the coordinate system xyz into coincidence with XYZ. The transformation can be decomposed into three elementary steps. First, xyz is rotated by α around the z axis. Then, the new coordinate system is rotated by β around $N = z \wedge Z$, the common perpendicular to z and Z. Finally, the obtained system is rotated by γ around Z. See 'Other sources' for the image reference.

the coordinate system Σ_i has been twisted around its own bond $\boldsymbol{u}_i$ before to coincide with Σ_{i+1}. Given a molecular configuration with known vectors $\{\boldsymbol{r}_i, \boldsymbol{f}_i, \boldsymbol{v}_i, \boldsymbol{u}_i\}_{i=1\dots N}$, such twisting angle can be computed from

$$\begin{cases} \cos(\alpha_i + \gamma_i)\cos^2(\frac{\beta_i}{2}) = \frac{1}{2}(\boldsymbol{f}_{i+1} \cdot \boldsymbol{f}_i + \boldsymbol{v}_{i+1} \cdot \boldsymbol{v}_i) \\ \sin(\alpha_i + \gamma_i)\cos^2(\frac{\beta_i}{2}) = \frac{1}{2}(\boldsymbol{f}_{i+1} \cdot \boldsymbol{v}_i - \boldsymbol{v}_{i+1} \cdot \boldsymbol{f}_i) \end{cases} \tag{5.10}$$

To conclude, we modeled the torsional energy contribution $H_t(\{\alpha_i, \gamma_i\}_{i=1\dots N})$ as that of an harmonic potential

$$\beta H_t(\{\alpha_i, \gamma_i\}_{i=1\dots N}) = \frac{\kappa_t}{2} \sum_{i=1}^{N} (\alpha_i + \gamma_i)^2 \tag{5.11}$$

with κ_t the torsional stiffness parameter. After a comparison with [6] we set $\kappa_t = 6$.

5.2 Geometric observables

To attempt a better quantification of the phenomenology presented in Chapter 4, we decided to assess the compaction and shape of both knotted and unknotted molecules, with and without supercoiling. To do that, we focus our attention on the geometrical observables described in what follows.

5.2.1 Gyration radius and tensor

The gyration radius is a rather common quantity in polymer physics, and it is supposed to give an idea of a polymer typical spatial size. It is defined as

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_{CM})^2 \tag{5.12}$$

where $\boldsymbol{r}_{CM}$ is the center of mass position (all beads are assumed to have unitary mass):

$$\boldsymbol{r}_{CM} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{r}_i \tag{5.13}$$

The gyration radius is proportional to the root mean square distance between two monomers, as shown by the alternative formula

$$R_g^2 = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_j)^2 \tag{5.14}$$

proved in the Appendix B. A more general quantity to describe the molecule spatial arrangement is the gyration tensor $\boldsymbol{S}$, defined as

$$S_{mn} = \frac{1}{N}\sum_{i=1}^{N}(r_{i,m} - r_{CM,m})(r_{i,n} - r_{CM,n}) \tag{5.15}$$

with $m, n \in \{x, y, z\}$ and $r_{i,m}$ the component along the m axis of $\boldsymbol{r}_i$. Since the gyration tensor $\boldsymbol{S}$ is real and symmetric, it can always be diagonalized

$$\boldsymbol{S} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^T \tag{5.16}$$

with R orthogonal and Λ the diagonal matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{pmatrix} \tag{5.17}$$

The columns of $\boldsymbol{R}$ provide the principal axis of the molecule, while the eigenvalues $\lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2$ are its principal moments. Moreover, since the tensor trace is invariant under change of basis, the eigenvalues sum gives the molecule gyration radius

$$R_g^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \tag{5.18}$$

5.2.2 Anisotropy and asphericity

To characterize polymer asphericity b and anisotropy κ we used shape measures proposed by [25]. In particular, both of them are defined as a function of the molecule principal moments. The asphericity

$$b = \lambda_1^2 - \frac{1}{2}(\lambda_2^2 + \lambda_3^2) \tag{5.19}$$

assumes low values whenever the beads spatial distribution is symmetric with respect to the principal axis. The relative shape anisotropy, instead, is bounded between $[0, 1]$ and is defined as

$$\kappa = \frac{3}{2}\frac{\lambda_1^4 + \lambda_2^4 + \lambda_3^4}{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^2} - \frac{1}{2} \tag{5.20}$$

5.3 Monte Carlo Sampling

The configurational space of our DNA closed chain at canonical equilibrium has been sampled through Monte Carlo simulations. Once initialized in a starting configuration γ_0, the system undergoes a state transition at each time step, defining a stochastic markovian process. The transition probability from a state $\gamma_i \in \Gamma$ to a state $\gamma_j \in \Gamma$ is given by the matrix element $\Pi_{\gamma_i \to \gamma_j}$, with Γ the set of all possible system configurations. Let us define with

$$\boldsymbol{p}^{(t)} = \left.\begin{pmatrix} p(\gamma_1, t) \\ p(\gamma_2, t) \\ \cdot \\ \cdot \\ \cdot \\ p(\gamma_{|\Gamma|}, t) \end{pmatrix}\right\} |\Gamma| \text{ elements} \tag{5.21}$$

the vectors whose i^{th} element is the probability to find the system in γ_i at time t and with

$$\Pi^{|\Gamma|\times|\Gamma|} = \begin{pmatrix} \Pi_{\gamma_1\to\gamma_1} & \Pi_{\gamma_2\to\gamma_1} & . & . & . & \Pi_{\gamma_{|\Gamma|}\to\gamma_1} \\ \Pi_{\gamma_1\to\gamma_2} & \Pi_{\gamma_2\to\gamma_2} & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ \Pi_{\gamma_1\to\gamma_{|\Gamma|}} & \cdot & . & . & . & \Pi_{\gamma_{|\Gamma|}\to\gamma_{|\Gamma|}} \end{pmatrix} \tag{5.22}$$

the stochastic matrix (normalized along the columns) containing the transition probabilities. In this way, the probability vector evolution with time can be simply expressed in matrix form

$$\boldsymbol{p}^{t+1} = \Pi \cdot \boldsymbol{p}^t = \Pi^{t+1} \cdot \boldsymbol{p}^0 \tag{5.23}$$

From this equation, one can write

$$\begin{aligned} p(\gamma, t+1) &= \sum_{\alpha\in\Gamma} p(\alpha, t)\Pi_{\alpha\to\gamma} = \sum_{\alpha\neq\gamma} p(\alpha, t)\Pi_{\alpha\to\gamma} + p(\gamma, t)\Pi_{\gamma\to\gamma} = \\ &= \sum_{\alpha\neq\gamma} p(\alpha, t)\Pi_{\alpha\to\gamma} + p(\gamma, t) - \sum_{\alpha\neq\gamma} p(\gamma, t)\Pi_{\gamma\to\alpha} \end{aligned} \tag{5.24}$$

As the time increases, the state probabilities $p(\gamma, t)$ tend to reach a stationary value $p_{st}(\gamma)$ if the following balance condition holds

$$\sum_{\alpha\in\Gamma} p_{st}(\alpha)\Pi_{\alpha\to\gamma} = \sum_{\alpha\in\Gamma} p_{st}(\gamma)\Pi_{\gamma\to\alpha} \tag{5.25}$$

One way to fulfill this equation is to separately equate the probability fluxes of each couple of states involved in the sum, namely

$$p_{st}(\alpha)\Pi_{\alpha\to\gamma} = p_{st}(\gamma)\Pi_{\gamma\to\alpha} \quad \forall\alpha \tag{5.26}$$

This more restrictive condition, satisfied for system at equilibrium, is called detailed balance. Usually, the transition probability $\Pi_{\gamma\to\alpha}$ is expressed as the product of a proposal term $T_{\gamma\to\alpha}$ and an acceptance term $A_{\gamma\to\alpha}$. Being our target equilibrium distribution the Boltzmann one

$$p_{st}(\gamma) = \frac{e^{-\beta H(\gamma)}}{Z} \tag{5.27}$$

the Eq. 5.26 can be rewritten as:

$$\frac{A_{\alpha\to\gamma}}{A_{\gamma\to\alpha}} = \frac{T_{\gamma\to\alpha}}{T_{\alpha\to\gamma}} e^{-\beta(H(\gamma)-H(\alpha))} \tag{5.28}$$

A common choice to verify Eq. 5.28, and so the detailed balance condition, is the well known Metropolis-Hastings criterion. This corresponds to choose the acceptance probabilities in the following way

$$A_{\gamma\to\alpha} = \min\left(1, \frac{T_{\alpha\to\gamma}}{T_{\gamma\to\alpha}} e^{-\beta(H(\alpha)-H(\gamma))}\right) \tag{5.29}$$

In the particular case of symmetric proposals, the previous expression reduces to

$$A_{\gamma\to\alpha} = \min\left(1, e^{-\beta(H(\alpha)-H(\gamma))}\right) \tag{5.30}$$

From the algorithmic point of view, the sampling is performed by stochastically proposing a transition from the actual configuration γ_{old} to a trial configuration γ_{trial}, drawing a random number $x \in [0,1]$ and accepting γ_{trial} only if

$$x < \frac{T_{\gamma_{trial}\to\gamma_{old}}}{T_{\gamma_{old}\to\gamma_{trial}}} e^{-\beta(H(\gamma_{trial})-H(\gamma_{old}))} \tag{5.31}$$

Whenever this inequality is not satisfied, the trial configuration is discarded and the system stays in γ_{old}.

5.3.1 Type of moves

The trial configuration proposal has been divided in two steps: a local cartesian displacement $\delta\boldsymbol{r}_i$ of a randomly chosen bead i and the subsequent rotation $\delta\phi_i$ of its body fixed coordinate system around its bond. Both $\delta\boldsymbol{r}_i$ and $\delta\phi_i$ are stochastically extracted from a uniform distribution with a suitably tuned support, and the way they affect the molecular configuration is described in the following. For clarity, they are treated separately.

Cartesian displacement $\delta \boldsymbol{r}_i$: Let us denote with $(\boldsymbol{r}_i, \boldsymbol{f}_i, \boldsymbol{v}_i, \boldsymbol{u}_i)$ the vectors relative to the i^{th} bead in the actual state, and with $(\boldsymbol{r}_i^{(1)}, \boldsymbol{f}_i^{(1)}, \boldsymbol{v}_i^{(1)}, \boldsymbol{u}_i^{(1)})$ the same vectors after the first step of the trial transition. In particular $\boldsymbol{r}_i^{(1)}$ and $\boldsymbol{u}_i^{(1)}$ can be readily computed

$$\begin{cases} \boldsymbol{r}_i^{(1)} = \boldsymbol{r}_i + \delta \boldsymbol{r}_i \\ \boldsymbol{u}_i^{(1)} = \frac{\boldsymbol{r}_{i+1} - \boldsymbol{r}_i^{(1)}}{||\boldsymbol{r}_{i+1} - \boldsymbol{r}_i^{(1)}||} \end{cases} \tag{5.32}$$

To update the remaining vectors, let us define the angle σ as

$$\cos(\sigma) = \boldsymbol{u}_i \cdot \boldsymbol{u}_i^{(1)} \tag{5.33}$$

and the unit vector $\hat{\boldsymbol{k}}$ perpendicular to both $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^{(1)}$:

$$\hat{\boldsymbol{k}} = \frac{\boldsymbol{u}_i \wedge \boldsymbol{u}_i^{(1)}}{||\boldsymbol{u}_i \wedge \boldsymbol{u}_i^{(1)}||} \tag{5.34}$$

The vectors $\boldsymbol{f}_i^{(1)}$ and $\boldsymbol{v}_i^{(1)}$ can be obtained by rotating $\boldsymbol{f}_i$ and $\boldsymbol{v}_i$ about $\hat{\boldsymbol{k}}$ by the angle σ. To do so, one can build the matrix

$$\boldsymbol{K} = \begin{pmatrix} 0 & -k_z & k_y \\ k_z & 0 & -k_x \\ -k_y & k_x & 0 \end{pmatrix} \tag{5.35}$$

and apply the Rodrigues' rotation formula

$$\boldsymbol{f}_i^{(1)} = \big(\mathbb{I} + \sin(\sigma)\boldsymbol{K} + (1 - \cos(\sigma))\boldsymbol{K}^2\big)\boldsymbol{f}_i \tag{5.36}$$

The vector $\boldsymbol{v}_i^{(1)}$ can be computed either by applying the same transformation to $\boldsymbol{v}_i$ or, more simply, from the external product $\boldsymbol{v}_i^{(1)} = \boldsymbol{u}_i^{(1)} \wedge \boldsymbol{f}_i^{(1)}$. An equivalent formula to perform the same rotation is given by

$$\boldsymbol{f}_i^{(1)} = \cos(\sigma)\boldsymbol{f}_i + (1 - \cos(\sigma))(\hat{\boldsymbol{k}} \cdot \boldsymbol{f}_i)\hat{\boldsymbol{k}} + \sin(\sigma)(\hat{\boldsymbol{k}} \wedge \boldsymbol{f}_i) \tag{5.37}$$

Notice that even though we did not perform the rotation of $\Sigma_i^{(1)}$ around $\boldsymbol{u}_i^{(1)}$ yet, the i^{th} and $(i+1)^{th}$ twisting angles may already be changed at this stage.

Rotation $\delta \phi_i$: As anticipated, the second step involve a rotation of the i^{th} bfc about the bond $\boldsymbol{u}_i^{(1)}$ by a randomly extracted angle $\delta\phi_i$. The vectors corresponding to the final trial configuration, indicated with $(\boldsymbol{r}_i^{(2)}, \boldsymbol{f}_i^{(2)}, \boldsymbol{v}_i^{(2)}, \boldsymbol{u}_i^{(2)})$, can be computed in this way

$$\begin{cases} \boldsymbol{r}_i^{(2)} = \boldsymbol{r}_i^{(1)} \\ \boldsymbol{u}_i^{(2)} = \boldsymbol{u}_i^{(1)} \\ \boldsymbol{f}_i^{(2)} = \cos(\delta\phi_i)\boldsymbol{f}_i^{(1)} + (1 - \cos(\delta\phi_i))(\boldsymbol{u}_i^{(1)} \cdot \boldsymbol{f}_i^{(1)})\boldsymbol{u}_i^{(1)} + \sin(\delta\phi_i)(\boldsymbol{u}_i^{(1)} \wedge \boldsymbol{f}_i^{(1)}) \\ \boldsymbol{v}_i^{(2)} = \boldsymbol{u}_i^{(2)} \wedge \boldsymbol{f}_i^{(2)} \end{cases} \tag{5.38}$$

where the Rodrigues' rotation formula has been applied again to find $\boldsymbol{f}_i^{(2)}$. Notice that since the rotation axis $\hat{\boldsymbol{k}}$ and $\boldsymbol{u}_i^{(1)}$ are perpendicular to each other, the moves proposal is symmetric and the

Eq. 5.30 holds.
Alternatively to the cartesian displacement, one could implement the so called crankshaft move to vary the molecule geometry. It consists in randomly selecting two beads i and j and rotating the sub-chain in between about the axis determined by $\boldsymbol{r}_j - \boldsymbol{r}_i$ by a stochastic angle. Although this global approach would allow to propose significantly modified trial configurations, it may result in an excessively high rejection ratio.

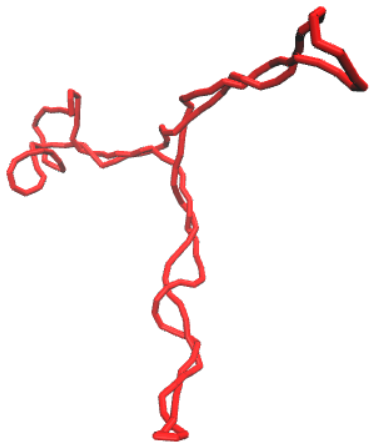

Figure 5.3: Supercoiled unknotted configuration with three clearly visible plectonemic structures.

The main reason that leads us to prefer local cartesian displacements is that the mentioned drawback would be enhanced by the fact that we deals with supercoiled molecules that exhibits plectonemes (see Figure 5.3). These highly wound superhelical structures are quite compact and crankshaft moves would be very likely rejected because of excluded volume interaction violations.

5.3.2 Autocorrelation time

In general, analysis involving Monte Carlo Markov Chains (MCMC) can help the expectation value computation of observables relevant to the problem under study. Given a certain observable f, its statistical average corresponds to

$$< f >_p = \int_\Gamma f(\gamma) p(\gamma)\, d\gamma \tag{5.39}$$

with Γ the system configurational space and $p(\gamma)$ the statistical weight of γ. If one was able to sample a sequence of independent configurations $\gamma_1, \gamma_2, ..., \gamma_N$ from $p(\gamma)$, the statistical average $< f >_p$ would be approximated by the estimator

$$\hat{f} = \frac{1}{N} \sum_{i=1}^{N} f(\gamma_i) \tag{5.40}$$

Indeed, the random variable $\hat{f}$ has average $< f >_p$ and variance that decreases as N gets larger. However, since the MCMC provides a sequence of correlated states, one needs to evaluate how many

steps are required to consider the random variables $f(\gamma_i)$ and $f(\gamma_{i+t})$ as practically independent. To address this problem, one can compute the autocorrelation function

$$C(t) = < (f_i - < f >_{p(\gamma_i)})(f_{i+t} - < f >_{p(\gamma_{i+t})}) >_{p(\gamma_i, \gamma_{i+t})} \tag{5.41}$$

which, at equilibrium, is actually independent of i due to the time translational invariance. The autocorrelation function asymptotic behavior for large t is usually given by an exponential decay

$$C(t) \overset{t \to \infty}{\sim} \exp\left(-\frac{t}{\tau}\right) \tag{5.42}$$

where τ is called the autocorrelation time. From an algorithmic point of view, a fist approximation of the autocorrelation function can be given by the estimator

$$\hat{C}_1(t) = \frac{1}{N-t} \sum_{i=1}^{N-t} (f_i - \hat{f})(f_{i+t} - \hat{f}) \tag{5.43}$$

This expression relies on the previously mentioned time translational invariance but it does not take into account the initial burn-in required by the MCMC to reach the equilibrium distribution. So, as a rule of thumb, one can discard a number of initial configurations comparable to the autocorrelation time τ_1 found with $\hat{C}_1(t)$, and compute again the estimator

$$\hat{C}(t) = \frac{1}{N-t-\tau_1} \sum_{i=\tau_1+1}^{N-t} (f_i - \hat{f})(f_{i+t} - \hat{f}) \tag{5.44}$$

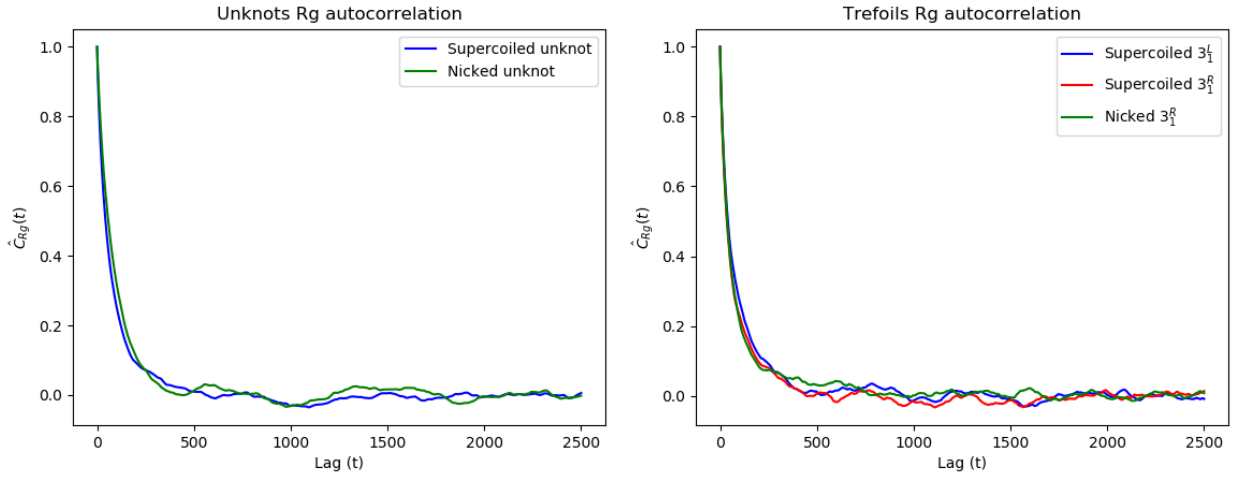

Figure 5.4: Gyration radius autocorrelation functions $\hat{C}_{Rg}(t)$ of knotted/unknotted supercoiled/nicked configurations. A unit in the lag time axis corresponds to 10^7 Monte Carlo steps. As the lag time increases, oscillations become wider due to the reduced number of elements in the sum 5.44.

5.3.3 Tuning the moves amplitude

To perform the cartesian displacement $\delta \boldsymbol{r_i}$ we uniformly extracted a random point from a cube centered in $\boldsymbol{r}_i$ with edges length A. The A magnitude has been established by considering the two following aspects:

1. If big A values would enlarge the configurational space region potentially reachable from a certain state γ_{old} with a single move, they could also considerably increase the rejection ratio because of the higher (on average) energy difference between γ_{old} and γ_{trial}. For this reason, too high A values would induce an increasing of the autocorrelation time. On the other hand, although generally accepted, too small variations between γ_{old} and γ_{trial} would also slow down the system decorrelation. Thus, the autocorrelation time minimum has been obtained with the compromise value $A = 0.1 \cdot b_0$.

2. Fortunately, the same moves amplitude that minimized the system autocorrelation was already suitable to prevent strand passages to occur. In this way, chain topology variations were disallowed and the Lk conservation was guaranteed.

5.4 Validation of the model

Tangent-tangent correlation for planar rings Let us consider a DNA planar ring confined in two dimensions characterized by a low flexibility ratio $\chi = \frac{L}{l_p}$. In these conditions, an analytical expression for the tangent-tangent correlation function $G(s)$ (see Appendix B for $G(s)$ definition) has been derived by Sakaue et al. [21]:

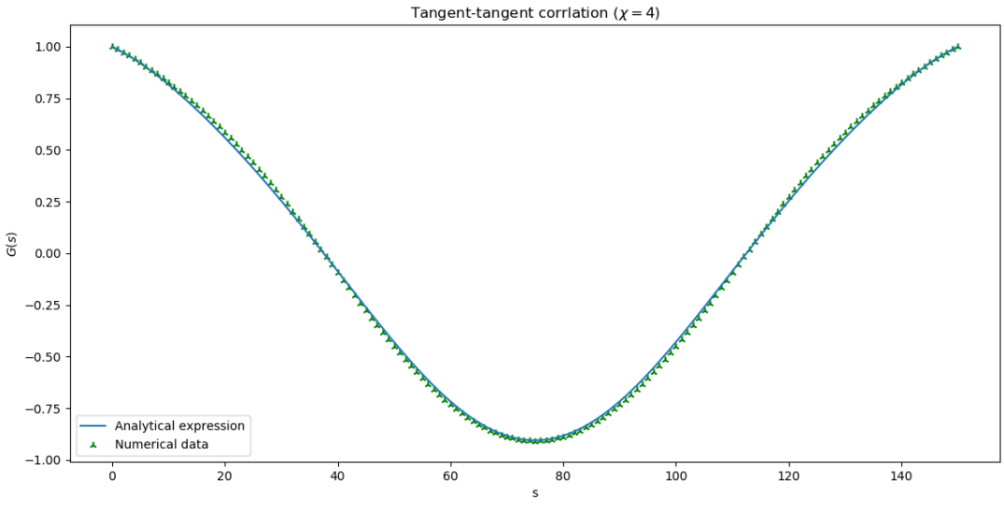

Figure 5.5: Comparison between the tangent-tangent correlation $G(s)$ obtained with our model (green marker) and the one predicted by the analytical expression 5.45 (blue curve), for a DNA two-dimensional closed ring with flexibility ratio $\chi = 4$. The curvilinear coordinate s is expressed in simulation units, obtained dividing every length measure by $b_0 = 10$ nm.

$$G(s) = \left[1 + \frac{L}{2\pi^2 l_p}\left(g\left(\frac{2\pi s}{L}\right) - g(0)\right)\right]\cos\left(\frac{2\pi s}{L}\right) \tag{5.45}$$

with

$$g(z) = \frac{(\pi - z)^2}{4} - \frac{\pi^2}{12} - \cos(z) \tag{5.46}$$

As one approaches the infinitely stiff regime $\chi = 0$, the function $G(s)$ tends to

$$G(s) \xrightarrow{\chi \to 0} \cos\left(\frac{2\pi s}{L}\right) \tag{5.47}$$

which is indeed the case of a perfectly circular molecule. Increasing the DNA flexibility, the expression has shown to be in excellent agreement with the numerical simulations and experiments up to $\chi \simeq 6$, but then it lose its precision. Notice that the $G(s)$ function no longer depends on the index i because of the translational invariance introduced by the curve closure.

To validate our model, we simulated a two-dimensional ring in stiff regime $\chi = 4$ with the torsional contribution turned off, and we compared the numerically obtained tangent-tangent correlation to the one predicted by Eq. 5.45 (see Figure 5.5).

Torsional stiffness in planar rings To further validate our model and methods we simulated a rigid circular planar ring at canonical equilibrium with only twisting energy, and we computed its $\boldsymbol{f}$-vectors correlation

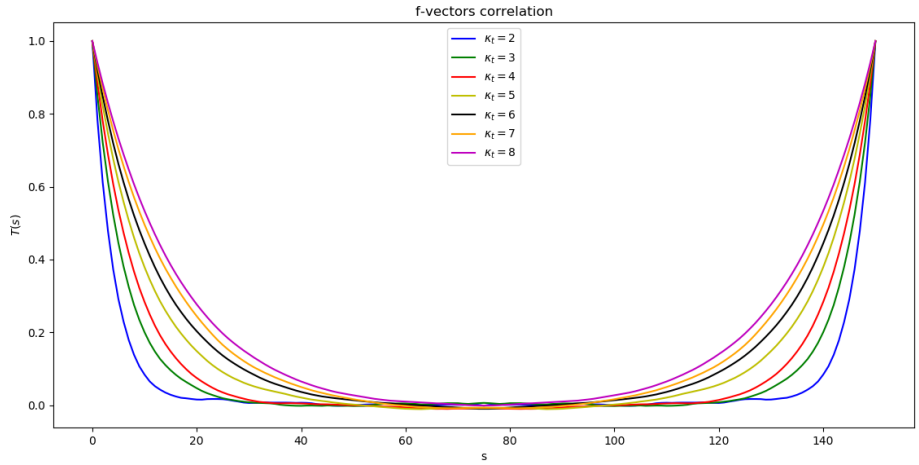

Figure 5.6: $\boldsymbol{f}$-vectors correlation for different values of the torsional stiffness parameter κ_t. The characteristic length of the initial decay is related to the torsional persistence length. The curvilinear coordinate is expressed in simulation unit.

$$T_i(s) = < \boldsymbol{f}_i \cdot \boldsymbol{f}_{i+s} > \tag{5.48}$$

For moderate values of torsional stiffness, the ring local behavior (in terms of $\boldsymbol{f}$-vectors orientation memory) cannot differ too much from that of a rigid linear chain. For that reason, to derive a

relationship between the stiffness parameter κ_t and the characteristic length of the curves decay in Figure 5.6, we use the following model

$$H(\{\boldsymbol{f}_i\}_{i=1\dots N}) = -\frac{\kappa_t}{\beta}\sum_{i=1}^{N-1}\boldsymbol{f}_i\cdot\boldsymbol{f}_{i+1} \tag{5.49}$$

For calculations convenience, let the couplings temporarily depend on the chain sites. The partition function reads:

$$Z = \prod_{i=1}^{N}\int_{-\pi}^{\pi} d\omega_i \prod_{j=1}^{N-1} e^{\beta J_j \cos(\omega_j)} = 2\pi\prod_{i=1}^{N-1} 2\pi\, I_0(\beta J_i) \tag{5.50}$$

with

$$I_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{x\cos\theta} d\theta \tag{5.51}$$

the first kind modified Bessel function of order 0. Similarly to the tangent-tangent correlation in Appendix B, we can compute $T(s)$ as

$$\begin{aligned} T_i(1) &= \frac{1}{\beta}\frac{\partial \ln Z}{\partial J_i} = \frac{1}{I_0(\beta J_i)}\frac{\partial I_0(\beta J_i)}{\partial(\beta J_i)} \\ T_i(s) &= \prod_{k=i}^{i+s-1} T_k(1) = \prod_{k=i}^{i+s-1}\frac{1}{I_0(\beta J_k)}\frac{\partial I_0(\beta J_k)}{\partial(\beta J_k)} \end{aligned} \tag{5.52}$$

Imposing again $J_i = \frac{\kappa_t}{\beta}\ \forall i$, rearranging $T_i(s)$ in exponential form and defining the modified Bessel

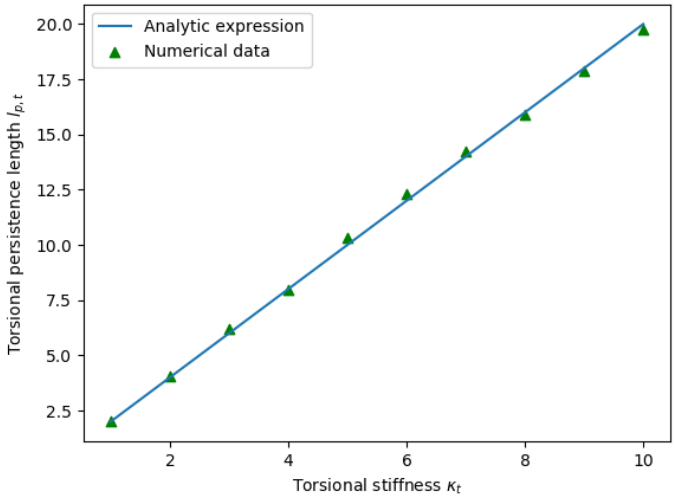

Figure 5.7: Relation between torsional persistence length $l_{p,t}$ and torsional stiffness parameter κ_t. The persistence length is expressed in simulation unit.

function of first kind and first order $I_1(x) = \frac{\partial I_0(x)}{\partial x}$, we obtain the so called torsional persistence length

$$l_{p,t} = \frac{b_0}{\ln\left(\frac{I_0(\kappa_t)}{I_1(\kappa_t)}\right)} \simeq \frac{b_0}{\ln(1+\frac{1}{2\kappa_t})} \simeq 2b_0\kappa_t \tag{5.53}$$

This relation has been found to be in very good agreement with the numerical data provided by the simulations (as shown in Figure 5.7).

Chapter 6

Results and open challenges

In recent experiments, both yeast minichromosomes and bacterial plasmids have been used to estimate the incidence of various knot types *in vivo* [28, 27]. Thanks to high resolution two-dimensional electrophoresis, different species have been separated in the gel according to their topological complexity and chirality. Indeed, it has been shown that the molecules' migration velocity in the gel is affected by the knot complexity and the interplay between handedness and supercoiling (see Figure 6.1).

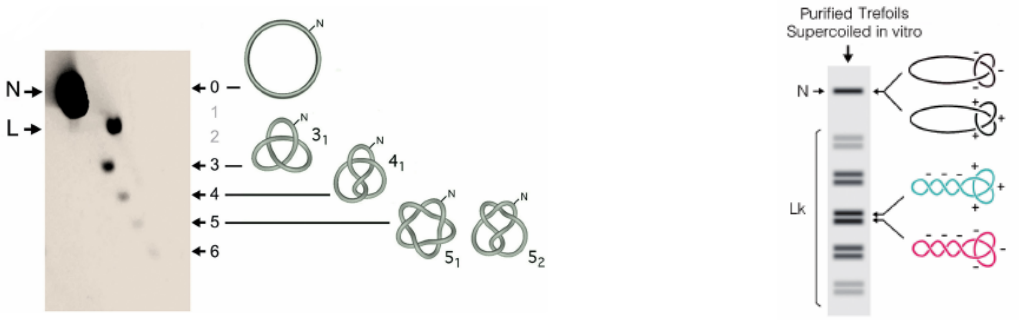

Figure 6.1: High resolution two-dimensional electrophoretic characterization of 4.4 kbp bacterial plasmids *in vitro*. On the left: knots separation on the basis of their topological complexity. On the right: bands splitting into doublets due to the different migration velocity of supercoiled trefoil enantiomers. Images from [28, 27].

In particular, the following main aspects emerged:

1. As the knot crossing number increases, the molecules migrate faster. By suitably tuning the electric fields in vertical and horizontal directions, knotted molecules can be isolated from linear fragments, and the knots ladder in Figure 6.1 can be obtained.

2. Trefoil knots of different handedness and approximately same supercoiling level split into doublets. Conversely, nicked enantiomers concentrates in a single band. This means that the coupling between chirality and supercoiling can play a role in the molecules migration velocity.

3. As the supercoiling level increases, molecules acquire greater migration velocities in the gel (see the separation between different doublets in Figure 6.1).

Model and simulations With this work, we intended to give an interpretation to the set of measurements just mentioned. Even though it is a very common experimental technique, the electrophoretic migration is still not completely understood from a theoretical point of view. Indeed, the phenomenology provided by electrophoresis experiments does not give a direct insight into the way knotting and supercoiling affect the migration velocity and into their separated effects, which we were able to better control with our model.
To investigate such phenomenology without simulating the molecules dynamics with an electric field during the electrophoresis, we sampled the configurational space of molecules in bulk (outside the gel) and we tried to ascribe the differences in migration velocities to the systems conformational properties. In particular, we systematically profiled geometric observables as gyration radius, asphericity, writhe and average crossing number to characterize the molecular compaction and shape.

The species reported in Table 6.1 have been analyzed, and by properly comparing them we were able to separately evaluate the extent to which knotting and supercoiling affect the molecule conformational properties. Being mirror images of each other, only one of the nicked trefoil enantiomers

Species	Number of runs	Snapshots per run	Dump interval	Independent samples
Nicked 0_1	5	≈ 38.000	10^7	≈ 1300
Supercoiled 0_1	5	≈ 70.000	10^7	≈ 2300
Nicked 3_1^R	5	≈ 30.000	10^7	≈ 1000
Supercoiled 3_1^R	5	≈ 46.000	10^7	≈ 1500
Supercoiled 3_1^L	5	≈ 38.000	10^7	≈ 1300

Table 6.1: The simulated species are listed in this table, together with the duration of each run, the number of Monte Carlo steps after which we saved a configuration (dump interval) and the number of independent samples acquired.

Figure 6.2: Initialization of supercoiled unknot and right handed trefoil. The $\{\boldsymbol{f}_i\}_{i=1\ldots N}$, represented in blue, give the idea of how much the molecule has been twisted to obtained the desired supercoiling level.

has been simulated. To initialize the supercoiled species, we started from the planar ring and the (almost) ideal trefoil configurations and we gradually twisted them by rotating the body fixed coordinate systems along the chain up to reach the target Lk (see Figure 6.2, where also the $\{\boldsymbol{f}_i\}_{i=1...N}$ vectors have been represented). For the nicked molecules, we simply removed the torsional energy contribution coming from the last twisting angle $\alpha_N + \gamma_N$. To get a feeling of how the various species evolved during the Monte Carlo simulations, some snapshots sequences are reported.

Supercoiled 0_1 (red) and nicked 0_1 (green):

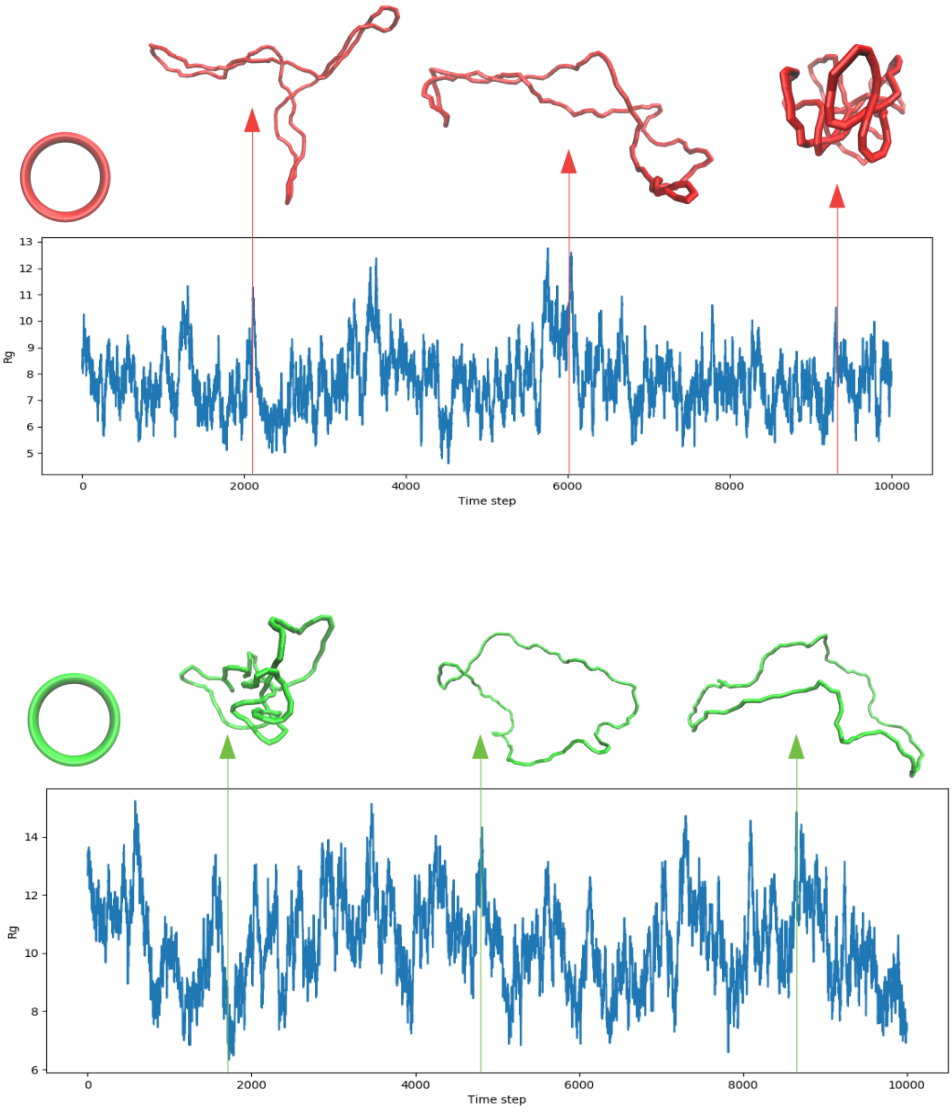

Figure 6.3: Gyration radius time series relative to supercoiled (red) and nicked (green) 0_1 with some snapshots taken at the Monte Carlo time steps corresponding to the arrows positions.

Supercoiled 3_1^R (red), supercoiled 3_1^L (orange):

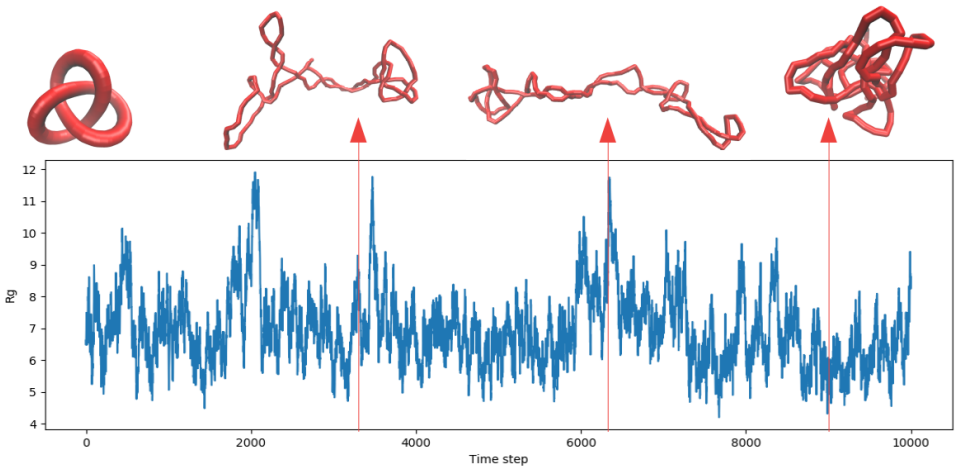

Figure 6.4: Gyration radius time series relative to supercoiled 3_1^R with some snapshots taken at the Monte Carlo time steps (in units of dump interval) corresponding to the arrows positions.

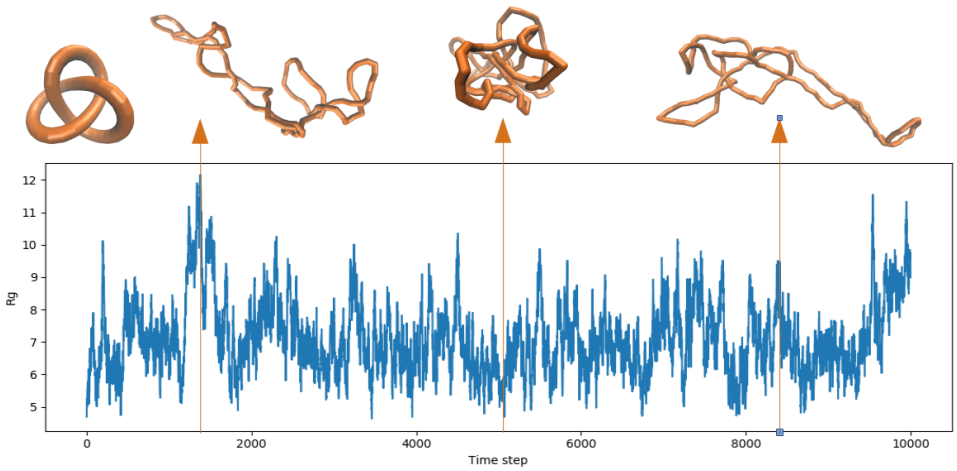

Figure 6.5: Gyration radius time series relative to supercoiled 3_1^L with some snapshots taken at the Monte Carlo time steps (in units of dump interval) corresponding to the arrows positions.

Nicked 3_1^R :

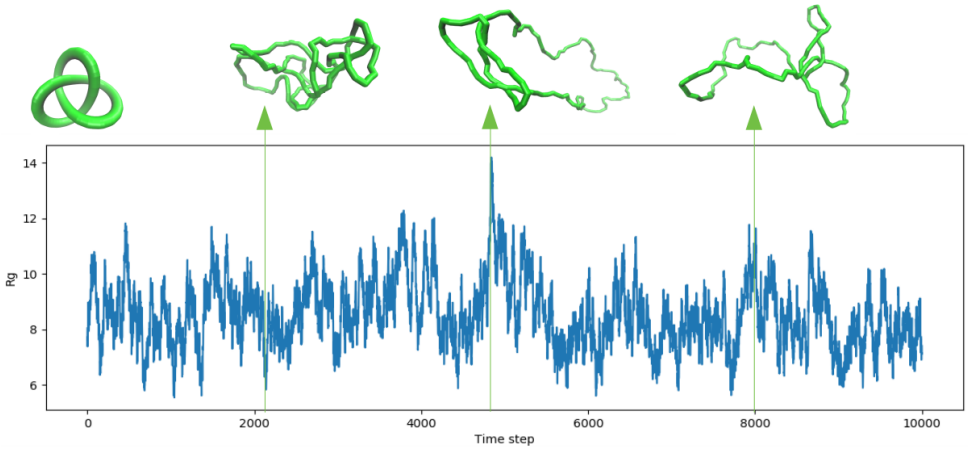

Figure 6.6: Gyration radius time series relative to nicked 3_1^R with some snapshots taken at the Monte Carlo time steps (in units of dump interval) corresponding to the arrows positions.

6.1 Purely topological effects

To isolate the topological contribution to the molecular compaction and shape, we compared the nicked unknot and trefoil (as in Figure 6.1). Indeed, these species are free to dissipate their torsional stress, which consequently does not come into play in their spatial arrangement.

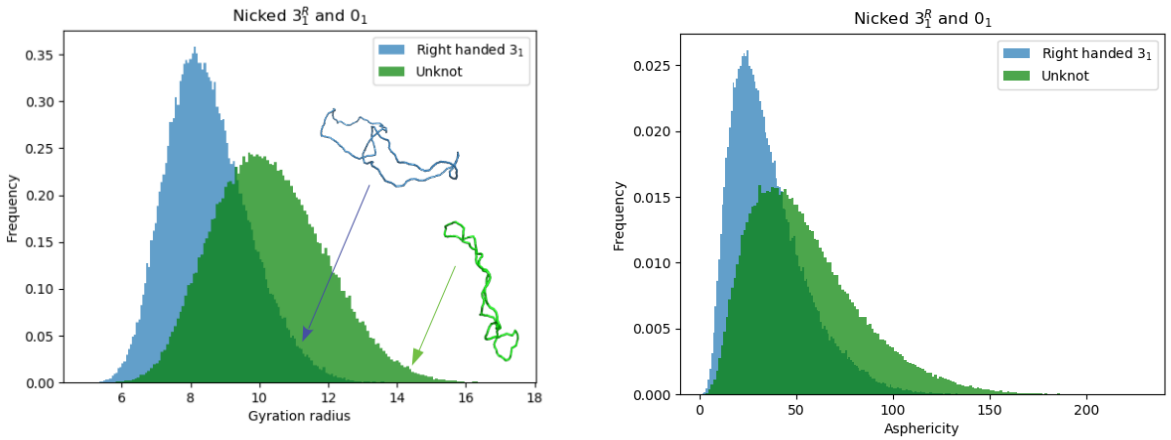

Figure 6.7: Gyration radius and asphericity distributions of nicked unknot (green) and right-handed trefoil (blue). Two typical configurations give an idea of why the knot constraint prevents the trefoil to elongate as the unknot.

As expected, the unknotted chain can more easily assume elongated configurations with higher R_g since it lacks the topological constraint due to the knot presence. For this reason, also its

asphericity distribution decays much more slowly than the trefoil one, which tends to concentrate on more compact conformations.

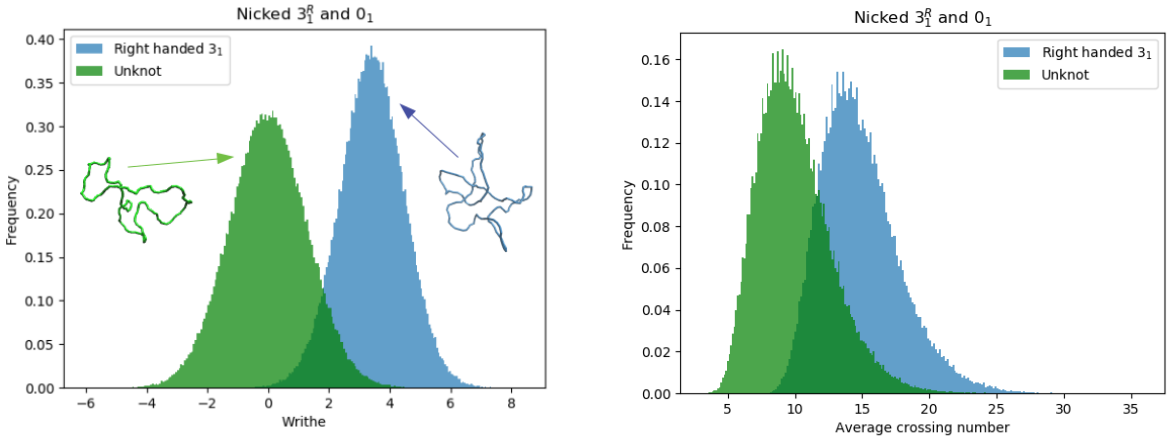

Figure 6.8: Writhe and average crossing number distributions of nicked unknot (green) and right-handed trefoil (blue).

The writhe and ACN difference between the two species induced by the knot intrinsic crossings, is reported in Figure 6.8. The constraint imposed by the knot presence affects also the distributions variance, reducing the range of variability of R_g, asphericity and writhe for the trefoil. The analysis of these geometrical observables is in agreement with the experimental results that show a greater gel migration velocity for knotted DNA rings (see Figure 6.1).

6.2 Supercoiling effects

To evaluate the extent to which supercoiling affects molecular compaction and shape, we compared both the results of nicked and supercoiled chains at fixed topology, and those of supercoiled unknot and trefoil. The superhelical density of all species has been set to $\approx -2\%$, namely the value at which experimentalists found a clearer separation between the doublet bands corresponding to trefoil enantiomers [27]. More precisely, the ΔLk_e of the various species were

$$\begin{cases} \Delta Lk_e(0_1) = -10 \\ \Delta Lk_e(3_1^L) \simeq -9.6 \\ \Delta Lk_e(3_1^R) \simeq -10.4 \end{cases} \tag{6.1}$$

where the difference between the trefoil enantiomers is due to the extra-twisting that we added to satisfy the closure conditions.

6.2.1 Unknot topology

The plectonemic structures that appear when supercoiling is added, induce a reduction of the molecule effective size as shown in Figure 6.9. Because of this compaction, the DNA ring tends to populate microstates with higher spherical symmetry and the asphericity distribution becomes

more peaked. The plectonemes superhelical arrangement and the high entanglement degree due to supercoiling substantially contribute to increase the average crossing number (see the net separation of the ACN distributions in Figure 6.10).

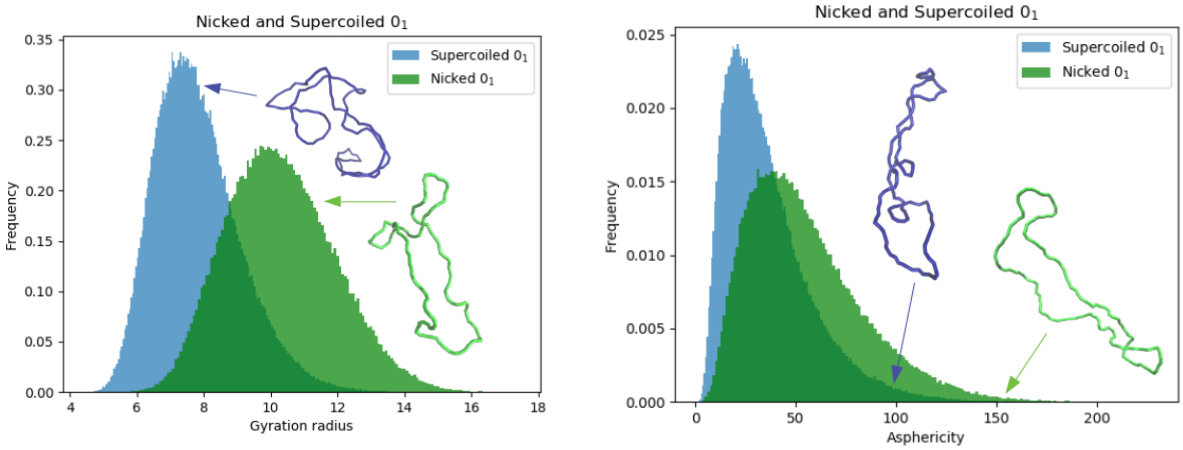

Figure 6.9: Gyration radius and asphericity distributions of nicked unknot (green) and supercoiled unknot (blue).

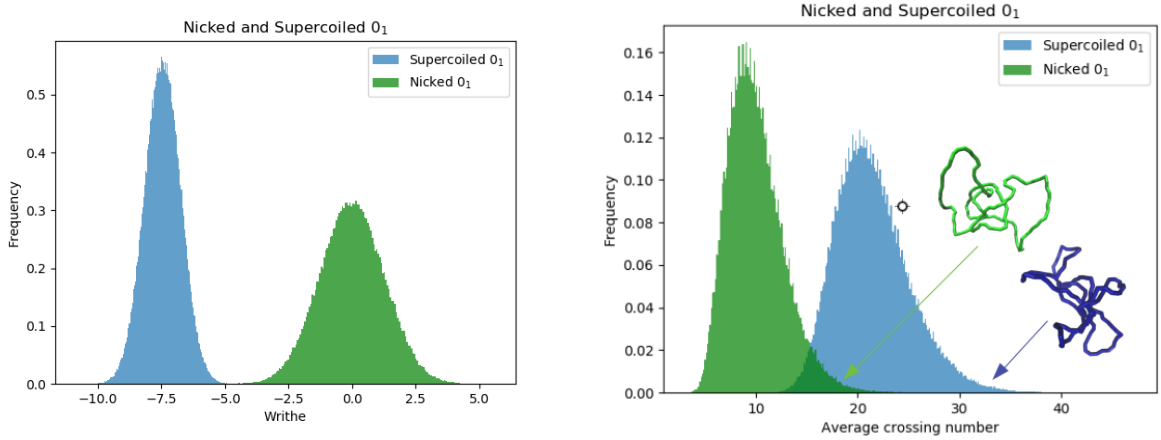

Figure 6.10: Writhe and average crossing number distributions of nicked unknot (green) and supercoiled unknot (blue).

6.2.2 Trefoil topology

An analogous behavior has been observed for knotted molecules. The energy excess introduced in the inizialization by twisting the chain is subsequently distributed into torsional and bending contributions, accordingly to their stiffness parameters. The ratio between κ_t and κ_b is a key aspect for determining the extent to which supercoiling is able to make molecules more compact. In our case, the gyration radius distributions relative to the nicked and supercoiled trefoils can

be clearly distinguished (see plot on the left in Figure 6.11). This conformational difference is in agreement with the clear doublets separation obtained with high resolution two dimensional gel electrophoresis [27]. Moreover, configurations with high values of asphericity are further penalized by the supercoiling, as can be evinced by comparing the distribution decays in the right part of Figure 6.11.

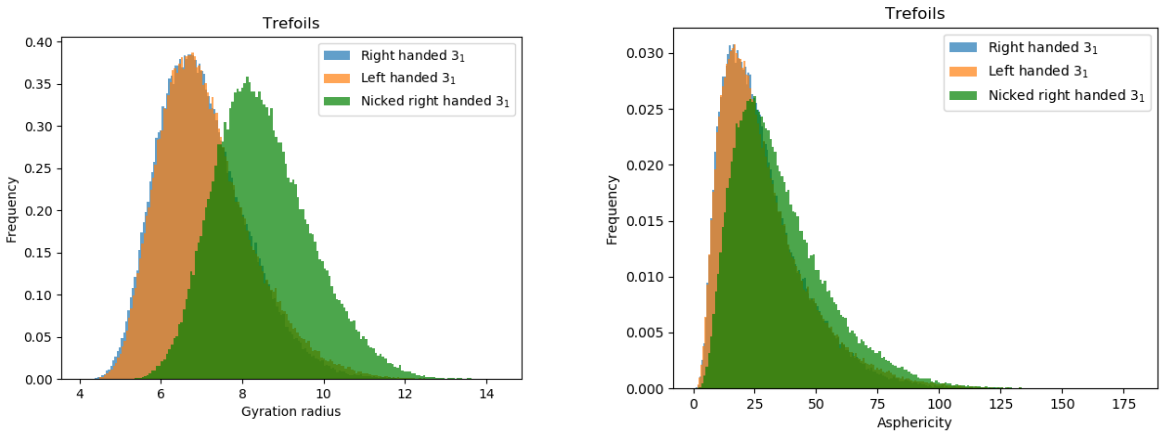

Figure 6.11: Gyration radius and asphericity distributions of nicked trefoil (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange).

However, we did not find structural elements with which we could justify the different gel migration velocity for the two supercoiled enantiomers. Indeed, we observed an almost perfect overlap of their gyration radius and asphericity distributions. A part from the writhe, whose average values are perfectly distinguishable due to the opposite sign of the enantiomers intrinsic crossings, the only observable that exhibited some difference is the average crossing number (see right part of Figure 6.12).

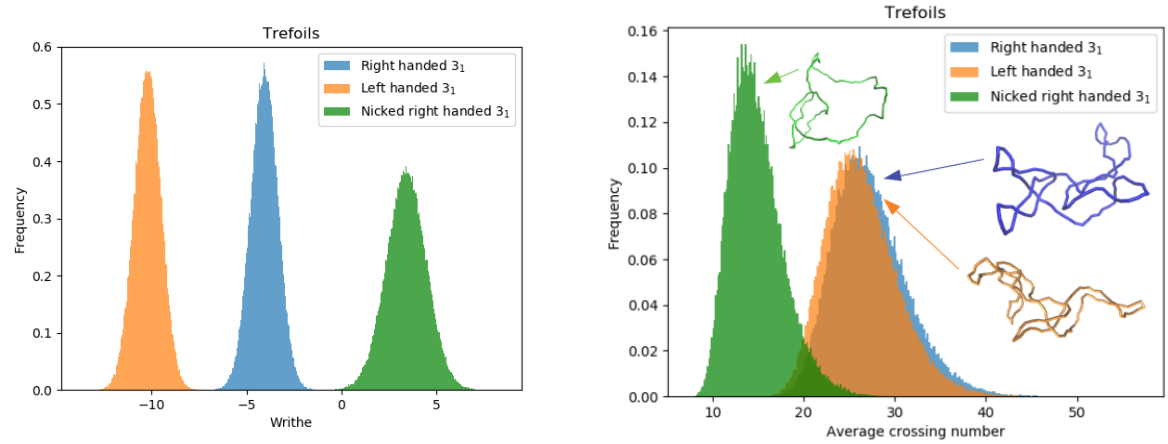

Figure 6.12: Writhe and average crossing number distributions of nicked trefoil (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange).

However, we attributed the ACN distributions shift to the fact that the two enantiomers do not present exactly the same supercoiling level for reasons explained at the beginning of this section. In particular, their ΔLk_e differ by almost one unit: being the positive-noded trefoil a bit more supercoiled, its tendency to assume higher ACN value is justified.

6.2.3 Comparing supercoiled 0_1 and 3_1

Finally we compared the results for molecule with different topology and, approximately, the same supercoiling level. Also in this case, the gyration radius distributions confirm the compaction induced by the knot presence. From the shape point of view, we observed that trefoils favored more spherically symmetric configurations than the unknot. Moreover, as expected, the intrinsic crossings due to the knotting provoked an increasing of the average crossing number with respect to the unknotted chain.

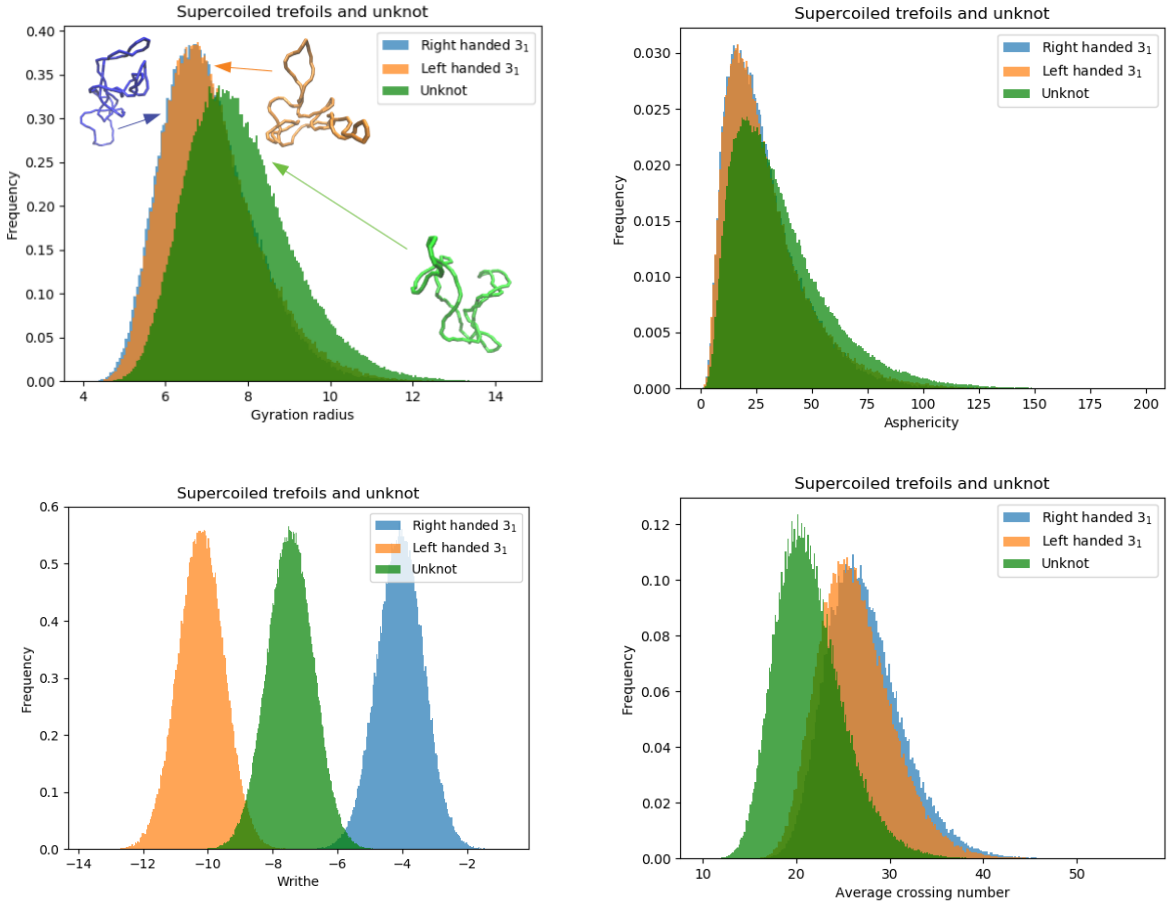

Figure 6.13: Gyration radius, asphericity, writhe and average crossing number distributions of supercoiled unknot (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange).

6.3 Conclusions

In this work, we developed a coarse grained model to describe closed DNA chains with different topology and supercoiling levels, and we carried out Monte Carlo simulations to sample their configurational space at canonical equilibrium.
Through the analysis of geometric observables we gave an insight into how knotting and supercoiling can affect the molecular compaction and shape. In particular, our results support the following experimental evidences resulting from agarose gel electrophoresis [28, 27]:

1. Knotted molecules acquire a greater gel migration velocity with respect to unknotted ones (see left part of Figure 6.1).

2. At fixed topology, the more DNA chains are supercoiled, the faster they migrate through the gel pores.

Open challenge: We found that the phenomenology relative to the splitting of electrophoretic bands corresponding to supercoiled trefoil enantiomers (see Figure 6.1) cannot be elucidated by the conformational properties we studied. We hypothesized that the static geometrical quantities we looked at are not enough to capture the key aspect behind such bands splitting and that it could be worth to consider how the gel architecture interfere with the DNA chain dynamics. We will further investigate the appearance of trefoil doublets in electrophoretic measurements by trying to improve our model in such a way that it takes into account the interplay between system and gel structure during the migration.

Appendix A

As reported in Chapter 3, the writhe Wr of an arbitrary curve $\gamma : \mathbb{R} \to \mathbb{R}^3$ can be computed by means of the Gauss integral

$$Wr = \frac{1}{4\pi} \int_\gamma \int_\gamma \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1)$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are two points on γ. In the case of a polygonal closed curve with N segments $\gamma = \bigcup_{i=1}^N \gamma_i$, the latter expression can be rewritten as a finite sum

$$\begin{aligned} Wr &= \frac{1}{4\pi} \int_{\bigcup_{i=1}^N \gamma_i} \int_{\bigcup_{j=1}^N \gamma_j} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) = \frac{1}{4\pi} \sum_{i=1}^N \sum_{j=1}^N \int_{\gamma_i} \int_{\gamma_j} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{||\boldsymbol{r}_2 - \boldsymbol{r}_1||^3} \cdot (d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) \\ &= 2 \sum_{i=2}^N \sum_{j<i} \frac{\Omega_{ij}}{4\pi} \end{aligned}$$

with Ω_{ij} the Gauss integral contribution coming from the segments i and j. Notice that $\Omega_{ii} = 0$ because $d\boldsymbol{r}_1$ and $d\boldsymbol{r}_2$ would be parallel and so $(d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1) = 0$ over the whole domain, and $\Omega_{i,i+1} = 0$ since $(d\boldsymbol{r}_2 \wedge d\boldsymbol{r}_1)$ and $\boldsymbol{r}_2 - \boldsymbol{r}_1$ would be perpendicular. At this stage, an analytical expression for

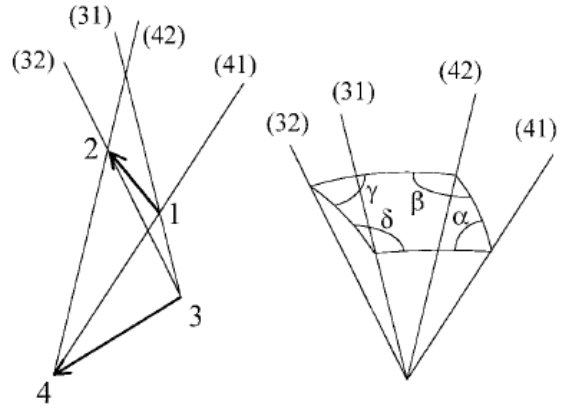

Figure 6.14: On the left: sketch of the observation directions which limit the solid angle needed to compute the Gauss integral Ω_{ij}. On the right: the quadrangle whose area gives $\tilde{\Omega}_{ij}$. Image from [14].

Ω_{ij} has to be found. Let's call $1, 2$ the ends of the segment i, and $3, 4$ the ones of j. In this way the two segments will correspond to the vectors $\boldsymbol{r}_{12}$ and $\boldsymbol{r}_{34}$. The Gauss integral absolute value $\tilde{\Omega}_{ij}$ is given by the solid angle which includes all the observation directions such that $\boldsymbol{r}_{12}$ and $\boldsymbol{r}_{34}$ form an apparent crossing, with $\boldsymbol{r}_{12}$ closer to the observation point. By looking at the Figure 6.14, one can evince that such a solid angle is delimited by the planes $(3, 4, 1)$, $(3, 4, 2)$, $(1, 2, 4)$ and $(1, 2, 3)$, and can be built by letting the lines (32), (31), (42) and (41) pass through the origin and taking their intersection with the unitary sphere. In this way, the solid angle value corresponds with the quadrangle area depicted on the right of Figure 6.14:

$$\tilde{\Omega}_{ij} = \alpha + \beta + \gamma + \delta - 2\pi$$

To compute these angles one define the following versors:

$$\boldsymbol{n}_1 = \frac{\boldsymbol{r}_{13} \wedge \boldsymbol{r}_{14}}{|\boldsymbol{r}_{13} \wedge \boldsymbol{r}_{14}|} \qquad \boldsymbol{n}_2 = \frac{\boldsymbol{r}_{14} \wedge \boldsymbol{r}_{24}}{|\boldsymbol{r}_{14} \wedge \boldsymbol{r}_{24}|} \qquad \boldsymbol{n}_3 = \frac{\boldsymbol{r}_{24} \wedge \boldsymbol{r}_{23}}{|\boldsymbol{r}_{24} \wedge \boldsymbol{r}_{23}|} \qquad \boldsymbol{n}_4 = \frac{\boldsymbol{r}_{23} \wedge \boldsymbol{r}_{13}}{|\boldsymbol{r}_{23} \wedge \boldsymbol{r}_{13}|}$$

So, one has:

$$\begin{cases} \alpha = \frac{\pi}{2} + \arcsin(\boldsymbol{n}_1 \cdot \boldsymbol{n}_2) \\ \beta = \frac{\pi}{2} + \arcsin(\boldsymbol{n}_2 \cdot \boldsymbol{n}_3) \\ \gamma = \frac{\pi}{2} + \arcsin(\boldsymbol{n}_3 \cdot \boldsymbol{n}_4) \\ \delta = \frac{\pi}{2} + \arcsin(\boldsymbol{n}_4 \cdot \boldsymbol{n}_1) \end{cases}$$

and

$$\tilde{\Omega}_{ij} = \arcsin(\boldsymbol{n}_1 \cdot \boldsymbol{n}_2) + \arcsin(\boldsymbol{n}_2 \cdot \boldsymbol{n}_3) + \arcsin(\boldsymbol{n}_3 \cdot \boldsymbol{n}_4) + \arcsin(\boldsymbol{n}_4 \cdot \boldsymbol{n}_1)$$

At this point the crossing sign should be introduced to get Ω_{ij}. In particular, a right handed crossing occurs whenever $(\boldsymbol{r}_{34} \wedge \boldsymbol{r}_{12}) \cdot \boldsymbol{r}_{13} > 0$, so that

$$\Omega_{ij} = \tilde{\Omega}_{ij} \ \mathrm{sign}\big((\boldsymbol{r}_{34} \wedge \boldsymbol{r}_{12}) \cdot \boldsymbol{r}_{13}\big)$$

Other alternatives to evaluate the writhe of a closed polygonal curve are proposed in [14].

Appendix B

Tangent-tangent correlation function and bending persistence length

A relation between bending persistence length and κ_b can be found by evaluating the extent to which the bonds directionality is influenced by the bending energy in a linear DNA chain. For doing that, we consider the Kratky-Porod model

$$H_{KP} = -\frac{\kappa_b}{\beta} \sum_{i=1}^{N-1} \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} \tag{6.2}$$

with $\{\boldsymbol{u}_i\}_{i=1\ldots N-1}$ the bond vectors already defined in subsection 5.1.2, and we compute the tangent-tangent correlation

$$G_i(s) =< \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+s} > \tag{6.3}$$

For calculation convenience let us temporarily suppose to have a site dependent coupling

$$H_{KP} = -\sum_{i=1}^{N-1} J_i \, \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} \tag{6.4}$$

Being the partition function

$$\begin{aligned} Z &= \left(\prod_{i=1}^{N} \int d\Omega_i \right) \exp(\beta \sum_{j=1}^{N-1} J_j \, \boldsymbol{u}_j \cdot \boldsymbol{u}_{j+1}) = \left(\prod_{i=1}^{N} \int d\Omega_i \right) \prod_{j=1}^{N-1} \exp(\beta J_j \, \boldsymbol{u}_j \cdot \boldsymbol{u}_{j+1}) = \\ &= 4\pi \prod_{j=1}^{N-1} 4\pi \frac{\sinh(\beta J_j)}{\beta J_j} \end{aligned} \tag{6.5}$$

we can readily find

$$G_i(s=1) = \frac{1}{\beta} \frac{\partial \ln Z}{\partial J_i} = \coth(\beta J_i) - \frac{1}{\beta J_i} = L(\beta J_i) \tag{6.6}$$

where $L(x)$ stands for Langevin function. The tangent-tangent correlation $G_i(s)$ with $s > 1$, instead, can be calculated by taking advantage of the following symmetry argument. Let us take $< \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+2} >$: the two bonds could be decomposed into their parallel and normal component to $\boldsymbol{u}_{i+1}$

$$G_i(2) =< \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+2} >=< u_{i,//} u_{i+2,//} > + < u_{i,\perp} u_{i+2,\perp} > \tag{6.7}$$

For each set of bonds $\{\boldsymbol{u}_i\}_{i=1\dots N-1}$, let us build $\{\boldsymbol{u}'_i\}_{i=1\dots N-1}$ as

$$\begin{cases} u'_{j,//} = u_{j,//} & \forall j \\ u'_{j,\perp} = u_{j,\perp} & \text{if } j \leq i+1 \\ u'_{j,\perp} = -u_{j,\perp} & \text{if } j > i+1 \end{cases} \tag{6.8}$$

Since these two configurations have same energy but opposite $u_{i,\perp}u_{i+2,\perp} = -u'_{i,\perp}u'_{i+2,\perp}$, we have $< u_{i,\perp}u_{i+2,\perp} >= 0$. Thus

$$\begin{aligned} G_i(2) &=< u_{i,//}u_{i+2,//} >=< (\boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1})(\boldsymbol{u}_{i+1} \cdot \boldsymbol{u}_{i+2}) >= \frac{1}{\beta^2 Z}\frac{\partial Z}{\partial J_i \partial J_{i+1}} = \\ &= L(\beta J_i)L(\beta J_{i+1}) =< \boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} >< \boldsymbol{u}_{i+1} \cdot \boldsymbol{u}_{i+2} >= \\ &= G_i(1)G_{i+1}(1) \end{aligned} \tag{6.9}$$

The same reasoning can be iterated to get

$$G_i(s) = \prod_{k=i}^{i+s-1} G_k(1) = \prod_{k=i}^{i+s-1} L(\beta J_k) \tag{6.10}$$

Finally, imposing $J_k = \frac{\kappa_b}{\beta}\ \forall k$ again, we have:

$$G_i(s) = L^s(\kappa_b) = e^{s\ln(L(\kappa_b))} \tag{6.11}$$

By rearranging the previous expression, the characteristic length l_p of the tangent-tangent correlation decay can be evaluated

$$l_p = -\frac{b_0}{\ln(L(\kappa_b))} \tag{6.12}$$

This quantity is called bending persistence length. For small bending angles, the l_p expression can be expanded as

$$l_p \simeq -\frac{b_0}{\ln(1-\frac{1}{\kappa_b})} \simeq b_0\,\kappa_b \tag{6.13}$$

Alternative R_g^2 formula

As anticipated in Chapter 5, the gyration radius R_g^2 can be described by the equivalent formula

$$R_g^2 = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_j)^2 \tag{6.14}$$

To prove it, let us start from the original definition

$$\begin{aligned} R_g^2 &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_{CM})^2 = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i^2 - 2\boldsymbol{r}_i \cdot \boldsymbol{r}_{CM} + \boldsymbol{r}_{CM}^2) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i^2 - \boldsymbol{r}_{CM}^2) = \\ &= \frac{1}{N^2}\sum_{i,j}(\boldsymbol{r}_i^2 - \boldsymbol{r}_i \cdot \boldsymbol{r}_j) = \frac{1}{2N^2}\sum_{i,j}(\boldsymbol{r}_i - \boldsymbol{r}_j)^2 \end{aligned} \tag{6.15}$$

$G(s)$ for a two-dimensional ring

To prove the expression 5.45 we start from the Kratky-Porod Hamiltonian for a two-dimensional planar ring

$$H_{KP} = -\frac{\kappa_b}{\beta}\sum_{i=1}^{N}\boldsymbol{u}_i\cdot\boldsymbol{u}_{i+1} = \frac{\kappa_b}{\beta}\sum_{i=1}^{N}\big(\frac{1}{2}(\boldsymbol{u}_{i+1}-\boldsymbol{u}_i)^2-1\big) = -\frac{\kappa_b N}{\beta}+\frac{l_p}{2\beta}\sum_{i=1}^{N}\frac{(\boldsymbol{u}_{i+1}-\boldsymbol{u}_i)^2}{b_0^2}b_0 \quad (6.16)$$

and we take the continuum limit (the constant part can be discarded)

$$H_{WLC} = \frac{k_B T\, l_p}{2}\int_0^L\left(\frac{\partial\boldsymbol{u}(s)}{\partial s}\right)^2 ds \quad (6.17)$$

to obtain a *worm-like-chain* Hamiltonian, with $s\in[0,L]$ the curvilinear coordinate that indicates the position along the molecule. A perfectly circular chain will be described by

$$\begin{cases}\boldsymbol{r}(s) = \big(\cos\theta_0(s),\sin\theta_0(s)\big)\\ \boldsymbol{u}(s) = \big(-\sin\theta_0(s),\cos\theta_0(s)\big)\\ \theta_0(s) = \frac{2\pi s}{L}\end{cases} \quad (6.18)$$

Since we would like to focus on the stiff regime, we will characterize molecule configurations by means of small deviation $\delta\theta(s) = \theta(s)-\theta_0(s)$ from the ideal angle $\theta_0(s)$. The Hamiltonian becomes

$$\begin{aligned}H_{WLC} &= \frac{k_B T\, l_p}{2}\int_0^L\left(\frac{\partial\boldsymbol{u}(s)}{\partial s}\right)^2 ds = \frac{k_B T\, l_p}{2}\int_0^L\left(\frac{\partial\theta(s)}{\partial s}\right)^2 ds\\ &= \frac{k_B T\, l_p}{2}\frac{(2\pi)^2}{L}+\frac{k_B T\, l_p}{2}\int_0^L\left(\frac{\partial\delta\theta(s)}{\partial s}\right)^2 ds\end{aligned} \quad (6.19)$$

where we used $\delta\theta(0) = \delta\theta(L)$. Again, the constant term can be discarded. The quantity $\delta\theta(s)$ can be decomposed in its Fourier components

$$\delta\theta(s) = \frac{1}{L}\sum_{n=-\infty}^{+\infty}\tilde{\theta}_n e^{\frac{2\pi i n s}{L}} \quad (6.20)$$

where $\tilde{\theta}_0 = \tilde{\theta}_{\pm1} = 0$ to satisfy the closure condition. So

$$\begin{aligned}H_{WLC} &= -\frac{k_B T\, l_p}{2L^2}\int_0^L\sum_{n=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty}\frac{2\pi n}{L}\frac{2\pi m}{L}\tilde{\theta}_n\tilde{\theta}_m e^{\frac{2\pi n i s}{L}}e^{\frac{2\pi i m s}{L}} =\\ &= \frac{k_B T\, l_p\,(2\pi)^2}{2L^3}\sum_{n=-\infty}^{+\infty}n^2|\tilde{\theta}_n|^2 = \frac{k_B T\, l_p\,(2\pi)^2}{L^3}\sum_{n=2}^{+\infty}n^2|\tilde{\theta}_n|^2\end{aligned} \quad (6.21)$$

The average value $<|\tilde{\theta}_n|^2>$ can be computed by applying the equipartition theorem:

$$<|\tilde{\theta}_n|^2> = \frac{L^3}{l_p(2\pi)^2 n^2}\quad |n|\geq 2 \quad (6.22)$$

At this point, we are ready to evaluate the tangent-tangent correlation function

$$\begin{aligned} G(s) &=< \boldsymbol{u}(s)\cdot\boldsymbol{u}(0) >= \sin\theta(s)\sin\theta(0)+\cos\theta(s)\cos\theta(0) = \\ &= \cos\big(\theta_0(s)+[\delta\theta(s)-\delta\theta(0)]\big) \end{aligned} \tag{6.23}$$

Expanding the cosine to second order we find

$$G(s)\simeq\cos\theta_0(s)-\frac{1}{2}<[\delta\theta(s)-\delta\theta(0)]^2>=\cos\theta_0(s)\big[1-<(\delta\theta(s))^2>+<\delta\theta(s)\delta\theta(0)>\big] \tag{6.24}$$

where the translational invariance $<(\delta\theta(s))^2>=<(\delta\theta(0))^2>$ has been used. To evaluate these averages we write

$$<\delta\theta(s)\delta\theta(r)>=\frac{1}{L^2}\sum_{n=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty}<\tilde{\theta}_n\tilde{\theta}_m>e^{\frac{2\pi ins}{L}}e^{\frac{2\pi imr}{L}}=\frac{L}{l_p(2\pi)^2}\sum_{n=-\infty}^{+\infty}\frac{1}{n^2}e^{\frac{2\pi in(s-r)}{L}} \tag{6.25}$$

which specializes in

$$<\big(\delta\theta(s)\big)^2>=\frac{L}{l_p 2\pi^2}\sum_{n=2}^{+\infty}\frac{1}{n^2} \tag{6.26}$$

and

$$<\delta\theta(s)\delta\theta(0)>=\frac{L}{l_p 2\pi^2}\sum_{n=2}^{+\infty}\frac{1}{n^2}\cos\left(\frac{2\pi ns}{L}\right) \tag{6.27}$$

Defining

$$g(z)=\sum_{n=2}^{+\infty}\frac{\cos(nz)}{n^2} \tag{6.28}$$

and putting all together we finally find

$$G(s)=\left[1+\frac{L}{2\pi^2 l_p}\left(g\left(\frac{2\pi s}{L}\right)-g(0)\right)\right]\cos\left(\frac{2\pi s}{L}\right) \tag{6.29}$$

List of Figures

1.1 Illustration of the knot operational definition. Image from [1]. 3
1.2 Knot disappearance with a continuous transformation which does not involve the ambient space. Image from [16]. 4
1.3 Type I Reidemeister move 5
1.4 Type II Reidemeister move 5
1.5 Type III Reidemeister move 5
1.6 Table of knots up to crossing number $c(K) = 7$. 6
1.7 The knot 7_2 can be unknotted performing a single strand passage where denoted by the dashed circle. Image from [1]. 7
1.8 Left- and right-handed version of the trefoil knot 8
1.9 Right-hand rule to establish the sign of a crossing 8
1.10 Connected sum of two left-handed trefoils. The resulting knot is called granny knot. See Other sources. 9
1.11 Square knot, obtained by doing the connected sum of two trefoils with different handedness. 10
1.12 Examples of alternating knots: 4_1 (figure-8-knot) on the left and 5_2^R on the right . . 10
1.13 On the left: the trefoil 3_1 is the easiest torus knot. On the right: sketch of meridians and longitudes. Image from [1]. 11
1.14 Example of a twist knot with 6 half twists (8_1). See Other sources. 12
1.15 Example of crossings numbering for the Dawker code computation (in this case 6, −14, 16, −12, 2, −4, −8, 10). Image from [1]. 12
1.16 Crossings and arcs numbering for the 5_1^L Alexander matrix computation. 13

2.1 Nitrogenous bases structure, with IUPAC numbering for purines and pyramidines. Image from [3]. 15
2.2 On the left: B-form DNA representation. The minor and major grooves are indicated on the left and the perpendicularity between nitrogenous bases and helical axis can be appreciated from this point of view. Image from [32] .On the right: chemical structure of AT and CG pairings, with hydrogen bonds indicated by the dashed lines. 16
2.3 DNA chemical structure. The phosphates groups with their negatively charged residues are highlighted with yellow circles. The deoxyriboses are colored in orange, the adenines in green, the thymines in purple, the guanines in blue and the cytosines in red. See Other sources. 17

2.4 Some configurations of neighboring base-pairs: the tilt and roll angles θ_1 and θ_2 describe the local molecule bending, while θ_3 refers to the local torsion. Image from [18]. 18
2.5 Electron micrograph of nucleosomal fibers and their "beads on a string" appearance. The nucleosomes are indicated by the black arrows. The size marker on the top-right correspond to 30 nm. Image taken from [17]. 18
2.6 Representation of the chromatin hierarchical organization. From left to right: a simple DNA duplex, a nucleosomal fiber, a solenoidal structure, some supercoiled loops of a decondensed interphase chromosome. Image from [19]. 19
2.7 (A): energetically relaxed linear DNA fragment, with 8 helical turns. (B): circular chain obtained by closing the linear one without introducing any twist (still torsionally relaxed). (C): circular chain obtained by under-winding of one turn the linear fragment and then circularizing the molecule. This configuration is not torsionally relaxed and it will evolve in (D), where the balance between bending and twisting stiffness results in a figure-eight contorted conformation. Image from [32]. 20
2.8 From left to right: a closed DNA duplex chain is characterized by 3 negative supercoils. A type IA topoisomerase cleave to one strand of the duplex and it forms a phopshotyrosine bond with the $5'$ end of the red filament. The blue backbone passes through the break, decreasing the number of negative supercoils by 1. See Other sources. 21
2.9 Cross inversion mechanism of the type IIA topoisomerase. The enzyme presents three gates, called from top to bottom N-gate, DNA-gate, C-gate. They are mechanically coupled in such a way that a gate can be open only if the other two are closed. The double strand which is going to be gated (broken) is called G-segment, while the intact duplex which passes through the gate is called T-segment. Image from [20]. . 22
2.10 Sketch of replication process and main enzymes which come into play. Topoisomerases relieves the supercoiling excess generated by the replication fork progression. DNA polymerases elongate the duplex both in the leading and lagging strands. RNA primase and primer are present every time that an Okazaki fragment is being synthesized on the lagging strand (and also at the beginning of the leading strand). See Other sources. 23
2.11 Transcription process representation. The RNA polymerase reads the $3' \to 5'$ template strand (the lighter one) and progressively adds the RNA nucleotides according to the base complementarity rule. Remember that in RNA the thymine is substituted by the uracil. As soon as the mRNA strand is synthesized, it detaches from the template strand. The DNA unwinding and rewinding highlighted on the two sides of the transcription bubble are responsible for the supercoiling accumulation in the rest of the duplex. See Other sources. 25
2.12 Translation process representation, with an active ribosome traversing the mRNA in $5' \to 3'$ direction. The exit E-site in correspondence of the right ribosome side is omitted, and only empty leaving tRNA molecules are depicted. The amino acid on the A-site is forming a peptide bond with the polypeptide chain, which is going to be transferred from P- to A-site. Other aminoacyl-tRNA molecules are waiting to reach their corresponding codon. See Other sources. 26

3.1 Sketch of the $\boldsymbol{\phi}$ backbone around the curved cylinder and its helical axis $\boldsymbol{\alpha}$. At the cross-section level parametrized by t, the unit vector tangent to the helical axis $\boldsymbol{u}(t)$ and $\boldsymbol{v}_{\alpha\to\phi}(t)$ are represented. Image adapted from [9]. 29
3.2 On the left: an under-wound DNA duplex which has just been circularized ($n_{extra} = -1$). The linking number Lk initially corresponds to the twist Tw ($Lk = Tw = n - 1$). During its evolution, the DNA duplex will release a bit of torsional strain by increasing its bending energy. In particular, on the right, it reaches a "figure-of-eight" arrangement. Since the two double strand portions are extremely close in correspondence of the crossing, one may estimate the following values for the final configuration: $Wr \approx -1$, $Tw \approx n$ such that Lk is exactly $n - 1$. Image adapted from [32]. 32
3.3 On the left: the Whitehead link is an example of intertwined curves whose $Lk = 0$. The crossings signs has been assigned according to the usual right hand rule. On the right: two chiral enantiomers with opposite linking number ± 2. In both images the crossings of a curve with itself (irrelevant for Lk calculation) are indicated with red arrows. Images adapted from [4, 9]. 33
3.4 On the left: representation of an ideal knots selection. All the knots are characterized by the same axial length and it is immediate to notice that their compaction decreases with the knot complexity. These ideal configurations have been obtained by means of Monte Carlo simulations [13]. On the right: a table summarizing some of the ideal conformations main properties, like axis length to tube diameter ratio, writhe and ACN. Image and table adapted from [13]. 34
3.5 A DNA duplex whose helical axis geometry is the one of an ideal left-handed trefoil 3_1^L. Image adapted from [33]. 35

4.1 On the left: one-dimensional gel electrophoresis with nicked bacterial plasmids of 4.4 kbp. Before the reaction with a molar excess of topo II, only unknotted (N) and linear (L) molecules are present (lane 1). After the reaction (lane 2), a knots ladder (Kn) appears. As can be evinced from the bands thickness and brightness, the population of a certain knot type decreases with its complexity. Moreover, since a low voltage is applied, the distance in the gel is proportional to the knot complexity and compaction. On the right: high resolution two-dimensional electrophoresis on the same samples. A higher voltage is applied in the left-to-right direction in order to retard the knot species compared to the linear fragments. The crossing number of each knot type is reported on the right. Image adapted from [28]. 38
4.2 Effect of the applied electric field on the electrophoretic migration of DNA samples obtained by bacteriophage P4 capsid. Lane 1: sample with no additional treatment after the extraction. Lane 2: sample heated to 75°C. Lane 3: sample treated with nicking endonuclease. Lane 4: sample treated with topoisomerases II. The one-dimensional electrophoresis on the left has been performed at low voltage (25V for 20h), whereas the one on the right at high voltage (100V for 1h). From a comparison of the two cases it emerges that, despite their greater compaction, knotted configurations can migrate more slowly than linear fragments when the applied voltage is high. Image adapted from [26]. 39

4.3 On the left: high resolution two-dimensional electrophoresis at low-high voltage with yeast minichromosomes YRp4 (4.4 kbp) after their reaction with topo I (lane 1) and topo II (lane 2). The bands disappearance in the lane 2 confirmed that the ladder was populated by knot species (labeled with their crossing number). Central plot: knotting probability vs DNA length profiles relative to trefoils (green), more complex knots with crossing number > 3 (red) and their sum (black). As the chromatin length increases, the curves saturate to a stationary value. On the right: knot probability for various knot species. A colors legend is reported to distinguish the minichromosomes types. Image adapted from [28]. 40
4.4 Trefoils doublets separation with one-dimensional gel electrophoresis. On the left: knotted supercoiled configuration of a 5.6 kbp DNA plasmid obtained by cyclization in diverse reaction conditions (lanes 1 and 2) or after a treatment with molar excess of topo II (lane 3). Image from [22]. On the right: trefoil enantiomers separation obtained with YRp4 plasmid (4.4 kbp). Image from [27]. 41

5.1 Closed DNA molecule modeled as an elastic rod made by $N = 150$ cylindrical segments. 42
5.2 Euler transformation bringing the coordinate system xyz into coincidence with XYZ. The transformation can be decomposed into three elementary steps. First, xyz is rotated by α around the z axis. Then, the new coordinate system is rotated by β around $N = z \wedge Z$, the common perpendicular to z and Z. Finally, the obtained system is rotated by γ around Z. See 'Other sources' for the image reference. 45
5.3 Supercoiled unknotted configuration with three clearly visible plectonemic structures. 50
5.4 Gyration radius autocorrelation functions $\hat{C}_{Rg}(t)$ of knotted/unknotted supercoiled/nicked configurations. A unit in the lag time axis corresponds to 10^7 Monte Carlo steps. As the lag time increases, oscillations become wider due to the reduced number of elements in the sum 5.44. 51
5.5 Comparison between the tangent-tangent correlation $G(s)$ obtained with our model (green marker) and the one predicted by the analytical expression 5.45 (blue curve), for a DNA two-dimensional closed ring with flexibility ratio $\chi = 4$. The curvilinear coordinate s is expressed in simulation units, obtained dividing every length measure by $b_0 = 10$ nm. 52
5.6 $\boldsymbol{f}$-vectors correlation for different values of the torsional stiffness parameter κ_t. The characteristic length of the initial decay is related to the torsional persistence length. The curvilinear coordinate is expressed in simulation unit. 53
5.7 Relation between torsional persistence length $l_{p,t}$ and torsional stiffness parameter κ_t. The persistence length is expressed in simulation unit. 54

6.1 High resolution two-dimensional electrophoretic characterization of 4.4 kbp bacterial plasmids *in vitro*. On the left: knots separation on the basis of their topological complexity. On the right: bands splitting into doublets due to the different migration velocity of supercoiled trefoil enantiomers. Images from [28, 27]. 56
6.2 Initialization of supercoiled unknot and right handed trefoil. The $\{\boldsymbol{f}_i\}_{i=1\dots N}$, represented in blue, give the idea of how much the molecule has been twisted to obtained the desired supercoiling level. 57

6.3 Gyration radius time series relative to supercoiled (red) and nicked (green) 0_1 with some snapshots taken at the Monte Carlo time steps corresponding to the arrows positions. . . . 58
6.4 Gyration radius time series relative to supercoiled 3_1^R with some snapshots taken at the Monte Carlo time steps (in unit of dump interval) corresponding to the arrows positions. . . . 59
6.5 Gyration radius time series relative to supercoiled 3_1^L with some snapshots taken at the Monte Carlo time steps (in unit of dump interval) corresponding to the arrows positions. . . . 59
6.6 Gyration radius time series relative to nicked 3_1^R with some snapshots taken at the Monte Carlo time steps (in unit of dump interval) corresponding to the arrows positions. . . . 60
6.7 Gyration radius and asphericity distributions of nicked unknot (green) and right-handed trefoil (blue). Two typical configurations give an idea of why the knot constraint prevents the trefoil to elongate as the unknot. . . . 60
6.8 Writhe and average crossing number distributions of nicked unknot (green) and right-handed trefoil (blue). . . . 61
6.9 Gyration radius and asphericity distributions of nicked unknot (green) and supercoiled unknot (blue). . . . 62
6.10 Writhe and average crossing number distributions of nicked unknot (green) and supercoiled unknot (blue). . . . 62
6.11 Gyration radius and asphericity distributions of nicked trefoil (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange). . . . 63
6.12 Writhe and average crossing number distributions of nicked trefoil (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange). . . . 63
6.13 Gyration radius, asphericity, writhe and average crossing number distributions of supercoiled unknot (green), supercoiled right handed trefoil (blue) and supercoiled left handed trefoil (orange). . . . 64
6.14 On the left: sketch of the observation directions which limit the solid angle needed to compute the Gauss integral Ω_{ij}. On the right: the quadrangle whose area gives $\tilde{\Omega}_{ij}$. Image from [14]. . . . 66

Bibliography

[1] C.C. Adams. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. W.H. Freeman, 1994.

[2] Stuart Allison, Robert Austin, and Mike Hogan. Bending and twisting dynamics of short linear dnas. analysis of the triplet anisotropy decay of a 209 base pair fragment by brownian simulation. *The Journal of Chemical Physics*, 90(7):3843–3854, 1989.

[3] G Michael Blackburn, Michael J Gait, David Loakes, Martin Williams, David M Egli, Andy Flavell, Anna Marie Pyle, W David Wilson, S Ihtshamul Haq, Ben Luisi, Julie Fisher, Charlie Laughton, Stephanie Allen, and Joachim Engels. *Nucleic Acids in Chemistry and Biology*. The Royal Society of Chemistry, 2006.

[4] Hans Boden and Cynthia Curtis. The sl(2, c) casson invariant for knots and the Â-polynomial. *Canadian Journal of Mathematics*, 68:1–25, 05 2015.

[5] Y. Burnier, Julien Dorier, and A. Stasiak. Dna supercoiling inhibits dna knotting. *Nucleic Acids Research*, 36:4956 – 4963, 2008.

[6] Giuseppe Chirico and Jörg Langowski. Kinetics of dna supercoiling studied by brownian dynamics simulation. *Biopolymers*, 34(3):415–433, 1994.

[7] G. Călugăreanu. L'integrale de gauss et i'analyse des noeuds tridimensionnels. *Rev. Math. Pures Appl.*, (1):5–20, 1959.

[8] G. Călugăreanu. Sur les classes d'isotopie des noeuds tridimensionnels et leurs invariants. *Czechoslovak Mathematical Journal*, 11(4):588–625, 1961.

[9] Michael S. Waterman Eric S. Lander. *Calculating the Secrets of Life: Contributions of the Mathematical Sciences to Molecular Biology.* The National Academies Press., 1995.

[10] Jim Hoste et al. The first 1,701,936 knots, 1998.

[11] FB Fuller. The writhing number of a space curve. *Proceedings of the National Academy of Sciences of the United States of America*, 68(4):815—819, April 1971.

[12] S Ishii, T Murakami, and K Shishido. Gyrase inhibitors increase the content of knotted dna species of plasmid pbr322 in escherichia coli. *Journal of Bacteriology*, 173(17):5551–5553, 1991.

[13] Vsevolod Katritch, Jan Bednar, Didier Michoud, Robert G. Scharein, Jacques Dubochet, and Andrzej Stasiak. Geometry and physics of knots. *Nature*, 384(6605):142–145, November 1996.

[14] Konstantin Klenin and Jörg Langowski. Computation of writhe in modeling of supercoiled dna. *Biopolymers*, 54:307 – 317, 10 2000.

[15] E. Z. Lynn and Nancy J. Crisona. *Coating DNA with RecA Protein to Distinguish DNA Path by Electron Microscopy*, pages 99–107. Humana Press, Totowa, NJ, 1999.

[16] Cristian Micheletti, Davide Marenduzzo, and Enzo Orlandini. Polymers with spatial or topological constraints: Theoretical and computational results. *Physics Reports*, 504(1):1 – 73, 2011.

[17] Donald Olins and Ada Olins. Chromatin history: Our view from the bridge. *Nature reviews. Molecular cell biology*, 4:809–14, 10 2003.

[18] Wilma Olson, David Swigon, and Bernard Coleman. Implications of the dependence of the elastic properties of dna on nucleotide sequence. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 362:1403–22, 08 2004.

[19] Huntington F Willard Robert L.Nussbaum, Roderick R.McInnes. *Thompson and Thompson, Genetics in Medicine*. Elsevier Health Sciences, 7 edition, 2007.

[20] Joaquim Roca. Topoisomerase ii: a fitted mechanism for the chromatin landscape. *Nucleic Acids Research*, 37:721 – 730, 2009.

[21] T. Sakaue, G. Witz, G. Dietler, and H. Wada. Universal bond correlation function for two-dimensional polymer rings. *EPL (Europhysics Letters)*, 91(6):68002, sep 2010.

[22] Stanley Y. Shaw and James C. Wang. Chirality of dna trefoils: Implications in intramolecular synapsis of distant dna segments. *Proceedings of the National Academy of Sciences*, 94(5):1692—1697, March 1997.

[23] Kazuo Shishido, Naoki Komiyama, and Shukuko Ikawa. Increased production of a knotted form of plasmid pbr322 dna in escherichia coli dna topoisomerase mutants. *Journal of Molecular Biology*, 195(1):215 – 218, 1987.

[24] Andrzej Stasiak, Vsevolod Katritch, Jan Bednar, Didier Michoud, and Jacques Dubochet. Electrophoretic mobility of dna knots. *Nature*, 384:122–122, 1996.

[25] Doros N. Theodorou and Ulrich W. Suter. Shape of unperturbed linear polymers: polypropylene. *Macromolecules*, 18(6):1206–1214, 1985.

[26] Sonia Trigueros, Javier Arsuaga, Mariel Vazquez, De Witt Sumners, and Joaquim Roca. Novel display of knotted dna molecules by two-dimensional gel electrophoresis. *Nucleic acids research*, 29:E67–7, 08 2001.

[27] Antonio Valdés, Belén Martínez-García, Joana Segura, Sílvia Dyson, Ofelia Díaz-Ingelmo, and Joaquim Roca. Quantitative disclosure of DNA knot chirality by high-resolution 2D-gel electrophoresis. *Nucleic Acids Research*, 47(5):e29–e29, 01 2019.

[28] Antonio Valdés, Joana Segura, Sílvia Dyson, Belén Martínez-García, and Joaquim Roca. DNA knots occur in intracellular chromatin. *Nucleic Acids Research*, 46(2):650–660, 11 2017.

[29] James D Watson and Francis H Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *The American journal of psychiatry*, 160(4):623—624, April 2003.

[30] James H. White. Self-linking and the gauss integral in higher dimensions. *American Journal of Mathematics*, 91(3):693–728, 1969.

[31] JH White and WR Bauer. Calculation of the twist and the writhe for representative models of dna. *Journal of molecular biology*, 189(2):329—341, May 1986.

[32] Guillaume Witz. Dna topology: A biophysical study of dna rings, knots, and catenanes by means of atomic force microscopy and numerical simulations. 01 2010.

[33] Guillaume Witz and Andrzej Stasiak. Dna supercoiling and its role in dna decatenation and unknotting. *Nucleic acids research*, 38:2119–33, 12 2009.

Other sources

Figure 1.10: Weisstein, Eric W. "Granny Knot." From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/GrannyKnot.html

Figure 1.14: Dror Bar-Natan, Scott Morrison, The Knot Atlas, http://katlas.org

Figure 2.3: By Madprime - Own work. Image created with Inkscape., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1848174

Figure 2.8: https://www.quora.com/What-is-the-function-of-topoisomerase

Figure 2.10: http://www.bio.miami.edu/dana/250/250SS10_7.html

Figure 2.11: Philpot Education, Biology support site, Essential idea 2.1.7 https://www.philpoteducation.com/mod/book/view.php?id=782&chapterid=1129

Figure 2.12: By LadyofHats - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=4889777

Figure 5.2: Euler angles, Lionel Brits, https://commons.wikimedia.org/wiki/File:Eulerangles.svg, 2008 (accessed September 28, 2020), licensed by CC BY 3.0

Acknowledgements

First of all I would like to express my gratitude to Cristian for offering me the opportunity to work alongside him in this project. His kindness, his unfailing availability, his meticulous attention to detail make him the supervisor I hoped to meet. His guidance and advice have been invaluable to my scientific and personal growth.
I would like to thank all my class mates of Physics of Complex Systems for giving me the opportunity to be part of a wonderful group, for giving me the strength to face this path and for encouraging me in the most difficult moments of these two unforgettable years that I will always carry in my heart.
Thanks to all the friends who have supported me over the years. Thanks to Kevin, Jessica, Luca, Marcello, Giuliana who have been always close to me despite my recent nomadism and for having filled my years in Turin with sincere joys.
Thanks to Stefano and Antonio for their brotherhood, the clarity of their soul, their unique ability to bring me closer to myself.
Thanks to Antonio Francesco, Martina, Gavino, Franca and Tonino for being my second family, for their advice and constant support.
A huge thank you to Carolina for always being by my side, for the incomparable strength she gives me, for her inexhaustible patience, for growing with me day after day.
Finally, thanks to all my family who have always supported me, and especially to my parents Marco and Angelica for their unconditional love, for their sacrifices, for encouraging me to face every difficulty. I owe everything to them.