DNA mechanics inside plectonemes, nucleosomes and chromatin fibers

PROEFSCHRIFT

ter verkrijging van de graad
van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus
prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 2 oktober 2013
klokke 16:15 uur

door

Giovanni Lanzani

geboren te Padua, Italië in 1984

Promotiecommissie

Promotor:	Prof. dr. H. Schiessel
Overige leden:	Prof. dr. J.-M. Victor *Université Pierre et Marie Curie, Paris*
	Prof. dr. G. Wuite *Vrije Universiteit Amsterdam*
	Dr. V. Vitelli
	Prof. dr. E.R. Eliel
	Prof. dr. C.W.J. Beenakker

Casimir PhD series, Delft-Leiden 2013-7

This work is part of the research program 'DNA in action: Physics of the genome' of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

Typeset in LaTeX, written in `VIM`, version controlled with `git` and printed by Proefschriftmaken.nl || Uitgeverij BOXPress.
The word cloud on the cover was partially realized thanks to an amazing tool by Jonathan Feinberg, `wordle` (`http://www.wordle.net`).

Contents

Introduction

> All general laws are imprecise and all precise laws are banal.
>
> It's much better to have a complicated Hamiltonian with a simple domain than a simple Hamiltonian in a complicated domain.
>
> I'll leave a small hole in the proof, but a finite one. For those of you not following the course about logic, it means that it can be filled with a countable amount of steps.
>
> If I say $1 - 1 < 2$, that's true, but not really optimal.
>
> Giancarlo Benettin

Living in a natural world means we are surrounded by things that are there regardless of our presence. We call these things "nature". A marked difference between humans and animals is whether we accept this "nature" as it is. Many animals build their homes by altering "nature", but no species knows which laws make it possible for the home not to fall apart. Humans are different. If I went to physics, it was because the "how" was far more interesting than any other question. Someone may object why I didn't became an engineer then. There are two reasons for that. The first is the one that I gave to my wife one of the first time I met her. That

happened when I was a freshman in physics, hence the slang:

> Because, you see, every time a law of nature is confirmed by experiments, I feel like the harmony of the universe is preserved. And this is, you know, cool.

I was probably referring to general relativity, a theory criticized by many, if not most, physicists back then, proven to be correct by experiments. Years later, a dear friend of mine, described the second reason much better than I could possibly do, so the next words are his

> In the modern world, the beauty and essence of physics tend to be assigned to the endeavor of finding a single, simplest, and unifying principle describing the root of everything we can observe around us.[1]

With such a premise, the reader should be surprised this thesis is about biophysics, a lesser "physics" when compared to string theory and cosmology, where the above principles are felt more strongly. To understand why that is, we need a detour. We'll have many throughout the thesis, but like every other tour, the intention is to have a good time while we're at it.

The everyday operations taking place in our body strikingly resemble the activity of a public library. Without much thinking we read books so full with words that, when aligned on a single line of text, would easily cover the distance between our home and our workplace. But instead of jogging while reading, people were smart enough to condense text in lines, lines in pages, pages in books and books in library shelves.

Books remain thus widely accessible and easily readable while being compact. But what part of our organism has a similar behaviour? It turns out that DNA, the molecule contained in the nucleus of each human cell and carrying our genetic information, is also stored in an extremely compact fashion. I am shorter (but thicker!) than the total DNA contained in every one of my cells. No wonder then that nature had to find a clever

[1] This is probably the line that separates physicists from engineers: *the single, simplest and unifying principle*.

way to compact DNA so that it fits inside our cells' nuclei (which are approximately one millionth of a meter wide). And things are more complicated than with written words, because DNA is a semi-flexible, negatively charged polymer, so it does not like to be bent, twisted or packed together.

A first hint of universality, the *unifying principle*, is already there. *Polymers* are ubiquitous around us: DNA, proteins, cellulose, PVC and many more. While they have the most disparate usages, their behavior is universally described by simple laws (see section 1.1). For example the entropy makes a polymer behave like a spring, even though the two objects have nothing else in common.

A second hint will come only later (in section 1.2): most of the shapes that DNA will be assumed to have are derived by looking at the motion of a *pendulum*. This fascinating analogy was first noted by Gustav Kirchhoff in 1859. The German physicist was not thinking at polymers though, but at elastic lines, or *elastica*. To understand the elastica we can use mechanical equilibrium, variational calculus and elliptic integrals. Moreover, besides the pendulum, it is analogous to a sheet holding a volume of water and the surface of a capillary [38].

Studying DNA then is not as narrow as it may seems. DNA as a polymer, or as an elastica, means that we can re-use centuries old results to study a relative newcomer in physics textbooks, without losing any generality. And without knowing, or liking, anything about biology or chemistry.[2]

But let me present in more details what the thesis tries to accomplish. With the aim of better understanding the compaction and de-compaction of DNA we will first, in chapter 1, introduce the reader to the basic physics behind DNA. Then, in chapter 2 we present the driving forces in the equilibrium of the chromatin fiber. The chromatin fiber is a cylinder that reduces the space needed to store the genetic code. However, since we still need to access the genetic code "trapped" in the chromatin, we will look at how to (transiently) unwrap DNA from the nucleosomes, chromatin's core constituents, in chapter 3. Last, but not least, we will look at the effects of torque and tension on naked DNA, in chapter 4.

[2] I can hear the sighs of relief from here.

Chapter 1

In which we lay the foundations for the rest of the thesis

> This is the course of Mathematical Physics, where physical problems are treated in a mathematical way, thus rigorous. This could cause pleasure or pain, depending on individual inclinations.
>
> Francesco Fassò

DNA is one of those objects that, in recent times, has become a buzzword, i.e. a word used outside its original contest often in an inaccurate manner and inappropriately.

To clean every bit of confusion out: plainly said, DNA, or deoxyribonucleic acid, is a molecule carrying the necessary information to produce proteins. Proteins, in turn, are the fundamental bricks that constitutes our body, along with water and, if your partner happens to be a marvelous cook, fats (alas!).

Since proteins come in a great variety, the quantity of DNA contained in our body shouldn't surprise anyone. With the help of four so-called nucleotides (bp), *adenine* (A), *guanine* (G), *cytosine* (C) and *thymine* (T), or ATCG, which are always paired together into base pairs (A with G and C with T) forming a double helix, DNA stores the genetic information.

To produce all the proteins present in our body, the base pairs are read in groups of three, giving 64 possible combinations. These sequences are then translated into 20 amino acids, proteins building blocks; for the mathematically more inclined reader, we note that the function translating between ATCG tuples and amino acids is surjective but not injective.

There is also another occasion when the genetic code is read from DNA, *cell replication*: the *daughter cell* needs to be identical to the *mother cell*. While replicating thus, the whole DNA molecule has to be read and a new copy is assembled in place! If that does not seem remarkable, think at the numbers involved: a two meters long molecule is being read and a copy, also two meters long at the end, is created in a portion of the cell whose diameter is about 5 millionth of a meter.

If you are still unimpressed (at this point it's safe to assume that you're a mathematician), this is the moment to tell you that DNA is negatively charged and does not like to be bent, since it is

1.1 A semiflexible polymer?

When I first heard physicists were studying DNA I immediately thought at how experimentalists were having fun in their labs, trying to manipulate our genetic code to make us live forever. I could not imagine how wrong I was: not only experimentalists were not having fun nor trying to live forever, haunted by immortal in-laws, but none other than theoretical physicists were busy day and night to catch the secrets of that small molecule, so simple in its components, but so complicated when in action. A DNA molecule behaves in fact as a *polymer* [23].

A polymer is an object composed of thousands (or more) of identical or similar units, called *monomers*. The monomers are connected to each other through flexible bonds. The thousand of monomers implies that a huge number of configurations are possible, each with approximately the same energy, regardless of the specific kind of bond. The number of configurations hints at a dominance of the entropy in the polymer behavior. This, together with the independence from the specific kind of bond, means that any reasonable model can describe the polymer on

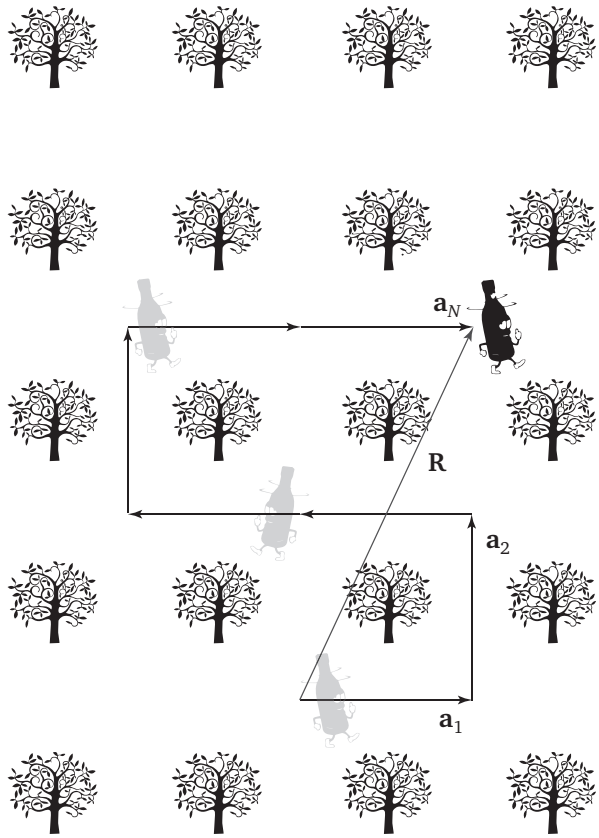


Figure 1.1: A drunk wanderer in a Dutch wood, i.e. a wood of equally spaced and perfectly equal trees. Without an external force, there is a high probability that the wanderer will walk randomly in the wood.

length scales much larger than the monomers' dimensions. The simplest model to describe a polymer is

The random walk, or the drunk wanderer

We describe the polymer as a sequence of monomers following a random walk (RW) on a periodic square lattice. The situation is analogous to a drunk wanderer in a Dutch wood, as depicted in figure 1.1. Its end-to-

end vector is

$$\mathbf{R} = \sum_{i=1}^{N} \mathbf{a}_i = b \sum_{i=1}^{N} \hat{a}_i \tag{1.1}$$

where N is the number of bonds of length b and $\hat{a}_i$ is their direction, in this case either $(\pm 1, 0)$ or $(0, \pm 1)$. The randomness of the walk implies

$$\begin{aligned} \langle \mathbf{R} \rangle &= 0 \\ \langle \mathbf{R}^2 \rangle &= b^2 \left\langle \sum_{i,j=1}^{N} \hat{a}_i \cdot \hat{a}_j \right\rangle = b^2 \left(\left\langle \sum_{i=1}^{N} \hat{a}_i^2 \right\rangle + \left\langle \sum_{i \neq j}^{N} \hat{a}_i \cdot \hat{a}_j \right\rangle \right) \\ &= b^2 \left\langle \sum_{i=1}^{N} \hat{a}_i^2 \right\rangle = b^2 N. \end{aligned} \tag{1.2}$$

At first, we would think that applying a force f would change the end-to-end vector to $\mathbf{R} = bN\hat{f}$, i.e. a completely stretched polymer. However, thinking at the drunkard analogy, it seems difficult that all his missteps would disappear if we try to enforce a direction on him. Some detours will still be present, even though with a different result than before. If C_2H_5OH is the reason behind the drunkard resistance to force, entropy is behind the polymer behavior.

To prove it, consider the probability for a RW to have an end-to-end vector equal to $\mathbf{R} = (x, y, z)^T$. If we denote the total number of RW's by $M \geq 1$, the probability is given by how many RW's end at $\mathbf{R}$, divided by M. By the central limit theorem, stating that a sufficiently large number of independent random variables is properly approximated by a Gaussian distribution, the probability can be written as

$$\begin{aligned} p(\mathbf{R}) &\simeq \text{const.}\, N^{-1/2} e^{-\frac{x^2}{2\langle x^2 \rangle}} N^{-1/2} e^{-\frac{y^2}{2\langle y^2 \rangle}} N^{-1/2} e^{-\frac{z^2}{2\langle z^2 \rangle}} \\ &= \text{const.}\, N^{-3/2} e^{-\frac{3R^2}{2b^2 N}}, \end{aligned} \tag{1.3}$$

where we used $\langle x^2\rangle = \langle y^2\rangle = \langle z^2\rangle = b^2N/3$. The entropy is then given by the Boltzmann relation $S(\mathbf{R}) = k_B \ln p(\mathbf{R})M$, from which the free energy follows:

$$\begin{aligned} S(\mathbf{R}) &= S_0 - \frac{3k_B}{2b^2N}R^2 \\ F(\mathbf{R}) &= E - TS(\mathbf{R}) = F_0 + \frac{3k_BT}{2b^2N}R^2. \end{aligned} \tag{1.4}$$

The free energy of a RW has the same form of *Hooke*'s law, i.e. it describes the small deformation of an elastic spring. For example applying a force in the $\hat{x}$-direction gives the end-to-end distance along $\hat{x}$ through

$$f = \frac{\mathrm{d}F(\mathbf{R})}{\mathrm{d}x} = \frac{3k_BT}{b^2N}x = K(T)x \tag{1.5}$$

$$\Longrightarrow$$

$$x = \frac{fb^2N}{3k_BT} \tag{1.6}$$

where $K(T)$ is the temperature-dependent *entropic spring* constant of the chain.

Equation (1.5) might seems artificial since it gives results for $x > Nb$ (the maximum extension the polymer reaches before breaking) and for values of x not belonging to the lattice. Moreover requiring a drunkard to wander on a grid is quite ambitious. To solve these limitations we consider the *freely jointed chain*, i.e. a chain with completely flexible joints. Formally the chain is defined by $\{\mathbf{R}_i\}$, $i \in 1, \dots, N$, $\mathbf{R}_i = b\hat{R}_i$ with $\hat{R}_i$ a random vector on the unit sphere (if we are considering a three dimensional chain). In this case $\langle\mathbf{R}\rangle = 0$ and $\langle\mathbf{R}^2\rangle = b^2N$ as in equation (1.2) hinting at a universal behaviour for polymers. Applying a force f along the $\hat{z}$-direction gives the Hamiltonian

$$H = -\sum_{i=1}^{N} bf\cos\vartheta_i \tag{1.7}$$

where ϑ_i is the angle between $\mathbf{R}_i$ and $\hat{z}$.

The partition function Z follows

$$\begin{aligned} Z &= \int_0^{2\pi} \mathrm{d}\varphi_1 \ldots \mathrm{d}\varphi_N \int_0^{\pi} \mathrm{d}\vartheta_1 \sin\vartheta_1 \ldots \mathrm{d}\vartheta_N \sin\vartheta_N e^{-\beta H} \\ &= (2\pi)^N \int_{-1}^{1} \mathrm{d}\cos\vartheta_1 \ldots \mathrm{d}\cos\vartheta_N e^{\beta b f \sum_i \cos\vartheta_i} \\ &= \left(\frac{4\pi}{\beta b f}\right)^N \sinh^N \beta b f. \end{aligned} \tag{1.8}$$

The equivalent of equation (1.6) then is

$$\begin{aligned} \langle z\rangle &= \left\langle \sum_i b\cos\vartheta_i \right\rangle = \frac{1}{\beta}\left(\frac{1}{Z}\frac{\partial Z}{\partial f}\right) = \frac{1}{\beta}\frac{\partial}{\partial f}\ln Z \\ &= bN\left(\coth\beta b f - \frac{1}{\beta b f}\right) \simeq \begin{cases} \frac{b^2 N}{3k_B T} f & \text{for } \beta b f \ll 1 \\ bN - \frac{N}{\beta f} & \text{for } \beta b f \gg 1. \end{cases} \end{aligned}$$

The paradoxes of equation (1.6) are now gone, as $\langle z\rangle < bN$, even for large forces, and continuous values of z are now possible as the polymer is not restricted by a lattice.

Another interesting case is the freely rotating chain model, defined by $\{\mathbf{R}_i\}$ where each $\mathbf{R}_i$ forms a fixed angle ϑ with $\mathbf{R}_{i-1}$, as depicted in figure 1.2, i.e. a vector should lie on the surface of a cone centered on the previous vector. This requirement implies

$$\langle \mathbf{R}_i \cdot \mathbf{R}_{i+1}\rangle = b^2 \cos\vartheta,$$

i.e. the vector will, on average, be exactly in the centrum of the cone, whose height is $b\cos\vartheta$. For more consecutive vectors, by induction we have

$$\langle \mathbf{R}_i \cdot \mathbf{R}_{i+2}\rangle = b^2 \cos^2\vartheta;$$

$$\langle \mathbf{R}_i \cdot \mathbf{R}_{i+j}\rangle = b^2 \cos^j\vartheta. \tag{1.9}$$

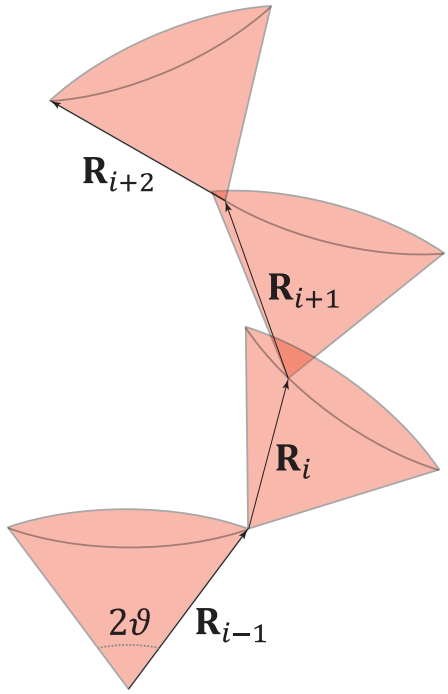


Figure 1.2: A section of the freely rotating chain. Each **R** lies on the surface of its own cone.

This is enough to compute the equivalent of equation (1.2)

$$\begin{aligned}\langle \mathbf{R}^2 \rangle &= \sum_{i,k=1}^{N} \langle \mathbf{R}_i \mathbf{R}_k \rangle = \sum_{i=1}^{N} \sum_{j=-i+1}^{N-i} \langle \mathbf{R}_i \mathbf{R}_{i+j} \rangle \\ &\approx \sum_{i=1}^{N} \sum_{j=-\infty}^{\infty} \langle \mathbf{R}_i \mathbf{R}_{i+j} \rangle = b^2 \sum_{i}^{N} \left(1 + 2 \sum_{j=1}^{\infty} \cos^j \vartheta \right) \\ &= b^2 N \left(-1 + 2 \sum_{j=0}^{\infty} \cos^j \vartheta \right) = \frac{1+\cos\vartheta}{1-\cos\vartheta} b^2 N \equiv b_{\text{eff}}^2 N.\end{aligned}$$

The approximation used is acceptable because, thanks to equation (1.9), the correlation decays exponentially for large j's. Comparing the mean-

squared end-to-end distance of the three models presented we see a common scaling behaviour, i.e. $\langle \mathbf{R}^2 \rangle \sim N$ (with a pre-factor depending on the details of the model). As promised in the introduction, the knowledge of the specific polymers' chemistry is not needed to understand it behaviour: its mean-squared end-to-end distance always scales with N.

What about DNA?

Up to now we treated monomers as points, without volume. Real polymers, however, have a finite volume. This finiteness forbids the presence of two monomers at the same place (at the same time). This is an effect of the *excluded volume interactions*. As a consequence the mean-squared end-to-end distance increases changing from $\sqrt{\langle \mathbf{R}^2 \rangle} \sim N^{1/2}$ to $\sqrt{\langle \mathbf{R}^2 \rangle} \sim N^{3/5}$.

To derive the new scaling behaviour a variation to the RW model is used, the self-avoiding walk (SAW). While similar to a RW, a SAW is more difficult to solve, because the excluded volume interactions are long ranged: pieces of the polymer separated by many monomers could still overlap in a RW, and therefore need to be kept apart in a SAW.

Although DNA is a real polymer, for the length scales considered in this thesis we can safely ignore excluded volume effects. In fact DNA is half way in between a completely flexible polymer, for which we expect strong excluded volume effects, and a stiff rod, difficult to bend and for which excluded volume is only relevant over very long distances. Such a polymer is called a *semi-flexible polymer* and is studied with the *worm-like chain* framework (WLC).[1] The WLC model can describe semi-flexible polymers and, using a coarse grained approximation, also long strands of DNA where the particular sequence of base pairs (bp) is ignored.

To see what is the threshold between the flexible and the stiff regimes for a DNA molecule, we consider the energy needed to bend it. Within the WLC model, the curvature $\kappa(s)$ is used to quantify the bending energy. Here $0 \leq s \leq L$ is the arc length of the polymer with countour

[1] The framework was first introduced in 1949 by Kratky and Porod [32].

length L. More specifically

$$E_p = \frac{A}{2}\int_0^L \kappa^2(s)\mathrm{d}s \tag{1.10}$$

where A is the *bending modulus* whose value ($\approx 50\,\mathrm{nm}\,k_BT$) is experimentally determined by measuring the energy needed to deform a portion of DNA from the straight state to another state, with a well defined $\kappa(s)$ (easy when $\kappa(s)$ is constant). The curvature $\kappa(s)$ that minimizes the energy is given, through the Euler-Lagrangian equations, by $\dot{\kappa}(s) = 0$ i.e. $\kappa(s) = m/L$, m constant; the resulting energy is $E_p = Am^2/2L$; including thermal fluctuations the equipartition theorem yields $\langle E\rangle = k_BT/2$ so that $\langle m^2\rangle = Lk_BT/A$. Considering the orientation of the polymer between s and $s + l$ we can write

$$\begin{aligned}\langle \mathbf{t}(s)\cdot\mathbf{t}(s+l)\rangle &= \langle \cos\kappa(s)l\rangle = \langle \cos m^2\rangle \approx 1 - \frac{1}{2}\langle m^2\rangle \\ &= 1 - \frac{l}{2}\frac{k_BT}{A}\end{aligned}$$

where $\mathbf{t}(s)$ represent the tangent of the polymer at s. With the same reasoning between s and $s + 2l$ using the independence of the bending between s and $s + l$, and between $s + l$ and $s + 2l$

$$\begin{aligned}\langle \mathbf{t}(s)\cdot\mathbf{t}(s+2l)\rangle &= \langle \cos(2\kappa l)\rangle \\ &= \langle \cos\kappa l\rangle^2 - \langle \sin\kappa l\rangle^2 \\ &= (1 - \frac{l}{2}\frac{k_BT}{A})^2 - 0.\end{aligned}$$

By induction when $nl = L$ and $n \to \infty$ we can write

$$\begin{aligned}\langle \mathbf{t}(s)\cdot\mathbf{t}(s+L)\rangle &= \lim_{n\to\infty}\left(1 - \frac{L}{n}\frac{k_BT}{2A}\right)^n \\ &= e^{-L/2l_p}\end{aligned} \tag{1.11}$$

where $l_p \equiv A/k_BT$ is the *bending persistence length*. The interpretation of l_p using equation (1.11) is that points l_p apart along the chain have uncorrelated orientation. Equation (1.11) gives the mean-squared end-to-end

distance of a DNA molecule[2]

$$\langle \mathbf{R}^2 \rangle = \left\langle \left(\int_0^L \mathbf{t}(s) \mathrm{d}s \right)^2 \right\rangle = \int_0^L \mathrm{d}s \int_0^L \mathrm{d}s' \langle \mathbf{t}(s) \cdot \mathbf{t}(s') \rangle$$
$$= \int_0^L \mathrm{d}s \int_0^L \mathrm{d}s' e^{-|s-s'|/l_p} = 2 \int_0^L \mathrm{d}s \int_0^S \mathrm{d}s' e^{-(s-s')/l_p} =$$
$$= 2l_p^2 \left(\frac{L}{l_p} + e^{-L/l_p} - 1 \right) \tag{1.12}$$
$$\approx \begin{cases} L^2 & \text{for } L \ll l_p \\ 2l_p L & \text{for } L \gg l_p. \end{cases} \tag{1.13}$$

When $L \ll l_p$ the polymer behaves as a stiff rod, where no bending takes place, while when $L \gg l_p$ we recover the ideal chain result, $\langle \mathbf{R}^2 \rangle \sim N$. We can thus describe DNA at larger length scales as a random walk with step size equal to $l_p \approx 50\,\mathrm{nm}$ (at room temperature).

Obviously at some point the excluded volume interactions will play a role, but the disproportion between length and diameter of the DNA molecule make the use of the RW justified up to $L \leq 5\,\mu\mathrm{m}$. For further details we invite the reader to buy a copy of the book about biophysics authored by my supervisor.

Besides l_p, DNA has another persistence length, the *torsional persistence length*, $l_t = C/k_B T \approx 100\,\mathrm{nm}$. Usually C is called the *torsional modulus*. As the origin of l_p lies in the bending resistance, the origin of l_t lies in the resistance to twist that DNA opposes when its twist deviates from the natural value[3] of $2\pi/10$ bp. The total energy then results in

$$H_0 = \frac{1}{2} \int_0^L \mathrm{d}s \left[A\kappa^2(s) + C \left(\frac{\mathrm{d}\eta}{\mathrm{d}s} \right)^2 \right]. \tag{1.14}$$

We stress that the twist $\mathrm{d}\eta/\mathrm{d}s$ in equation (1.14) is the twist exceeding the natural twist.

[2]We switch here silently to three dimensions. One can show that in this case the persistence length is twice as short than in two dimensions, because the chain can bend in two independent directions.

[3]We use the word "natural value" because DNA is naturally twisted when relaxed.

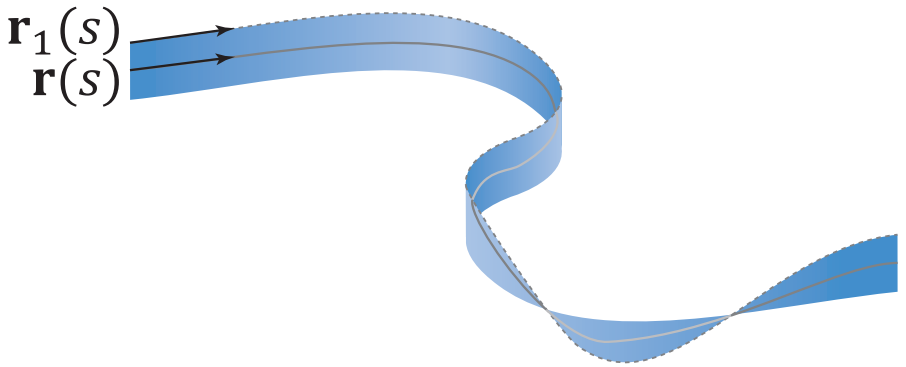


Figure 1.3: An example of a ribbon.

Twist is easier to understand when we visualize DNA as a ribbon: the centerline of DNA, $\mathbf{r}(s)$, represents the axis of the ribbon, and with one of the two strands, represented by $\mathbf{r}_1(s)$, we completely determine the ribbon geometry, see figure 1.3. The two vectors are enough to compute the curvature and the twist, and therefore the energy of the DNA chain.

1.2 The Euler angles

Starting from $\mathbf{r}(s)$ we can conveniently represent DNA through the *Euler angles*. Consider $\mathbf{t}(s) = \dot{\mathbf{r}}(s)/|\dot{\mathbf{r}}(s)|$, $\mathbf{n}(s)$, pointing towards $\mathbf{r}_1(s)$, and $\mathbf{m}(s) = \mathbf{n}(s) \times \mathbf{t}(s)$. The three vectors, $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{m}$, are, respectively, the tangent, normal and binormal of $\mathbf{r}(s)$. They form a coordinate system that moves along the chain (hence the s dependency).

Once the three vector at $s = 0$ are specified, $\mathbf{t}_0$, $\mathbf{n}_0$ and $\mathbf{m}_0$, we can define three angles, $\varphi(s)$, $\vartheta(s)$ and $\psi(s)$ such that $\mathbf{t}(s)$, $\mathbf{n}(s)$ and $\mathbf{m}(s)$ are given by a rotation of $\varphi(s)$ around $\mathbf{t}_0$ followed by a rotation of $\vartheta(s)$ about the new $\hat{n}$-axis, and finished by a rotation of $\psi(s)$ about the new $\hat{t}$-axis. In other words in terms of the rotation matrices the transformation

matrix is[4]

$$O(s) = O_{\mathbf{t}_s}(\psi_s) O_{\mathbf{n}_s}(\vartheta_s) O_{\mathbf{t}_0}(\varphi_s) \tag{1.15}$$

where $O_i(\alpha)$ represents a rotation of α radians about the $\hat{i}$-axis. These three angles are called the *Euler angles*. Through them the vector $\mathbf{t}_s$ can be expressed as

$$\mathbf{t}_s = (\sin\vartheta_s \cos\varphi_s, \sin\vartheta_s \sin\varphi_s, \cos\vartheta_s)^T \tag{1.16}$$

and the Hamiltonian in equation (1.14) takes the form

$$\begin{aligned} H_0 &= \frac{1}{2}\int_0^L \mathrm{d}s \left[A\,\mathbf{t}_s^2 + C\,(\,\mathbf{u}_s \times \dot{\mathbf{u}}_s \cdot \mathbf{t}_s)^2\right] \\ &= \frac{1}{2}\int_0^L \mathrm{d}s \left[A\left(\dot{\varphi_s}^2 \sin^2\vartheta_s + \dot{\vartheta_s}^2\right) + C\left(\dot{\varphi_s}\cos\vartheta_s + \dot{\psi_s}\right)^2\right]. \end{aligned} \tag{1.17}$$

We define here $\Delta\psi_s = \mathbf{u}_s \times \dot{\mathbf{u}}_s \cdot \mathbf{t}_s$ related to the twist of the polymer by

$$Tw = \int_0^L \mathrm{d}s \frac{\Delta\psi_s}{2\pi}. \tag{1.18}$$

Adding a force F along the $\hat{z}$-axis changes equation (1.17) to

$$H = H_0 - F\int_0^L \mathrm{d}s \cos\vartheta_s. \tag{1.19}$$

Equation (1.19) is similar to the Hamiltonian of a symmetric spinning top with a fixed point on a gravitational field. The analogy is so powerful that it is called, after its inventor, the *Kirchhoff kinetic analogy*. A complete classification of its solutions exists (see [51]).

In the Kirchhoff analogy ϑ_s is the *precession*, φ_s the *nutation* and ψ_s the *rotation* of the top.

We outline how to solve the system in the planar case ($\dot{\varphi}_s = 0$). When $\dot{\varphi}_s = 0$ the Hamiltonian is

$$H_p = \frac{1}{2}A\int_0^L \mathrm{d}s\dot{\vartheta}_s^2 - F\int_0^L \mathrm{d}s\cos\vartheta_s + \frac{C}{2}\int_0^L \dot{\psi}_s^2. \tag{1.20}$$

[4]From now on we use φ_s to indicate $\varphi(s)$ (and similarly with other symbols, when the notation does not create confusion).

The angle ψ_s has a trivial solution. Instead ϑ_s has, as Lagrangian,

$$\mathcal{L} = \frac{A}{2}\dot{\vartheta}_s^2 + F\cos\vartheta_s \tag{1.21}$$

that can be interpreted as the Lagrangian of a *pendulum*. Depending on the total energy the pendulum could be revolving or oscillating. The Lagrangian remains as in equation (1.21), but the interpretation of the parameters changes. Defining $\vartheta = 0$ to be the pendulum at rest and $\vartheta = \pi$ the upside-down pendulum, the boundary condition $\vartheta_0 = 0$ yields

$$\mathcal{L} = Ml^2\frac{\dot{\vartheta}_s^2}{2} + Mgl\cos\vartheta_s. \tag{1.22}$$

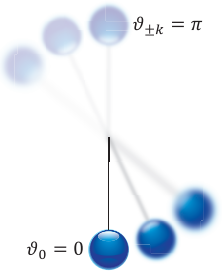


Figure 1.4: The revolving pendulum with the relevant boundary conditions.

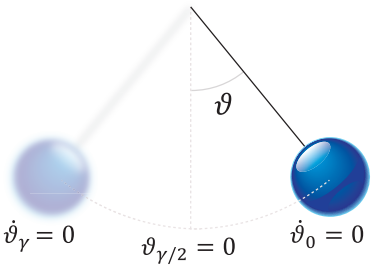


Figure 1.5: The oscillating pendulum with the relevant boundary conditions. Here γ is half the pendulum's period.

If the total energy of the system E_{tot} is bigger than $E_{\max} = 2Mgl$ (the maximum potential energy of an oscillating pendulum) then the pendulum is revolving, otherwise it will oscillate. The Lagrangian eq. (1.22) gives

$$\ddot{\vartheta}_s = -\frac{g}{l}\sin\vartheta_s \tag{1.23}$$

that can be rewritten as

$$\left(\frac{\dot{\vartheta}_s}{2}\right)^2 = \frac{g}{l}\left(m - \sin\frac{\vartheta_s}{2}\right) \tag{1.24}$$

where m is an integration constant. Multiplying by $2l^2M$ gives the kinetic energy on the LHS. When $\vartheta = 0$ the total energy is purely kinetic and equation 1.24 gives

$$\begin{aligned} E_{\text{tot}} &= 2glmM \\ &= mE_{\max} \\ &\longrightarrow \\ m = \frac{E_{\text{tot}}}{E_{\max}} &= \begin{cases} > 1 & \text{for } E_{\text{tot}} > E_{\max} \text{ (revolving)} \\ \in]0,1[& \text{for } E_{\text{tot}} < E_{\max} \text{ (oscillating).} \end{cases} \end{aligned} \tag{1.25}$$

This elegantly links m with the energy of the system, allowing for an immediate physical interpretation of the equations. From eq. (1.24), calling $g/l \equiv \lambda^{-2}$, we get

$$\begin{aligned} \left(\frac{\dot{\vartheta}_s}{2}\right) &= \sqrt{m}\lambda^{-1}\sqrt{1 - \frac{1}{m}\sin\frac{\vartheta_s}{2}} \\ \frac{\mathrm{d}\frac{\vartheta_s}{2}}{\sqrt{1 - \frac{1}{m}\sin^2\frac{\vartheta_s}{2}}} &= \mathrm{d}s\frac{\sqrt{m}}{\lambda}. \end{aligned} \tag{1.26}$$

To proceed we first make a distinction for the boundary condition in the two different cases. When the pendulum is revolving the conditions are illustrated in figure 1.4 while the oscillating one is depicted in figure 1.5. In the first case integrating from $s = 0$ to s gives

$$\int_0^{\frac{\vartheta_s}{2}} \frac{\mathrm{d}\frac{\vartheta}{2}}{\sqrt{1 - \frac{1}{m}\sin^2\frac{\vartheta_s}{2}}} = \frac{\sqrt{m}}{\lambda}s.$$

The integral results in the *elliptic function* F, whose inverse function is am, so that

$$\begin{aligned} \mathrm{F}\left(\frac{\vartheta_s}{2}\middle|\frac{1}{m}\right) &= \frac{\sqrt{m}}{\lambda}s \\ &\rightarrow \\ \frac{\vartheta_s}{2} &= \mathrm{am}\left(\frac{\sqrt{m}}{\lambda}s\middle|\frac{1}{m}\right). \end{aligned} \tag{1.27}$$

Using the *Jacobi elliptic function* sn defined as $\mathrm{sn}(x|y) \equiv \sin(\mathrm{am}(x|y))$, the well known identity $\cos x = 1 - 2\sin^2 x/2$ yields

$$\cos\vartheta_s = 1 - 2\,\mathrm{sn}^2\left(\frac{\sqrt{m}}{\lambda}s\middle|\frac{1}{\sqrt{m}}\right). \tag{1.28}$$

This is the solution for the revolving pendulum. Note that eq (1.27) gives the $\bar{s}$ at which $\vartheta_{\bar{s}} = \pi$ (upside-down pendulum) as a function of m

$$\bar{s} = \mathrm{K}\left(\frac{1}{m}\right)\frac{1}{\sqrt{m}}\lambda.$$

When the pendulum is oscillating, starting from eq. (1.26) we find

$$\cos\vartheta_s = 1 - 2m\,\mathrm{sn}^2\left(\frac{s}{\lambda}\middle|m\right) \tag{1.29}$$

where we used the equality $\mathrm{sn}(\sqrt{m}x|m^{-1}) = \sqrt{m}\,\mathrm{sn}(x|m)$.

When the tangent vector is defined as in eq. (1.16), with $\varphi_s = 0$, its path will be given by $x_s = \int \sin\vartheta_s \mathrm{d}s$, $z_s = \int \cos\vartheta_s \mathrm{d}s$. In figure 1.6 we plot the resulting shapes for different values of m. The boundary between the two cases, $m = 1$, is the homoclinic loop, which has ends aligned with the $\hat{z}$-axis (i.e. in the force direction) and is described by

$$\cos\vartheta_s = 1 - 2\,\mathrm{sech}^2\frac{s}{\lambda}. \tag{1.30}$$

An interesting aspect of paths with ends aligned with the $\hat{z}$-axis is that they are, not without some efforts [51], also solvable in the non-planar case, i.e. $\varphi_s \neq 0$. The solution, for $t \in [0, 1]$, is

$$\cos\vartheta_s = 1 - 2t^2\,\mathrm{sech}^2\frac{st}{\lambda} \tag{1.31}$$

$$\varphi_s = \arctan\left(\frac{t}{\sqrt{1-t^2}}\tanh\frac{st}{\lambda}\right) + \sqrt{1-t^2}\frac{s}{\lambda}. \tag{1.32}$$

Ignoring the C-term, irrelevant now (but it will be included later), the elastic and potential energy contributions follows from eq. (1.19) adding

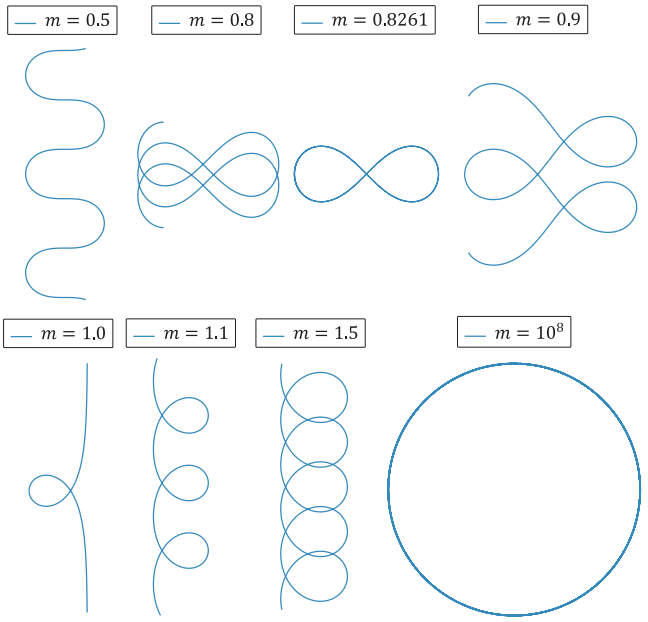


Figure 1.6: The various orbits obtained through integration of eqs. (1.28–1.30). The plot for different m are not in scale.

up to

$$E_{\text{loop}} = 2FL_{\text{loop}} \tag{1.33}$$

$$L_{\text{loop}} = \int_{-\infty}^{+\infty} \mathrm{d}s(1 - \cos\vartheta) = 4\lambda t \tag{1.34}$$

where L_{loop} is the length lost to the loop when compared to the straight chain. Fuller's equation, eq. (1.39) below gives the writhe of the path using the $\hat{z}$-axis as reference

$$Wr_{\text{loop}}(t) = \frac{2}{\pi} \arcsin t. \tag{1.35}$$

1.3 Twist and shout

To speak about writhe, we need to first address *twist* more carefully. Twist added or removed is only relevant when a polymer cannot relax the inserted turns, as in the case of DNA held with a magnetic or optical wrench. This is a typical experimental setup used to study the polymer torsional response. The bead is attached to one end of the DNA, while the other end is anchored to a surface. As the bead is turned, the polymer is over- or under-twisted. In this case the Hamiltonian of equation (1.19) becomes

$$H_\tau = H - 2\pi n \tau_F \tag{1.36}$$

where n is the number of turns inserted by the beads and τ_F is the torque in the direction of the force and acts here as a Lagrange multiplier (*number of turns clamp*).

Ignoring the natural twist of DNA ($2\pi/10$ bp) we interpret n as the *linking number*. The linking number indicates how two closed, oriented curves are linked with each other. Abbreviated with Lk, it is an integer normally fairly easy to compute for two curves that lie in the same plane, except when crossing. Take in fact the curves A and B and examine the points where they cross each other. For every crossing, use the right-hand rule with your right forefinger aligned with the direction of the curve that passes above and your long finger aligned with the direction of the other. If your thumb, stretched in a natural position, points up, then assign to that crossing a $+1$, otherwise -1. Define n_+ as the sum of all $+1$'s and n_- as the sum of all -1's. Then $Lk = n_+ + n_-$. Since a picture goes a long way, in figure 1.7 an example with two simple curves is depicted. Another way to compute the linking number, especially useful when the curves do not live on the same plane, is to use the Gauss integrals of the two closed curves, i.e.

$$Lk = \frac{1}{4\pi} \oint_A \oint_B \frac{\dot{\mathbf{a}}(s) \times \dot{\mathbf{b}}(t) \cdot (\mathbf{a}(s) - \mathbf{b}(t))}{\left|\mathbf{a}(s) - \mathbf{b}(t)\right|^3} \mathrm{d}s\mathrm{d}t. \tag{1.37}$$

Since a DNA molecule can be interpreted as a ribbon, we take the two curves $\mathbf{r}$ and $\mathbf{r}_1$ (see figure 1.3) that define the ribbon, and compute the linking number through eq (1.37). However, the energy of a polymer does

not depend on Lk, but on its twist Tw (see eq. (1.14)), experimentally difficult to measure. Luckily White [70] found a relation, now going by its name, that relates twist and linking number with the writhe Wr

$$Lk = Tw + Wr. \tag{1.38}$$

For closed curves we can compute Wr through equation (1.37) in the limit $\mathbf{b} \to \mathbf{a}$. However equation (1.38) is of little help to compute the energy, even if Lk is experimentally controlled. In fact to compute Tw (and thus the energy), we still have to use the cumbersome equation (1.37) for Wr.[5]

An alternative method is to calculate the writhe of a curve with respect to another, by using a relation provided by Fuller [25]:

$$Wr_B - Wr_A = \frac{1}{2\pi} \int \frac{\mathbf{t}_A \times \mathbf{t}_B \cdot \frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{t}_A + \mathbf{t}_B)}{1 + \mathbf{t}_A \cdot \mathbf{t}_B} \mathrm{d}s. \tag{1.39}$$

Here $\mathbf{t}_{A,B}(s)$ is the unit tangent vector for A, B. Both curves share the same parameter s, one is deformable into the other in a continuous way and the two tangents are never anti-parallel[6]. If they are antiparallel, the denominator of equation (1.39) diverges and the integral gives the correct answer $\mod 2$.

Applying eq. (1.39) to "open" DNA molecules requires attention, as the formula can only be applied when the curves are closed. However Starostin [64] showed how to "close" the polymer by connecting its ends, aligned with an axis at infinity, by using a geodesic on the tangent sphere. A common way to make use of equation (1.39) then is by taking $\mathbf{t}_A \parallel F \parallel \hat{z}$ (i.e. $Wr_A = 0$) and $\mathbf{t}_B = \mathbf{t}$; since the "closing" is the same, we only need equation (1.39) when $\mathbf{t}_A \neq \mathbf{t}_B$.

The fact that the curves are open explains why we identified Lk with n above. For a closed ribbon Lk is fixed and deforming it only changes Wr and Tw, leaving Lk unaffected. However when the ribbon (polymer in our case) is "open", if we insert n turns inside it, Tw and/or Wr will increase, making Lk equal to n (if it was 0 when $n = 0$).

[5] To clarify: equation (1.37) can be solved in the limit $\mathbf{b} \to \mathbf{a}$, but the process is long and error prone.

[6] More precisely they should be homotopic as non-intersecting space curves and the tangent of the homotopy should never be anti-parallel to one of the end curves.

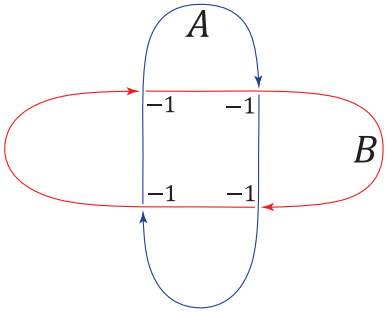


Figure 1.7: An example of two curves with $Lk = -4$.

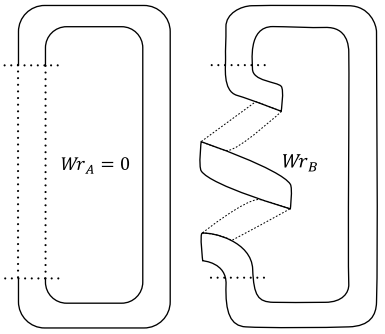


Figure 1.8: An example where Fuller equation can be fruitfully applied.

To further clarify these concepts, we can apply equation (1.39) to figure 1.8: while curve A has zero writhe, curve B has some. The common part of the two curves will not contribute to the integral of equation (1.39), since the cross product vanishes when $\mathbf{t}_A = \mathbf{t}_B$. We restrict thus the integral where $\mathbf{t}_B$ differs from $\mathbf{t}_A$. In the case of Figure 1.8 the two tangent vectors are

$$\begin{aligned} \mathbf{t}_A &= (0, 0, 1)^T \\ \mathbf{t}_B &= \frac{\partial}{\partial s}\mathbf{r}_B = \frac{\partial}{\partial s} r(\sin \pi s, -\cos \pi s, \pi s \tan \alpha)^T. \end{aligned} \tag{1.40}$$

Renormalizing $\mathbf{t}_B$ equation (1.39) gives

$$Wr_B - Wr_A = -s^*(1 - \sin \alpha) \tag{1.41}$$

where s^* is the number of helical turns ($s^* = 1$ indicates one full turn, $s^* = 2$ indicates two full turns, etc.).

We now rewrite equation (1.36) using equation (1.38)

$$H_\tau = H - 2\pi(Tw + Wr)\tau_F. \tag{1.42}$$

Here Wr does not depend on $\Delta\psi_s$: therefore using the Euler-Lagrange equations for $\Delta\psi_s$ through equation (1.18), we find that the twist rate is

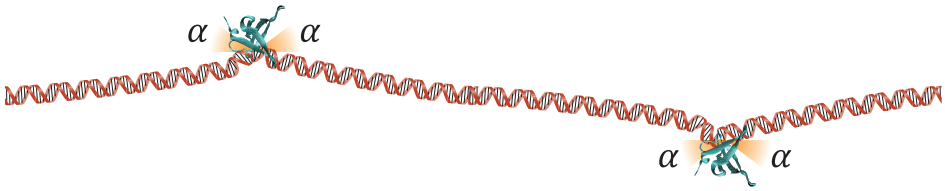


Figure 1.9: How proteins that enforce an angle on a polymer binds to DNA. The situation is analogous to the experiments performed in [18].

constant and equal to $\Delta\psi_s = \Delta\psi = 2\pi Tw/L$. This makes the torsional energy equal to

$$E_T = \frac{C}{2}\int_0^L \left(\frac{2\pi Tw}{L}\right)^2 \mathrm{d}s = \frac{2\pi^2 C}{L} Tw^2 = \frac{2\pi^2 C}{L}(n - Wr)^2 \tag{1.43}$$

1.4 DNA binding proteins.

An interesting application of the various path that DNA assumes in figure 1.6 is the following: suppose we have a DNA molecule held with a force F, with n_p proteins bound to it. Each protein enforces a kink with a certain angle α on the DNA, as in figure 1.9. If we know the difference ΔL between the contour length of the molecule and its end-to-end distance when it has n_p proteins bound, we can estimate n_p through

$$\Delta z n_p = \Delta L \tag{1.44}$$

where Δz is the length lost per protein in the force direction. The problem is interesting because knowing how n_p changes with the force is a fundamental step to understand the binding behavior of these proteins [18]. Calling l the contour length of the DNA between consecutive proteins, we use eq. (1.29) to find a Δz_e to be used in eq. (1.44). The resulting length lost per protein is

$$\Delta z_e = l - \int_{\bar{s}}^{\bar{s}+l} \cos\vartheta(s, \lambda, m)\mathrm{d}s \tag{1.45}$$

for $\bar{s}$ and m found through

$$\cos\vartheta(\bar{s}+l,\lambda,m)=\cos\vartheta(\bar{s},\lambda,m)=\cos\alpha \tag{1.46}$$

and $\cos\vartheta$ given by eq. (1.29). This approach means that we are "cutting" the paths in figure 1.6 at the points where they are orientated in the α direction. If l, however, is large enough, the DNA between two proteins will be approximately straight, and we can use the simpler eq. (1.30) instead of eq. (1.29). In that case the length lost per protein is

$$\Delta z_h = 2\int_{\bar{s}}^{\infty}(1-\cos\vartheta_h(s,\lambda))\,\mathrm{d}s. \tag{1.47}$$

with $\cos\vartheta_h$ given by eq. (1.30) and $\bar{s}$ given by

$$\cos\vartheta_h(\bar{s},\lambda)=\cos\alpha. \tag{1.48}$$

The relative difference between the two Δ's, Δz_h and Δz_e turns out to be lower than 20%, and thus acceptable, as long as $l>\lambda$. Using these ideas in ref. [18], the numbers of bound crenarchaeal chromatin proteins Cren7 and Sul7 (for which the kink angles are known from molecular dynamics simulation) were determined.

1.5 Buckling

When dealing with a water hose[7] one source of continuous stress is the coiling of the hose around itself. Pulling on it will hopefully uncoil it, but eventually the hose will coil again.

A similar coiling happens with a DNA molecule: when twisted, a certain torque will be reached, causing the coiling of the molecule. The point when a straight molecule becomes energetically unstable is the *bifurcation point*. Its value can be determined as follows. The Hamiltonian of the

[7]For the Dutch readers: a water hose is a flexible tube utilized for watering plants; widely used in countries where it does not rain every other day, it is practically unknown in the Netherlands.

system can be written as (see eqs. (1.17, 1.18, 1.19, 1.42)):

$$H = \int_0^L \mathrm{d}s \left[\frac{A}{2}\dot{\mathbf{t}}_s^2 + \frac{C}{2}\Delta\psi_s^2 - \mathbf{F}\cdot\mathbf{t}_s\right] - 2\pi\left(Wr(\mathbf{t}_s) + \int_0^L \mathrm{d}s\frac{\Delta\psi_s}{2\pi}\right)\tau_F. \tag{1.49}$$

If we choose $\mathbf{F} \parallel \hat{x}$ and

$$\mathbf{t}_s = (\cos\varphi_s\cos\vartheta_s, \sin\varphi_s\cos\vartheta_s, \sin\vartheta_s)^T \tag{1.50}$$

then eq. (1.49) changes to

$$\begin{aligned} H =& \frac{1}{2}\int_0^L \mathrm{d}s\left[A\left(\dot{\varphi}_s^2\cos^2 + \dot{\vartheta}_s\right) + C\Delta\psi_s^2\right] \\ & - \int_0^L \mathrm{d}sF\cos\varphi_s\cos\vartheta_s \\ & - 2\pi\left(Wr(\mathbf{t}_s) + \int_0^L \mathrm{d}s\frac{\Delta\psi}{2\pi}\right)\tau_F \end{aligned}$$

with

$$Wr(\mathbf{t}_s) = \frac{1}{2\pi}\int \mathrm{d}s\frac{\dot{\vartheta}_s\sin\varphi_s + \dot{\varphi}_s\cos\varphi_s\sin\vartheta_s\cos\vartheta_s}{1+\cos\varphi_s\cos\vartheta_s} \tag{1.51}$$

the writhe computed with Fuller's equation (1.39) using $\hat{x}$ as reference axis. Since the twist rate, $\Delta\psi_s$, is constant in the number of turns clamp case (see section 1.3) we can drop the $\Delta\psi_s$-term in eq. (1.51) when analyzing fluctuations $\mathrm{d}\varphi$, $\mathrm{d}\vartheta$ on top of the straight solution $\varphi = 0$, $\vartheta = 0$ (i.e. $(1,0,0)^T$).

The energetic contributions of fluctuations $\mathrm{d}\varphi, \mathrm{d}\vartheta \ll 1$ sums up to

$$\mathrm{d}E = \int_0^L X^T\hat{T}X\mathrm{d}s \tag{1.52}$$

$$X^T = (\mathrm{d}\varphi_s, \mathrm{d}\vartheta_s) \tag{1.53}$$

$$\hat{T} = \frac{1}{2}\begin{pmatrix} -A\frac{\mathrm{d}^2}{\mathrm{d}s^2} + F & \tau_F\frac{\mathrm{d}}{\mathrm{d}s} \\ \tau_F\frac{\mathrm{d}}{\mathrm{d}s} & -A\frac{\mathrm{d}^2}{\mathrm{d}s^2} + F \end{pmatrix}. \tag{1.54}$$

Fourier-modes with wave-number $k_m = \sqrt{\tau_F^2/2A^2 - F/A}$ minimize the determinant of $\hat{T}$:

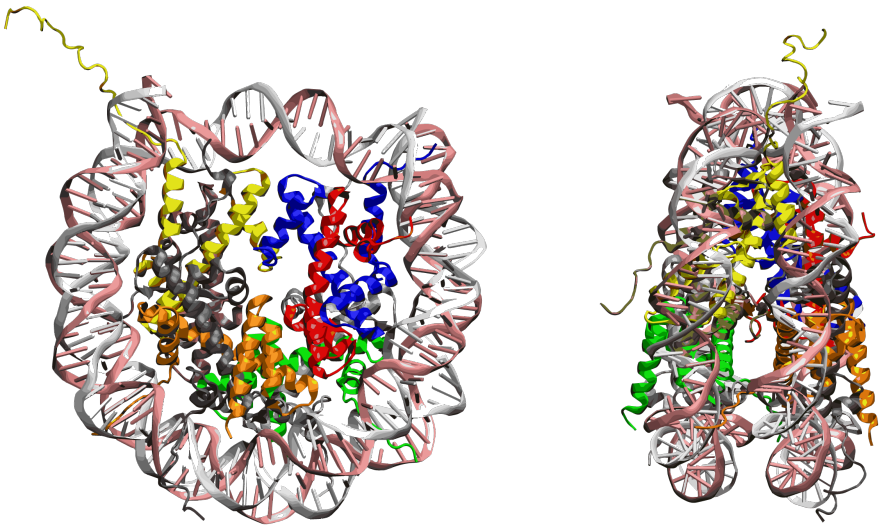
$$(\det \hat{T})|_{k_m} = \frac{\tau_F}{4A^2}\left(AF - \frac{\tau_F^2}{4}\right). \tag{1.55}$$

For $\tau_{\text{crit}} = 2\sqrt{AF}$ the determinant changes sign, and the straight rod solution becomes unstable. At this point, where $n = n_{\text{crit}} = \tau_{\text{crit}} L/2\pi C = \sqrt{AF}\, L/\pi C$, the ground state, for an infinite long chain with tangents at infinity aligned with the force F, shifts from the straight rod to a homoclinic solution (see section 1.2). As said above, we call this point the *bifurcation point*. This point is independent from the force direction or from the particular parametrization of $\mathbf{t}_s$. In fact, in chapter 4 we will use this results even though the force will be in the $\hat{z}$-axis, with $\mathbf{t}_s$ parametrized as in eq. (1.16).

1.6 Nucleosomes

The *nucleosome* core particle (NCP) (fig. 1.10) is composed by 147 bp of DNA wrapped ≈ 1.7 turns along a left-handed superhelical wrapping path of 4.75 nm radius around an octamer of *histone* proteins. Each octamer consists of four pairs of H2A, H2B, H3 and H4, called the four core histones. The shape of the NCP is similar to a *wedge*, i.e. a cylinder with the two surfaces not parallel. The diameter of the octamer is approximately $a = 7.5\,\text{nm}$, with an average height of $b = 6\,\text{nm}$. Since the radius of the nucleosome is so small, the bending energy required for wrapping is extremely high. However 14 binding sites at the octamer surface provide electrostatic interaction and hydrogen bonding with the DNA, making the NCP stable with a net negative energy per binding site of about $\sim 1.5 \div 2k_BT$. Indicating with s^* the number of wrapped turns of DNA around the NCP, the net adsorption energy density results in

$$\frac{\mathrm{d}E_{\text{ads}}}{\mathrm{d}s} = \begin{cases} \varepsilon - \varepsilon_b & \text{if } |s^*| \leq 1 \\ \varepsilon - \varepsilon_b - \varepsilon_{\text{el}} & \text{if } 1.67 \geq |s^*| > 1 \\ -\varepsilon_b - \varepsilon_{\text{el}} & \text{if } |s^*| > 1.67; \end{cases} \tag{1.56}$$



(a) Top view (b) Side view

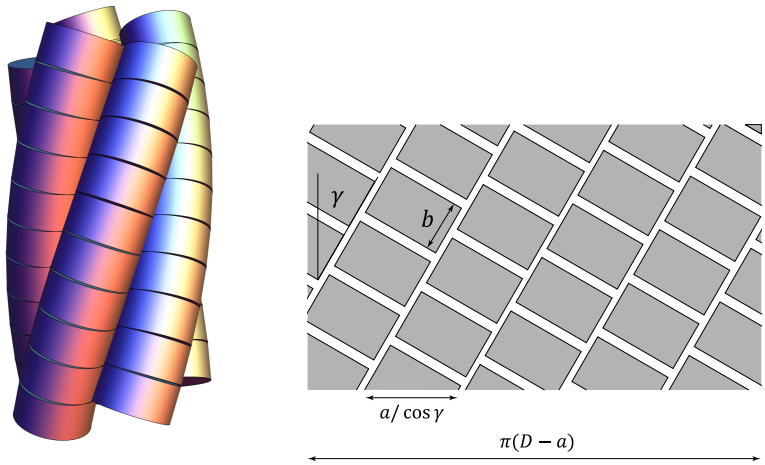
Figure 1.10: The geometry of the nucleosome [39].

Here ε is the pure adsorption energy density whereas ε_b accounts for the DNA bending cost and ε_{el} for the electrical repulsion between the two wrapped turns. These quantities are estimated to be $\varepsilon = 1.51 k_B T/\text{nm}$, $\varepsilon_b = 0.75\, k_B T/\text{nm}$ and $\varepsilon_{\text{el}} = 0.2 k_B T/\text{nm}$ [46].

For every DNA molecule in our cells, many nucleosomes are present, forming a beads-on-a-string like structure (the nucleosomes representing the beads). On average nucleosome are separated by $10 - 70$ bp of *linker DNA*. Summing the length of the linker DNA with the length of the DNA wrapped around the nucleosome gives the *repeat length*[8] which can vary from cell to cell (and from species to species).

To fit inside the cell, the nucleosomes sequence fold into a fiber (see figure 1.11a). This partially explains how a negatively charged polymer of two meters of contour length can fit inside the cells' nuclei: the positively charged octamer offset the charge on the DNA, and the hydrogen bonds help offsetting the bending energy required to compact the molecule.

[8]The relation between repeat and linker length is straightforward: repeat length = 147 bp + linker length.



(a) A picture of the chromatin fiber with diameter D. The DNA around each octamer is omitted.

(b) The nucleosome-footprint running through the nucleosomes centers, hence the $\pi(D - A)$ dimension, obtained by rolling out the cylinder of figure 1.11a on a plane.

Figure 1.11: The geometry of the chromatin fiber.

However many questions remain open. This structure, with diameter approximately equal to $D = 30\,\text{nm}$, is commonly called the *chromatin fiber*. Its constituents, NCP's and DNA, are known at atomic resolution but the fiber itself remains poorly understood, despite more than three decades of experiments. A wide range of models was put forward to explain the experiments that can be divided in roughly two classes: solenoid [22] and crossed-linker [74, 3] models. The former class assumes that consecutive nucleosomes along the DNA stack on top of each other, while in the latter the nucleosomes sit on opposite sides of the fiber. Neither class predicts, however, the optimal fiber conformation and the geometry of the fiber, i.e. its diameter, is fixed *ad hoc*. The resulting insight is then rather limited due to the huge number of possible configurations and hardly any experiments to distinguish between them. Since the groundbreaking study of the Rhodes groups [55] there is, however, more to explain than a single diameter. In these experiments regular fibers were re-

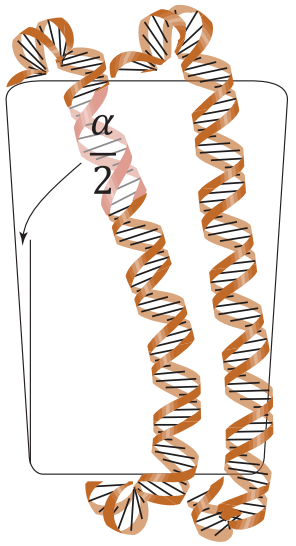


Figure 1.12: A side view of the nucleosome, showing the precise meaning of α.

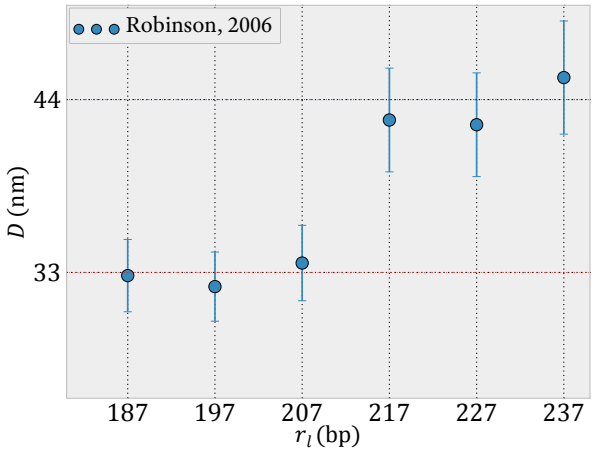


Figure 1.13: The diameters of the chromatin fiber as presented in [55].

constituted by placing about 50 nucleosomes equally spaced onto a piece of DNA. To space the nucleosomes the group used a *positioning sequence*, i.e. a certain combination of ATCG. The sequence used is called the 601 positioning sequence and the NCP have a higher affinity with it, meaning that they are most likely to bind there than in other places. If the 601 sequences are equally spaced then there is a high probability that also the nucleosomes will be equally spaced.

Varying the distance between the 601 sequences the Rhodes group studied repeat lengths from 187 to 237 bp in steps of 10 bp. The experimental findings were surprising (figure 1.13): for the three shorter repeat lengths, fibers with 33 nm diameter were reported. Even more remarkably, for the larger three repeats thick fibers, with a non-canonical 44 nm diameter, were observed. These findings point toward a discrete set of optimal nucleosome configurations that act as the main driving force for fiber formation. This leads to two questions: (*i*) which principle underlies

that discrete set of optimal nucleosome arrangements? (*ii*) Why does the rather stiff DNA double helix not affect the fiber diameter when the repeat length is varied over a range of at least 20 bp for the 33 and 44 nm wide fibers respectively? These two questions remain unanswered by the fiber model proposed in [55] and by models built upon it, like reference [73], where the fiber diameters are set *ad hoc*. In particular question (*ii*) remains unanswered by the two-angle models that predict fiber diameters that depend linearly on the DNA linker length (see for example references [74, 60, 37, 31, 15]).

A recent paper [14] gives a possible answer to (*i*). The authors start by assuming that nucleosomes pack densely inside the chromatin fiber, stacked on top of each other in a structure similar to figure 1.11. Each nucleosome belongs to a ribbon that follows an helical path with radius $R = (D - a)/2$ and pitch angle $\pm\gamma$.

In principle the nucleosomes could also stack side by side (rotating the nucleosomes footprints of figure 1.11b by 90°) as suggested in [55]. However NCP's are know to spontaneously stack face to face [19] as a consequence of the short-range attractive interaction between their faces.

A dense footprint packing implies that the area of the cylinder onto which the nucleosomes pack is equal to the total area of the footprints, i.e.

$$\sigma b a = \pi(D - a) \tag{1.57}$$

where σ is the nucleosome line density (NLD). The linear relation between the fiber diameter and σ can be experimentally verified since σ is a measurable quantity. In figure 1.14 we see how data from [55] fit equation (1.57).

The pitch angle γ is related to the fiber diameter by requiring the number of ribbons to precisely cover the periphery of the fiber

$$\pi(D - a) = a N_{\text{rib}} / \cos\gamma. \tag{1.58}$$

Without entering into the details, provided by [14], taking the three dimensional packing into account, the wedge angle α (see figure 1.12) is related to the pitch angle γ by

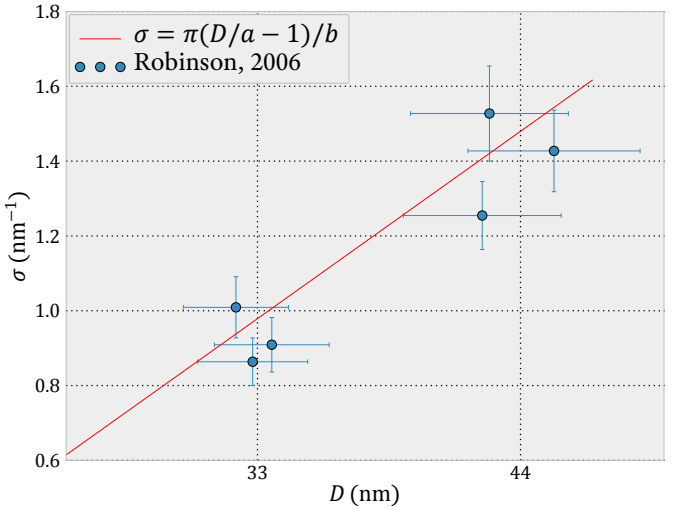


Figure 1.14: A comparison between the experimental data in [55] and equation (1.57).

$$\alpha \approx \frac{2b}{D-a}\left(1-\cos^2\gamma\right). \tag{1.59}$$

Another conditions on the possible fiber structure stems from the linker DNA. Denoting N_{step} the distance across ribbons between connected nucleosomes, the necessary and sufficient condition for a regular backbone winding (BW) — defined by $(N_{\text{rib}}, N_{\text{step}})$ — is the existence of two integers n and k such that

$$kN_{\text{step}} - nN_{\text{rib}} = 1, \quad 0 \le n \le k \le N_{\text{rib}}. \tag{1.60}$$

The condition ensures that, after k steps and n turns, neighboring ribbons are connected so that every ribbon is visited by the backbone. For example, the BW $(N_{\text{rib}}, 1)$ has consecutive nucleosomes in neighboring ribbons, and it obeys equation (1.60) for every N_{rib}. Figure 1.16, instead, represents BW $(5, 1)$. As noted in [14] this approach covers all major models for the fiber structure [28, 74, 72, 75, 41, 12, 60] i.e. the solenoid (BW $(1, 1)$ [22]), crossed-linker and interdigitated models, including new possibilities not considered previously.

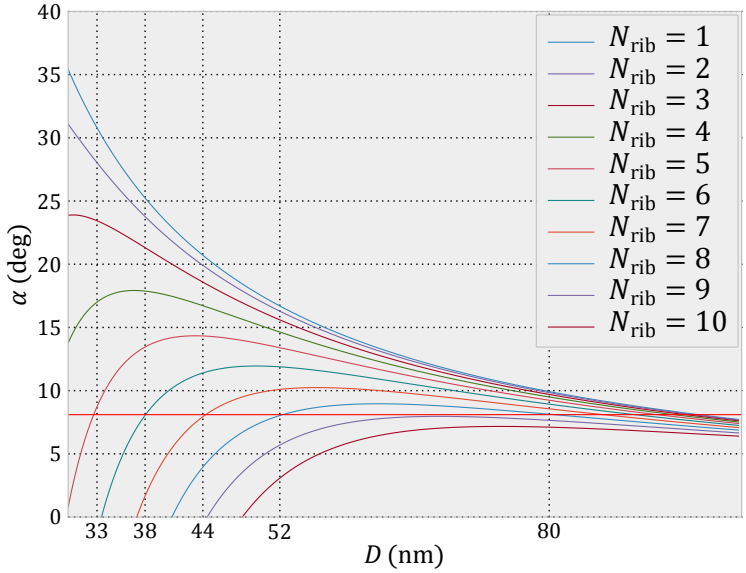


Figure 1.15: A plot of equation (1.59) and of $\alpha \approx 8°$. The intersections between the straight line and the curves mark the diameters compatible with the measured value of α.

Reference [14] provides thus an unifying framework for the chromatin fiber models presented in the past. These models were so successful because, once they set the diameter of the fiber, they could predict their NLD's. However equation (1.57) implies a linear relationship between the diameter of the fiber and the NLD meaning that setting the former implies the latter: therefore those models lose their predictive power in the dense-packing scenario.

To test the model we could use the connection that equations (1.58) and (1.59) provide between the wedge angle, $\alpha \approx 8°$, a microscopic parameter independently verified, the number of ribbons of the chromatin, N_{rib}, and its macroscopic diameter, D, experimentally accessible. If the dense-packing assumption is correct, the 33 and 44 nm diameters observed should correspond to the measured α. This is indeed the case, as the plot of equation (1.59) in figure 1.15 shows: of the four possible diameters compatible with $\alpha \approx 8°$ (for $D \lesssim 80$ nm) two are the one actually

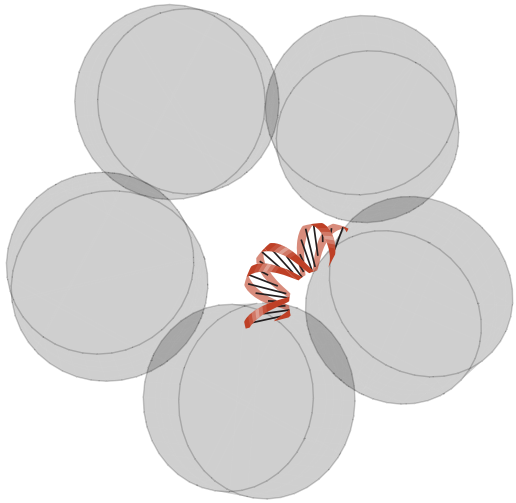


Figure 1.16: The (5, 1) BW.

measured.

N_{rib}	5	6	7	8
D	33	38	44	52

Table 1.1: Number of of nucleosome stacks, N_{rib}, in dense fibers together with their diameters in nm. The diameters follow from the geometry of the nucleosomes that are wedge shaped with a wedge angle of $\alpha = 8.1°$.

Table 1.1 identifies, for each N_{rib}, the diameter that the fiber with that number of ribbons should have to be consistent with the dense packing assumption and with $\alpha \approx 8°$. However there is no information about the sign of γ, the positive or negative backbone winding[9] and which N_{step} the observed fibers have. Moreover the reason for the jump from 33 nm to 44 nm is still unanswered. Chapter 2 will satisfy the curious reader.

[9]The sign of the backbone winding depends on whether, after starting from nucleosome x and visiting N_{rib} nucleosomes, we end up above or below nucleosome x. In case above (below), the backbone winding has positive (negative) helicity.

Chapter 2

The chromatin fiber

> I've made the calculation at home. That doesn't imply that the result is correct, but, merely, that it tends towards being correct.
>
> Giuseppe de Marco

This chapter is a direct continuation of section 1.6. You really should read it if you haven't. There we introduced the concept of dense packing of nucleosomes and used this concept to explain the results presented in [55]. The reader should always keep in mind that we assume that the nucleosomes are densely packed. Low nucleosome line densities (NLD), as typically observed for natural fibers *in vitro* [4] and *in situ* [59] indicate non-dense fibers and are not considered in our current study.

2.1 Introduction

In this chapter we study the linker DNA more precisely. Previously we presented the various structures compatible with the measured *wedge* angle, but the role of the linker DNA remained unclear. Since bending DNA can require tens of k_BT of energy, we expect the linker DNA to play a fundamental role in the chromatin fiber formation. Before continuing, we introduce the linker histones H1/H5. The first and last 10bp of every

linker DNA (at the entry and exit point of the nucleosome, see figure 2.1) are strongly bound to the globular domain of these histones [66]. The DNA that can be bent to go from one nucleosome to the consecutive one is therefore given by the repeat length shortened by the length wrapped around the nucleosome (147 bp) and the one bound to the linker histones (20 bp). When considering a repeat length of 207 bp the linker DNA is therefore only 40 bp. If these 40 bp have to connect nucleosomes in two consecutive ribbons, for nucleosomes in the same plane, their bending energy alone would lie in the $50\,k_BT$ range — even if one does not enforce a particular DNA entry-exit angle at the nucleosome.

Such energies would clearly overrule the stacking energy, the energy gain from putting one nucleosome on top of another, that has been estimated from chromatin fiber stretching experiments [11], theory [60] and simulations [42] to be on the order of $3\,k_BT$.

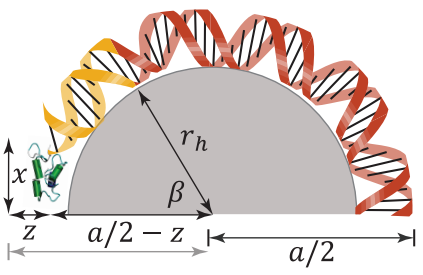


Figure 2.1: Half nucleosome with the nucleosomal DNA (red), the stem (orange), and the linker histone (green). Here $r_h = 3.75\,\text{nm}$ and $\beta = 0.33\pi$.

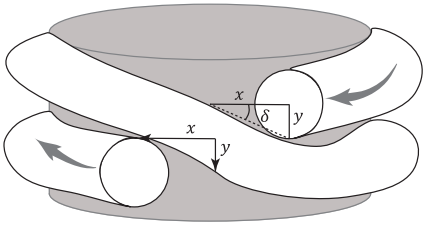


Figure 2.2: A cartoon of the nucleosome from another perspective. For simplicity, the linker histone is omitted. x and y are the distances between the centerline of the DNA and the dyad axis of the nucleosome.

Here we demonstrate how to solve this problem. Keeping the dense nucleosomal packing intact, the nucleosomal stacks can be shifted "out-of-register" in a way that reduces the elastic energy per linker to about one k_BT without changing the stacking energy. The predictions of our model — based only on geometrical constraints and DNA elasticity — agree remarkably well with the experimental data from ref. [55]. Our model ap-

plies to dense fibers that only form for perfectly spaced nucleosomes but not for native fibers like in ref. [71]. Also fibers with regularly spaced nucleosomes were excluded if the linker length was too short [58, 56], the total number of nucleosomes too small [17, 58, 61, 33, 27], or if there were no linker histones present [61, 16, 58].

2.2 Results

Since the elastic energy of the DNA is quite sensitive to the distance between two nucleosome, we discuss in detail the geometry of the nucleosome. Figure 2.1 shows a top view of a nucleosome half. The DNA exits the wrapped part to the left. The linker histone, close to the entry-exit point, binds the two DNA linkers together forming a stem region [66]. The tip of the stem has a distance $r = D/2 - a + z$ from the centerline of the fiber. If the centerline of the DNA is shifted from the dyad axis of the nucleosome by an x- and y-offset (see figure 2.2), the distance d that a linker has to span to connect two consecutive nucleosomes is

$$d(\Delta, N_{\text{rib}}, N_{\text{step}}) = \left\{\Delta'^2 + 2r^2\left[1 - \cos\left(2\pi\frac{N_{\text{step}}}{N_{\text{rib}}} + \Delta' \tan\frac{\gamma}{R} - 2\frac{x'}{r}\right)\right]\right\}^{\frac{1}{2}} \tag{2.1}$$

with

$$\begin{aligned}
\Delta'(\gamma) &= \Delta + 2y' \operatorname{sign}(\Delta) \operatorname{sign}(\gamma),\\
x'(\gamma) &= \sqrt{(x^2 + y^2)}\cos(\delta + \gamma),\\
y'(\gamma) &= \sqrt{(x^2 + y^2)}\sin(\delta + \gamma).
\end{aligned}$$

In these equations Δ is the vertical offset between the two nucleosomes and $\delta = \tan^{-1} y/x$. Note that γ and R in equation (2.1) depend on N_{rib} [14].

The derivation of equation (2.1) is quite simple if we define the centerline of any ribbon to be described by

$$\mathbf{r}(s) = \left[r\cos\left(s\frac{\sin\gamma}{R} + \vartheta_i\right), r\sin\left(s\frac{\sin\gamma}{R} + \vartheta_i\right) s\cos\gamma\right] \tag{2.2}$$

with $\vartheta_i = i2\pi N_{\text{step}}/N_{\text{rib}}$, $i = 0, \ldots N_{\text{rib}}-1$. Note that $R = (D-a)/2$ since the pitch of the ribbon depends on R instead of r. Taking into account the correction due to the x-, y- and z-offsets as in figures 2.1–2.2 we find equation (2.1).

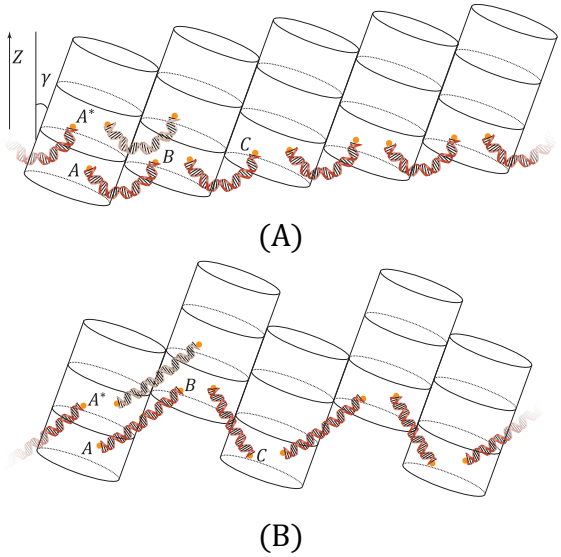


Figure 2.3: The 5 ribbon fiber rolled out in a plane, omitting the wrapped DNA in the figure for clarity. A) A constant vertical offset $b\cos\gamma/N_{\text{rib}}$ between connected nucleosomes (e.g. A-B) leads to highly bent linkers. B) A zig-zag geometry with vertical offsets $\Delta_\uparrow$ for A-B and $\Delta_\downarrow$ for B-C can have nearly straight DNA linkers, except close to the entry/exit points where we assume denaturation. Note that the fiber connectivity does not change from a) to b) and that the Z-axis indicates here the axis of the fiber and is not related to the z in figure 2.1.

The vertical offset Δ is not a free parameter. Starting from some arbitrary nucleosome after N_{rib} steps every ribbon has been visited once and the DNA ends up at the starting ribbon, just one nucleosome above or below. The sum of all the N_{rib} offsets between the connected ribbons must equal $h = \pm b\cos\gamma$ where $b = 6$ nm is the height of a nucleosome and h its height projected on the fiber axis. The sign of h determines the helicity of the linker path. We choose the geometry such that a positive h-

value leads to a positive helicity. The most obvious choice is $\Delta = h/N_{\text{rib}}$ for every vertical offset. But, as mentioned above, this would increase the bending energy too much, making the stacking of nucleosomes too costly. However, a vertical offset alternating between positive and negative values (still adding up to h after one round) circumvents this problem. Starting from a fiber where the offsets are uniform and the linkers highly bent, see figure 2.3A, one can arrive, by shifting stacks up and down, at a conformation where the linkers are almost straight, see figure 2.3B.

Before computing the energies we assume that a small DNA portion at the point where it enters/exits the linker histone is denaturated. This allows the linker DNA to point in any direction and to twist without further cost. Obviously the denaturation comes at some cost, typically about $1 - 3k_BT$ per base pair [57]. As a few base pairs need to be denatured, this might cost about $10k_BT$ in total. We justify this assumption by the fact that the resulting elastic energy per linker is substantially reduced, namely by several tens of k_BT. We furthermore speculate that the linker histone might facilitate the formation of the denaturated region, lowering its free energy cost. Recent experiments showing how the linker histone enhances the conformational flexibility of the DNA at the entry/exit point of the nucleosome [62] might support this idea.

Since the DNA freely rotates and twists at the nucleosome entry/exit points it assumes a planar shape. Its optimal energy is given by finding the minimum of

$$E = \int_0^l \frac{A}{2}\dot{\xi}^2(s) \tag{2.3}$$

where $\xi(s)$ represents the Euler angle parametrizing the linker DNA and l is the length of the linker DNA. The distance between entry and exit point is given by

$$x(l) - x(0) = \int_0^l \cos\xi(s)\mathrm{d}s = d \tag{2.4}$$

with d given by equation (2.1). This end-to-end distance clamp can be incorporated as a Lagrange multiplier, so that the minimization of the en-

ergy has to be done with the Lagrangian

$$L = \frac{1}{2}A\dot{\xi}^2(s) + \mu\cos\vartheta(s);$$

the Euler-Lagrange equations give

$$\begin{aligned} A\ddot{\xi}(s) &= -\mu\sin\vartheta(s) \\ \frac{A}{2}\frac{\mathrm{d}}{\mathrm{d}s}(\dot{\xi})^2 &= \mu\frac{\mathrm{d}}{\mathrm{d}s}\cos\xi; \end{aligned}$$

integrating gives (m_1 is an integration constant)

$$\begin{aligned} \frac{A}{2}\dot{\xi}^2 &= \mu\cos\xi + m_1 \\ &= \mu - 2\mu\sin^2\frac{\xi}{2} + m_1 \\ &\rightarrow \\ \left(\frac{\dot{\xi}}{2}\right)^2 &= \frac{\mu}{A}(m - \sin^2\frac{\xi}{2}) \end{aligned}$$

with $m = (1 + m_1/\mu)/2$. It is possible to rewrite this as an integral (only up to $l/2$, since the solution is symmetric)

$$\int_{\xi(l/2)/2}^{\vartheta(s/2)/2} \frac{\mathrm{d}\xi/2}{\sqrt{1 - \frac{1}{m}\sin^2\frac{\vartheta}{2}}} = \sqrt{\frac{\mu m}{A}}\left(s - \frac{l}{2}\right). \tag{2.5}$$

The solution is symmetric, with entry and exit angle opposite but arbitrary and $\xi(l/2) = 0$. Equation (2.5) becomes

$$\begin{aligned} \mathrm{F}\left(\frac{\xi(s)}{2}\middle|\frac{1}{m}\right) &= \sqrt{\frac{\mu m}{A}}\left(s - \frac{l}{2}\right) \\ \frac{\xi(s)}{2} &= \mathrm{am}\left(\sqrt{\frac{\mu m}{A}}\left(s - \frac{l}{2}\right)\middle|\frac{1}{m}\right) \\ \sin\frac{\xi(s)}{2} &= \mathrm{sn}\left(\sqrt{\frac{\mu m}{A}}\left(s - \frac{l}{2}\right)\middle|\frac{1}{m}\right) \end{aligned}$$

$$\cos\xi(s) = 1 - 2\,\mathrm{sn}^2\left(\sqrt{\frac{\mu m}{A}}\left(s - \frac{l}{2}\right)\middle|\,\frac{1}{m}\right)$$

$$\cos\xi(s) = 1 - 2m\,\mathrm{sn}^2\left(\frac{1}{\lambda}\left(s - \frac{l}{2}\right)\middle|\,m\right) \tag{2.6}$$

$$\xi(s) = \arccos\left(1 - 2m\,\mathrm{sn}^2\left(\frac{1}{\lambda}\left(s - \frac{l}{2}\right)\middle|\,m\right)\right) \tag{2.7}$$

with $\lambda = (A/\mu)^{1/2}$. Here F is the elliptic integral of the first kind, am its inverse and $\mathrm{sn} = \sin\mathrm{am}$. Equation (2.6) enables us to compute

$$x(s) = -\int_{l/2}^{s} \cos\xi(s)\mathrm{d}s = (s - \frac{l}{2}) - 2\lambda\,\mathrm{E}(\mathrm{am}((2s-l)/2\lambda|m)|m). \tag{2.8}$$

with E the elliptic integral of the second kind.

The derivative of $\xi(s)$ at $s = 0$ should also be 0, since the ends are free. This gives, using equation (2.6)

$$\begin{aligned}\dot{\xi}(s)\big|_{s=0} &= \frac{2\sqrt{m}}{\lambda}\,\mathrm{cn}((2s-l)/2\lambda|m)\big|_{s=0} \\ &= \frac{2\sqrt{m}}{\lambda}\,\mathrm{cn}(l/2\lambda|m) = 0.\end{aligned}$$

Therefore, for the periodic properties of cn ($\mathrm{cn} = \cos\mathrm{am}$)

$$\lambda = \frac{l}{2\,\mathrm{K}(m)(2n+1)}, \quad n \in \mathbb{N}. \tag{2.9}$$

Plugging this into Eq. (2.8) yields

$$x(s) = \left(s - \frac{l}{2}\right) - \frac{l}{\mathrm{K}(m)(2n+1)}\,\mathrm{E}\left(\mathrm{am}\left(\left(\frac{2s}{l} - 1\right)K(m)(2n+1)\middle|\,m\right)\middle|\,m\right); \tag{2.10}$$

here $\mathrm{K}(m) = \mathrm{F}(\pi/2|m)$. Equation (2.4) gives

$$x(0) = -\frac{l}{2} - \frac{l}{\mathrm{K}(m)(2n+1)} \mathrm{E}\left(\mathrm{am}\left(-K(m)(2n+1)\middle|\, m\right)\middle|\, m\right) = \frac{d}{2}. \tag{2.11}$$

Once the linker length l and the distance between entry/exit point d are known, choosing the desired n (as in eq. (2.9)) we solve Eq. (2.11) numerically to find $m(d, l)$. Once found, it can be used to compute the energy. Plugging infact $\dot{\xi}$ from Eq. (2.7) in Eq. (2.3), taking into account Eq. (2.9), results in

$$\begin{aligned} E(d,l) &= 2\int_s^{l/2} \frac{A}{2}\dot{\xi}^2(s)\mathrm{d}s\Big|_{s=0} \\ &= -\left(A\frac{8\,\mathrm{K}(m)}{l^2} l\,\mathrm{E}\left(\mathrm{am}\left(\frac{(2s-l)}{l}\,\mathrm{K}(m)\middle|\, m\right)\middle|\, m\right)\right)_{s=0} \\ &\quad -\left(A\frac{8\,\mathrm{K}(m)}{l^2}(1-m)(l-2s)\,\mathrm{K}(m)\right)_{s=0} \\ &= \frac{8A\,\mathrm{K}(m)}{l}\left(\mathrm{E}\left(\frac{\pi}{2}\middle|\, m\right) - (1-m)\,\mathrm{K}(m)\right) \end{aligned} \tag{2.12}$$

where the explicit dependency of m on d and l has been omitted for simplicity.

From eq. (2.12) we can calculate the average energy per linker DNA, $E_l(\{\Delta_i\}) = \sum_{i=1}^{N_{\mathrm{rib}}} E(\Delta_i, l)/N_{\mathrm{rib}}$. Considering the stacking energy, $E_{\mathrm{stack}} \approx -3k_BT$, this leads to the total energy of the fiber per nucleosome:

$$E_l(\{\Delta_i\}, n) = E_l(\{\Delta_i\}) + E_{\mathrm{stack}}\frac{n - N_{\mathrm{rib}}}{n}. \tag{2.13}$$

Here E_{stack} is multiplied by a factor that accounts for a finite size effect. For a sufficiently small number n of nucleosomes, fibers with less ribbons might be favored because they have less end nucleosomes. When comparing our model to experimental data we account in our calculations for this finite size effect.

Assuming that every fiber seen in the experiments correspond to the energetically most favorable geometry, we numerically minimize the total energy per nucleosome, eq. (2.13), with respect to $\{\Delta_i\}$, N_{rib} and N_{step}.

For each set we have to consider four cases since the ribbons and the linkers can be right- or left-handed, independent from each other. For an even number of ribbons the number of positive vertical offsets is the same as the number of negative ones. For an odd number of stacks and a positive (negative) helicity of the linker backbone, the number of positive vertical offsets exceeds the number of negative offsets by one (minus one). Moreover, for a given set of Δ_i that minimizes the energy, offsets with the same sign have equal values.

N_{rib}	5	6	7	8
D	33	38	44	52

Table 2.1: Number of of nucleosome stacks, N_{rib}, in dense fibers together with their diameters in nm. The diameters follow from the geometry of the nucleosomes that are wedge shaped with a wedge angle of $\beta = 8.1°$.

n	52	61	47	55	66	56
r_l(bp)	187	197	207	217	227	237
$E_l\,(k_BT\,)$	-1	-1.8	-1.4	-1.7	-2	-1.8
$\Delta_\uparrow(nm\,)$	2.2	5.9	7.7	11.2	12.6	15.1

Table 2.2: Optimal fibers for given number n of nucleosomes and repeat length r_l chosen as in the experiment [55]. The energy E_l per linker, eq. (2.13) with $E_{\text{stack}} = -3k_BT$, and the positive vertical offset $\Delta_\uparrow$ are presented for the case $(x, y, z) = (2.5, 0.5, 0.1)$ nm.

The dense fibers considered in our minimization are summarized in table 2.1. We only account for the case $N_{\text{step}} = 1$ since for any $N_{\text{step}} > 1$ one has strong steric interactions between the linkers. Also in the case $N_{\text{step}} = 1$ overlap between linkers can occur when the vertical offsets become too large. We consider in our minimization only allowed configurations. Having set $N_{\text{step}} = 1$ we have — for a given helicity of the ribbons and of the backbone — only one remaining degree of freedom, the

amount by which the ribbons are shifted with respect to each other. The energies per linker for infinite fibers with $E_{\text{stack}} = -3k_BT$ and $(x, y, z) = (2.5, 0.5, 0.1)$ nm are displayed in figure 2.5 in one bp steps between 177 and 237 bp repeat length. We show the energies for all possible numbers of ribbons. Curves for given N_{rib}-values are not smooth since the optimal helicity varies with the repeat length, see also figure 2.4. Note that for the chosen (x, y, z) -values there is no difference in structure between the infinite fibers and the finite ones from figure 2.4. The only role of the stacking energy is to make the energies negative, and therefore the fiber stable. Changing its value produces only a vertical shift in figure 2.5 (up to finite size effects).

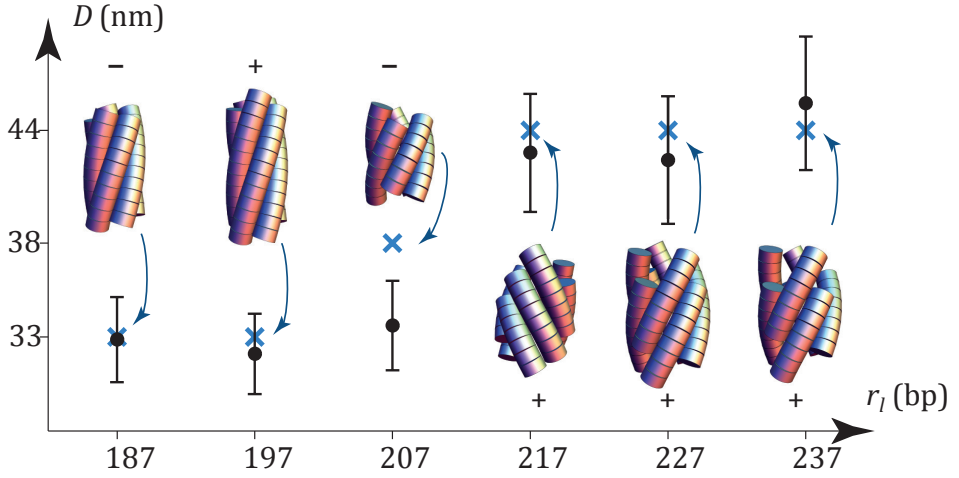


Figure 2.4: Fiber diameter as function of repeat length: experimental data [55] in black, our theoretical prediction for $(x, y, z) = (2.5, 0.5, 0.1)$ nm in blue.

The results for the six experimentally studied fibers [55] are presented in figure 2.4 along with table 2.2 for $(x, y, z) = (2.5, 0.5, 0.1)$ nm. Since these microscopic values are not known precisely we performed the minimization for a range of values (in nm) $0 < x < 3.5$, $0 < y < 2.5$ and $0 < z < 1$. For every set of (x, y, z)-values that gives the blue crosses in figure 2.4, the length of DNA in contact with the linker histone is about 10 bp (i.e., 20 bp per nucleosome), the length that has been shown to be

strongly bound to the globular domain of H1 [66]. We assume that H5, the linker histone used in [55], engages the same length. The helicity of each fibers can be seen directly from the artwork in figure 2.4, while the helicity of the linkers are indicated by + and − signs.

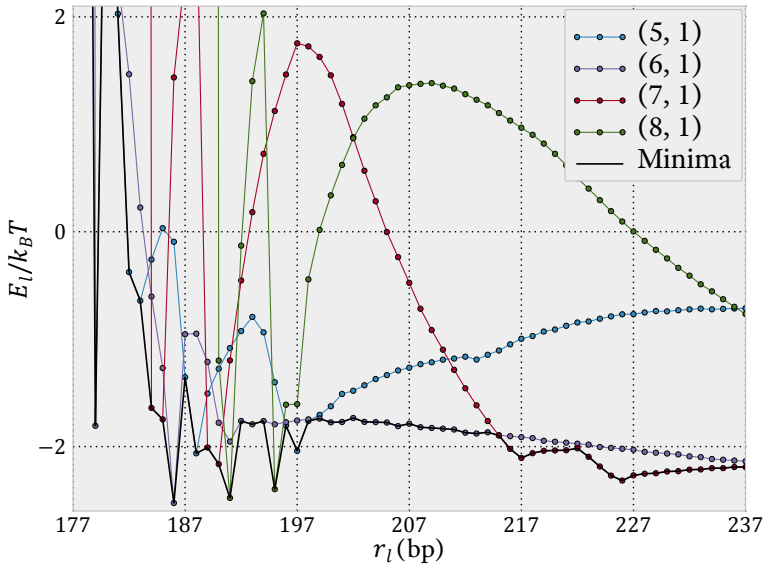


Figure 2.5: Energy E_l per linker, eq. 2.13, as a function of repeat length r_l for the four possible N_{rib}. We assume infinitely long fibers with $E_{\text{stack}} = -3k_BT$ and $(x, y, z) = (2.5, 0.5, 0.1)$ nm. We note that changes in the helicity of the fiber manifest themselves in kinks, as for the 7 ribbons structure at 222 bp (see also figure 2.4).

The predictions of our model are in agreement with the experiments, except for $r_l = 207$ bp, see figure 2.4. However, the electron micrographs (figure 1 of [55]) might indicate that fibers with $r_l = 207$ bp are thicker than the fibers with shorter repeat length and thinner than the ones with larger repeat length. From the five micrographs per repeat length shown in that figure we estimate $D \approx 33$ nm for $r_l = 197$ bp, $D \approx 38$ nm for $r_l = 207$ bp and $D \approx 44$ nm for $r_l = 217$ bp. Moreover, the variations in the diameters for fibers of the same repeat length are much smaller than the error bars, see figure 1. We speculate that the displayed fibers are examples of very regular and dense fibers for which our theory works best.

The whole ensemble of fibers shows larger variations in diameter, presumably reflecting less regular nucleosomal packings, and the average of the 207 bp repeat is even shifted close to 33 nm.

As can be seen from figure 2.5, formation of dense fibers for $r_l = 177$ bp is very expensive and might be even sterically impossible, depending on microscopic parameters. In fact, the Rhodes group found in a new study that 177 bp repeats form non-canonical 30 nm wide fibers (D. Rhodes, private communication, 2011).

We stress that short fibers, i.e. fibers with a small number of nucleosomes, might show different fiber geometries than long ones. E.g. the energies of the 6 and 7 ribbon fibers with $r_l = 237$ bp are so close (see figure 2.5) that $N_{\mathrm{rib}} = 6$ becomes cheaper already for $n = 50$. Very short fibers like e.g. $n = 10$ [17] and $n = 12$ [27] seem to prefer $N_{\mathrm{rib}} = 2$ compromising on perfect packing to have less end nucleosomes.

2.3 Discussion

We have presented a chromatin fiber model that predicts the fiber diameter as a function of the linker length. All the parameters that entered our model, the nucleosomal wedge angle and the DNA elastic modulus, were extracted from experiments that were performed on components of chromatin fibers, disconnected nucleosomes (so-called nucleosome core particles) and naked DNA, but not on chromatin fibers themselves. The first assumption of our model, the dense packing of the nucleosomes, leads to four different possible geometries. This together with a second assumption, namely that the experimental fibers are those with the lowest elastic energy per linker DNA, are already sufficient to predict the fiber diameters seen in the experiment.

To achieve constant fiber diameters over an extended range of linker lengths, the nucleosome stacking energy has to dominate over the elastic energy for linker bending. According to our study, this is achieved when (*I*) the DNA is locally denatured close to the entry-exit region. Extreme as it may sound it lowers the elastic energy by several tens of k_BT per linker. As the nearest-neighbor base pair free energy depends strongly on the base pair step, one might ask whether stretches with low cost for

denaturation are typically found just next to nucleosome positioning sequences. Linker histones are indeed known to preferably bind to AT rich regions [10]; (*II*) nucleosomes are equally spaced to achieve small energies through "out-of-register" sliding. If just two neighboring nucleosomes are connected by a shorter linker length than the rest, the stacks belonging to these two nucleosomes will not be able to shift by the optimal amount. In that case either the fiber cannot form, or a nucleosome has to disintegrate to allow the rest of nucleosomes to pack. On similar grounds we expect that optimal fibers are very stable against thermal fluctuations even though for certain linker lengths different fibers have similar energies. A thermal excitation in the form of a short fiber stretch with non-optimal geometry would be too costly to spontaneously occur as different sliding lengths would cause steric clashes at the boundaries.

The strict requirement of equal spacing of nucleosomes for the formation of dense fibers might have implications for living cells. Our model suggests that dense fibers would only form for equally spaced nucleosomes. Since linker lengths typically vary along DNA, perfectly dense fibers, as discussed in this paper, can hardly form. Instead one should expect less dense and less regular fibers as typically found when chromatin is isolated from cells, see e.g. [4]. Such less dense fiber stretches interdigitate with neighboring fibers, making them harder to detect *in vivo*. In the dense environment of the cell nucleus they may even disintegrate into a nucleosomal melt [40]. Nevertheless there are mechanisms that can cause an approximately equal spacing of nucleosomes *in vivo*, namely directly through mechanical signals in the underlying DNA sequence [30], or indirectly through statistical ordering in the vicinity of barriers [69, 44, 1]. Furthermore, one might speculate that the action of chromatin remodellers like ISWI that are known to repress transcription by forming equally spaced nucleosomes [26], make use of this phenomenon. Once they have equally spaced an array of nucleosomes, a dense fiber can form and the corresponding DNA stretch can no longer be accessed.

In the next chapter we will show how, combining torsion and tension, we can unwrap DNA from NCP's.

Chapter 3

Unwrapping

> If you ask me now, I cannot prove it, but I'm sure it's true.
>
> Giuseppe de Marco

In this chapter we investigate the effect of torque and force on a *nucleosome*. Using the worm-like chain model (WLC) we show how low negative torques eases the unwrapping of the DNA from the nucleosome. In some case a combination of high forces and high positive torques, surprisingly, favors the unwrapping as well. The theory presented provides an interesting insight on how to access the genetic code with tensions smaller than what previously thought.

To study the response, we consider a molecule of DNA bound to a single nucleosome. However, with due modifications, this applies to more general DNA spools, widely found in nature, as the Lac1 repressor [67], DNA gyrase [29] and RNA polymerase [54].

3.1 General model

We consider a nucleosome, with DNA legs at its ends, under tension f and torque, see fig. 3.1. In our model the DNA is described as a worm-like chain being wrapped around a cylinder that represents the histone

octamer. The wrapped section of the DNA molecule is described by the space curve $\mathbf{r}_n(s) = r(\pi s \tan\alpha, \cos\pi s, \sin\pi s)$ with $r = 4.3\,\text{nm}$ and $\alpha = -0.085$ and thus a pitch of $2\pi r \tan\alpha$. A nucleosome with s^*-turns of DNA adsorbed is described[1] by $\mathbf{r}_n(s)$ with $s \in [-s^*, s^*]$. To unwrap its DNA the nucleosome has to rotate around the $y-$axis by an angle β (fig. 3.1) resulting in $\mathbf{r}_r(s,\beta) = O_y(\beta)\mathbf{r}_n(s)$ with O_y denoting the corresponding rotation matrix [36]. To avoid collision between DNA and nucleosome we impose $\beta \in [-\pi + \alpha, -\alpha]$.

In the torsion-less case the shape of the planar DNA, where its ends are aligned with the force, has been worked out in [36] using the Kirchhoff kinetic analogy [51]. Adding torsion causes the DNA legs to bend out-of-plane. Since a non planar homoclinic loop is only favored when the inserted number of turns is between -1 and 1, and since most of the non-planarity is contained within the wrapped part of the DNA (see the section at the end of this chapter), we simplify our analysis by describing the legs by the planar homoclinic orbit with the tangent vector

$$\mathbf{t}_l(s) = O_z(\delta)(0, \sin\theta(s), \cos\theta(s))$$

with $\cos\theta(s) = 1 - 2\,\text{sech}^2(s/\lambda)$; here $\lambda = \sqrt{A/f}$ with A being related to the DNA persistence length $l_p = A/k_BT \approx 50\,\text{nm}$. From $-s_0$ to $+s_0$ we replace this curve with the wrapped nucleosomal DNA (see fig. 3.1). The δ-rotation of the DNA legs ensures continuity at the insertion point, without affecting the energy. In addition continuity requires

$$0 = \mathbf{t}_l(s_0) + t_n(-s^*) \tag{3.1}$$

therefore

$$s_0(s^*, \beta) = \frac{\lambda}{t}\,\text{arcsech}\,\frac{t_{\min}}{t} \tag{3.2}$$

$$t_{\min} = \sqrt{\frac{1 + \cos\alpha\cos\pi s^*\cos\beta - \sin\alpha\sin\beta}{2}}.$$

[1] Not counting an eventual translation, irrelevant for the energy.

and $t = 1$. In eq. (3.2) t represent the homoclinic parameter, which quantifies how "planar" the legs are. In this work we assume the legs to be perfectly planar ($t = 1$). This approximation is good for several reasons: first of all the domain of arcsech limits t to $[t_{\min}, 1]$. When $s^* \not\approx 0, 1, 2$, the β that minimizes the energy leads to $t_{\min} \approx 1$. On the other hand, when $s^* \approx 0, 1, 2$, the contribution by the legs to the energy is almost 0, since s_0 is very high (see eq. (3.7)). In principle eq. (3.1) determines δ as well, but the parameter is not relevant for our analysis.

As convention we assume that the point $\mp s^*$ of the adsorbed DNA is attached to the point $\pm s_0$ of its free counterpart so that the path of the DNA is described by

$$\mathbf{r}(s, s^*, \beta) = \begin{cases} \int \mathbf{t}_l(s, \delta_1)\mathrm{d}s & \text{if } s \in [-L_l(s^*, \beta), -s_0] \\ \int \mathbf{t}_l(s, \delta_2)\mathrm{d}s & \text{if } s \in [+s_0, +L_l(s^*, \beta)] \\ \mathbf{r}_n(s, \beta) & \text{if } s \in [-s^*, s^*]. \end{cases} \tag{3.3}$$

In the integrals one of the integration boundaries is the length $L_l(s^*, \beta) = (L + 2s_0 - L_n(s^*))/2$, where $L_n(s^*) = 2\pi r s^* \sec\alpha$ is the length of the DNA adsorbed by the nucleosome. The two angles δ_1, δ_2 are important to ensure continuity at the boundary between legs and nucleosome, but otherwise they do not influence the energy. Therefore we drop the δ argument of $\mathbf{t}_l$ from now on. Once s^* and β are known, the energy of the system can be computed from $\mathbf{t}_l(s_0)$ and $\mathbf{r}_r(s, \beta)$. The resulting structure is depicted in fig. 3.1.

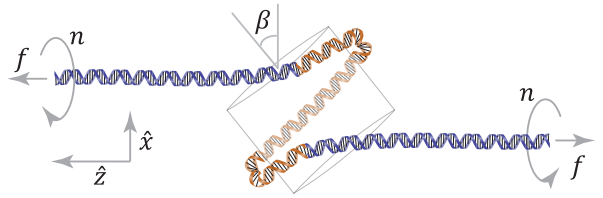


Figure 3.1: The nucleosome under tension and torque. In our model the histone octamer is represented by a cylinder. A part of the DNA molecule is wrapped around it (in an orange hue), the rest forms the legs (in a blue hue).

3.2 Writhe

To compute the writhe of the molecule, we use eq. (1.39) with respect to the $-\hat{z}$-axis:

$$\begin{aligned} Wr_{\mathrm{DNA}} = &\frac{1}{2\pi}\int_{-s_0}^{-L/2} \frac{-\hat{z}\times\mathbf{t}_l(s)}{1+(-\hat{z})\cdot\mathbf{t}_l}\cdot\frac{\mathrm{d}\mathbf{t}_l(s)}{\mathrm{d}s}\mathrm{d}s \\ &+\frac{1}{2\pi}\int_{L/2}^{s_0} \frac{-\hat{z}\times\mathbf{t}_l(s)}{1+(-\hat{z})\cdot\mathbf{t}_l}\cdot\frac{\mathrm{d}\mathbf{t}_l(s)}{\mathrm{d}s}\mathrm{d}s \\ &+\frac{1}{2\pi}\int_{0}^{s} \frac{-\hat{z}\times\mathbf{t}_r(s,\beta)}{1+(-\hat{z})\cdot\mathbf{t}_r}\cdot\frac{\mathrm{d}\mathbf{t}_r(s,\beta)}{\mathrm{d}s}\mathrm{d}s. \end{aligned} \tag{3.4}$$

The first two integrals give no contribution, while the third gives

$$Wr(s^*,\beta) = Wr^i(s^*,\beta) - Wr^i(-s^*,\beta); \tag{3.5}$$

$$\begin{aligned} Wr^i(s,\beta) = &\frac{\arctan\left(\cos\frac{\alpha-\beta}{2}\csc\frac{\alpha+\beta}{2}\tan\frac{\pi s}{2}\right)}{\pi} \\ &-\frac{1}{2}s\sin\alpha - n_{\mathrm{sol}}(s,\alpha). \end{aligned} \tag{3.6}$$

The function n_{sol} eliminates the (here of finite-size) discontinuities of the trigonometric function and it is -1 for $s\in[-3,-1]$, 0 for $s\in[-1,1]$ and 1 for $s\in[1,3]$ etc. Note that eq. (3.6) deviates from eq. (1.41), computed using the axis of the helix (that here rotates with β). The different behavior of the writhe is presented in fig. 3.2.

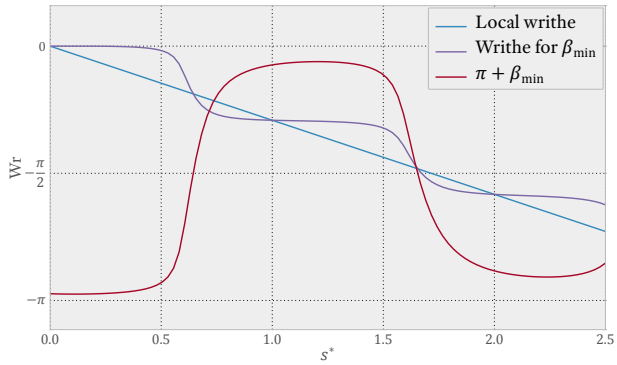


Figure 3.2: A comparison between the local writhe, i.e. using the helix axis in eq. 3.4, and the writhe for the β that minimized the energy, $\beta_{\min}$. For reference also $\pi + \beta_{\min}$ is plotted.

In eq. (3.4) we use $-\hat{z}$ instead of $\hat{z}$ to avoid a vanishing denominator for some values of $\beta < 0$. As required for the use of Fuller's relation, there is a homotopy between the straight $\hat{z}-$axis and any of the states (partially or fully wrapped nucleosome plus rotated legs) considered here. The continuity of the homotopy follows from the fact that the chain continuously changes from $s^* = 0$ (i.e. the $\hat{z}-$axis) to any subsequent state.

3.3 Energy

The total energy of a DNA chain of length L with s^* bound turns inside the nucleosome is the sum of the bending, potential, adsorption and torsional energy:

$$\begin{aligned} E_t(s^*, \beta) = 2 \times \frac{A}{2} \int_{s_0}^{L_l(s^*,\beta)} \dot{\mathbf{t}}_l^2(s) \mathrm{d}s \\ + f\Delta z(s^*, \beta) - 2 \int_0^{s^*} \frac{\mathrm{d}E_{\mathrm{ads}}(s)}{\mathrm{d}s} \mathrm{d}s \\ + \frac{2\pi^2 C}{L - L_n(s^*)} (n - Wr(s^*, \beta))^2. \end{aligned} \tag{3.7}$$

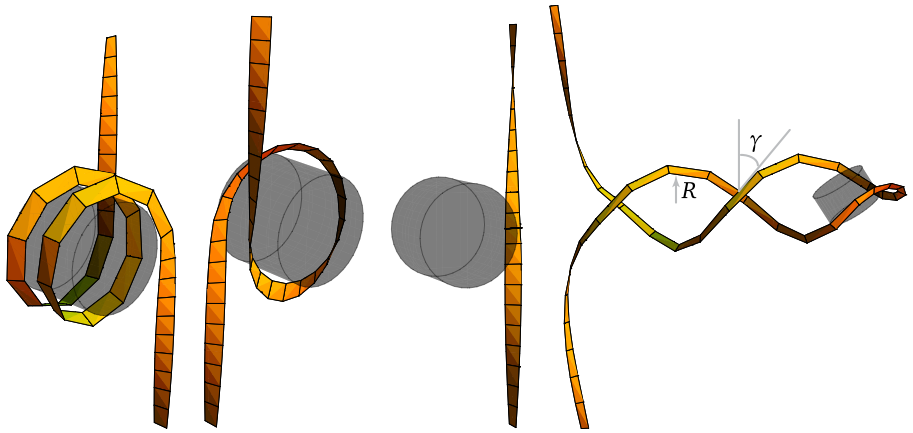


Figure 3.3: Schematic representation of the various stages of nucleosome unwrapping. The roman numerals indicate how many turns are approximately wrapped (N stands for 0, P stands for plectoneme). In order to show the effect of torque, the DNA double helix is here represented as a ribbon that is untwisted in the torsion-less case.

Here

$$\Delta z(s^*,\beta) = L_n(s^*) + (\mathbf{r}_r(-s^*,\beta) - \mathbf{r}_r(s^*,\beta)) \cdot \hat{z} + 2 \times \int_{s_0}^{L_l(s^*,\beta)} (1 - \mathbf{t}_l(s) \cdot \hat{z})\, \mathrm{d}s \tag{3.8}$$

is the shortening of the DNA end-to-end distance in the $\hat{z}$-direction due to the bending of the legs and the wrapping around the octamer. The adsorption energy is given, with the relevant details, by eq. (1.56) and in the last term of eq. 3.7 the quantity C is related to the torsional persistence length l_t via $l_t = C/k_BT$; we assume here $l_t = 100\,\mathrm{nm}$ [24].

To find the optimal configuration for given values of f and n the energy, eq. 3.7, needs to be minimized with respect to s^* and β. Since we neglect in our theory entropic contributions our results are only reliable for large enough forces, $f \gtrsim 0.5\,\mathrm{pN}$ [65].

3.4 Plectoneme

The unwrapping of the nucleosome is eased for moderate positive torques or, as shown later, for high negative torques. However, depending on the force the DNA can also form a structure called *plectoneme* (see fig. 3.3(P)) that adsorbs approximately all the linking number inserted in the system (see [24] and chapter 4). We do not expect nucleosome unwrapping in the presence of a plectoneme as the plectoneme can adsorb torsional stress more efficiently once formed. To estimate the parameter range where the plectoneme occurs, we give here the energy of a DNA molecule of length L that contains a plectoneme of length $p \geq 0$ with radius R and angle γ (fig. 3.3(P)):

$$E(p) = \frac{2\pi^2 C}{L}(n - \mathrm{d}Wp)^2 + (f + \mathrm{d}E_b)p. \tag{3.9}$$

Here $\mathrm{d}W = \cos\gamma \sin\gamma \operatorname{sign} n/2\pi R$ and $\mathrm{d}E_b = A\cos^4\gamma/(2R^2)$ are, respectively, the writhe density and the bending energy density of the plectoneme (see chapter 4).

As specified in chapter 4, in eq. 3.9 we ignore the energetic contribution of the end loop assuming that the nucleosome sits at the end of the plectoneme (fig. 3.3(P)). In principle a plectoneme could also appear somewhere else. However the high bending energy of an end loop makes it highly improbable.

By minimizing eq. 3.9 for p one finds that a plectoneme is expected, i.e. $p > 0$, for all values of n such that

$$n \notin \left[-\frac{(f + \mathrm{d}E_b)L}{4\pi^2 C \mathrm{d}W} + Wr(2), \frac{(f + \mathrm{d}E_b)L}{4\pi^2 C \mathrm{d}W}\right]. \tag{3.10}$$

Here $Wr(2) = -2.14$ is the writhe for 2 fully wrapped turns that for $s^* = 2$ is independent of β (see fig. 3.2). With this term we account for the writhe absorbed by the nucleosome that has around two fully wrapped turns for $n > 0$ and not too large forces. Following chapter 4 we use $\gamma \approx 1$ and $R \approx 1.8$ nm when the salt concentration is about 150 mM. We stress that when the plectoneme forms, it forms on top of the state the system had before the formation. E.g.: if the system has approximately two

turns wrapped, when we, for example, decrease n so that it is outside the interval describe by eq. (3.10), then a plectoneme will form with at its end a nucleosome wrapped twice. In this sense fig. 3.3(P) is only indicative of what really happens.

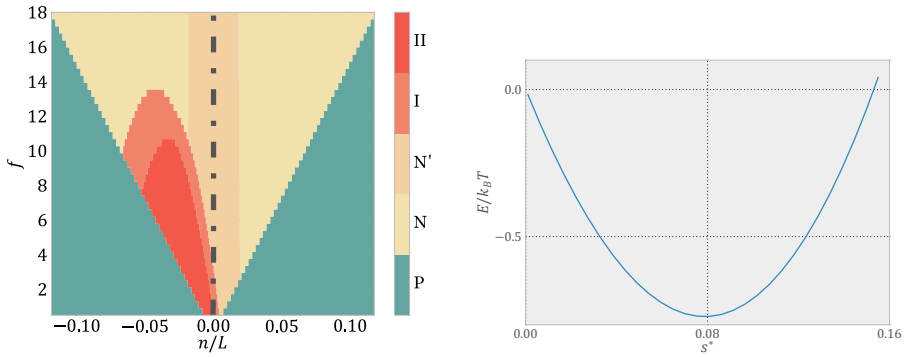
3.5 Twist defects

Apart from the plectoneme, twist defects [35, 45] can influence the nucleosome stability. A twist defects is present in a nucleosome when one DNA base pair is added or removed between two consecutive nucleosome binding sites, resulting in a local under- or overtwisting of the DNA. We can write an equation similar to eq. 3.9 for the twist defects if we replace $p \to m\Delta l$, $\mathrm{d}W \to k \operatorname{sign} n/\Delta l$, $\mathrm{d}E_b \to \mathrm{d}E_d/\Delta l$ and $f \to f \operatorname{sign} n$. m is an integer between 0 and 13 denoting the number of defects. $\Delta l = 0.34\,\mathrm{nm}$ and $k = 1/10$ are, respectively, the length and twist lost or gained by a defect. Finally $\mathrm{d}E_d = 9k_BT$ is the energetic cost of a defect [35]. Since $m \leq 13$ the shift in turns will be up to 1.3; a quick computation reveals that the 13 defects form before a plectoneme occurs. This changes the boundaries where the plectoneme forms, namely we need to subtract 1.3 from the left side of eq. 3.10.

The 1.3 turns per nucleosome are found in experiments where chromatin fibers are put under positive torsional stress [2]. It was suggested that this can be explained by a chiral transition of the nucleosome. Unfortunately a comparison of our model to these experiments is not possible as it involves a multinucleosome geometry and forces where thermal fluctuations cannot be neglected. It would be crucial to perform single nucleosome experiments to see whether the observed strong asymmetry in the response to positive and negative torsion is still present which favors the picture of a chiral transition.

3.6 Results

In fig. 3.4a we present the optimal nucleosome configurations for a wide range of forces and n/L-values of a DNA molecule with $L = 3500\,\mathrm{nm}$



(a) Diagram of state showing the configurations with the lowest energy for $L = 3500\,\text{nm}$ in the f-n/L-plane. The grey dashed-dotted line represents the writhe of the nucleosome when the legs are free to release torsional stress.

(b) The energy landscape near $s^* = 0$ for $f = 10\,\text{pN}$, $n = 0$. The minimum of energy is very close to the $s^* = 0$ case which makes it easy for the nuclesome to "evaporate."

Figure 3.4: Various results from the computations.

length. This diagram of states is nearly identical for all experimentally reasonable values of L, say for all $L > 500\,\text{nm}$. We find five different states, four of which are depicted in fig. 3.3 ((II) fully wrapped, (I) one turn wrapped, (N) unwrapped and (P) fully wrapped plus plectoneme). In addition, we indicate with (N') almost unwrapped configurations. That state is, however, not stable against thermal fluctuations as the global minimum is only tenths of k_BT away from the totally unwrapped state. A typical example is shown in fig. 3.4b. We therefore expect that the nucleosome typically falls apart once it has unwrapped its last turn.

The negative writhe of the wrapping path makes the nucleosome unwrapping highly asymmetric since the factor $(n - Wr(s^*, \beta))^2$ in the torsional energy, eq. 3.7, favors wrapping, $s^* > 0$, for $n < 0$ and unwrapping, $s^* = 0$, for $n > 0$. For large enough negative values of n, however, the nucleosome unwraps to have more twistable DNA available, see fig. 3.4a. The factor $1/(L - L_n)$ in the twist energy dominates then the behavior. In the diagram of states, fig. 3.4a, we indicate also by a dashed-dotted line the torsion-less case where the unbound DNA is free to rotate. This situation has been studied in Ref. [36].

So far we have only determined the optimal configurations via energy minimization. Of special experimental importance is, however, also to know the energy barrier between different states, especially at common boundaries in the diagram of states, fig. 3.4a. Choosing experimental parameters such that one has two minima between a large barrier, one can observe the hopping dynamics between them. This has been indeed observed in the torsionless case where a fast hopping between states (*II*) and (*I*) was observed at a certain force value manifesting itself in a change of the end-to-end extension [43, 33]. The boundaries and corresponding barriers between (*II*) and (*I*) and between (*I*) and (*N'*/*N*) are shown in fig. 3.5. Note that the system under torsion provides a much wider range of parameters where one can observe hopping as compared to the torsionless case. Especially for a wide range of forces we predict two values of n/L where hopping should be observed. It might be challenging to observe the branch with the transitions at the more negative n/L-values as these transition are associated with much higher barrier values (see fig. 3.5).

Appendix: why can we assume the legs to be planar?

The bending and force contribution of the legs in the $t = 1$ case basically always smaller than when $t < 1$. This leaves the writhe as the other possible source of error when excluding the non-planar solutions. However looking at the contribution of the legs to the torsional energy

$$\begin{aligned} E_T &\propto (n - Wr_l)^2 \\ Wr_l &= \frac{2}{\pi} \arcsin t \end{aligned} \tag{3.11}$$

we see that when $n \notin [-1, 1]$, the energy is minimized by $t = 1$. Considering the large amount of turns investigated (see figure 3.4a) we can safely ignore the non-planarity of the loop. Moreover eq. (3.11) is only partially true: when the legs are cut at s_0, which is generally high, at least for the case (*II-I-N*) of figure 3.3, the writhe contribution would decrease,

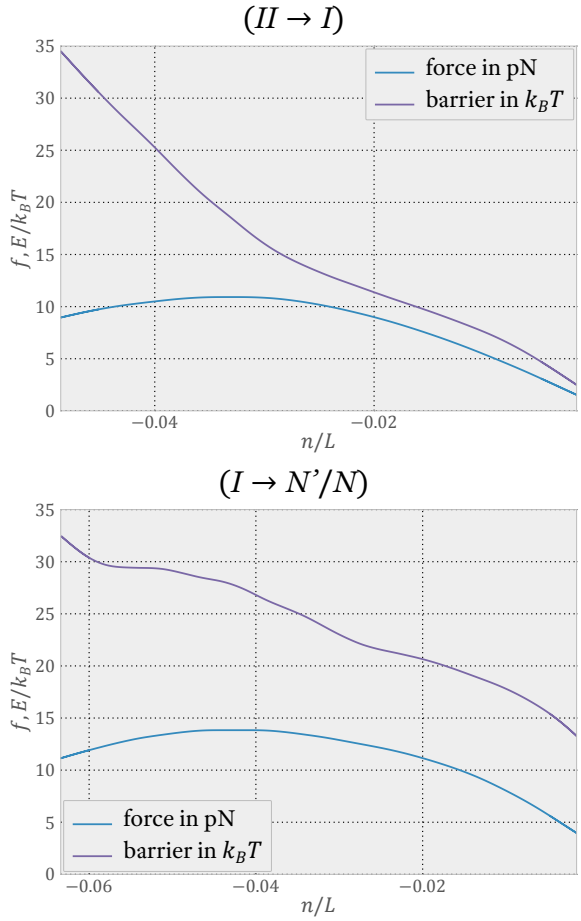


Figure 3.5: The force at which the minimum of the energy around $s^* = 2\,(1)$ and the one around $s^* = 1\,(0)$ have the same value, and the energy barrier necessary to cross from one state to the other, as a function of n/L. Here $L = 3500$ nm.

reducing even further the region in which a non-planar homoclinic loop matters.

Chapter 4

Plectonemes?

> Generally, things obtained without exertion
> are not that useful.
>
> Giuseppe de Marco

In the previous chapter we have mainly investigated the interaction between DNA molecules and nucleosomes. However also naked DNA behaves in an interesting fashion. Take, for example, the DNA persistence length, $l_p \approx 50\,\text{nm}$: it was found by measuring the extension of the molecule when stretched with different forces.

However when the molecule is also twisted, various theories were proposed, but a unifying framework to describe the experimental results was lacking. This is partly due to the important role of thermal fluctuations, extensively analyzed for low torques [51], but either left out [5, 49, 8] or partially added by hand for high torques [50].

In section 1.5, the *bifurcation point* of a straight rod (that is without writhe, see section 1.3 in the same chapter) was identified. In terms of the number of inserted turns, the bifurcation point is at $n_{\text{crit}} = \sqrt{Af}L_c/\pi C$ where L_c is the contour length of the DNA, f the force applied to it, and A and C are related to the bending and torsional persistence lengths (see subsection 1.1) by $A = k_B T l_p$, $C = k_B T l_t$.

Since the DNA molecule is self-avoiding, the cheapest way to produce writhe is the *plectoneme*, a stretch of the chain branching off in a

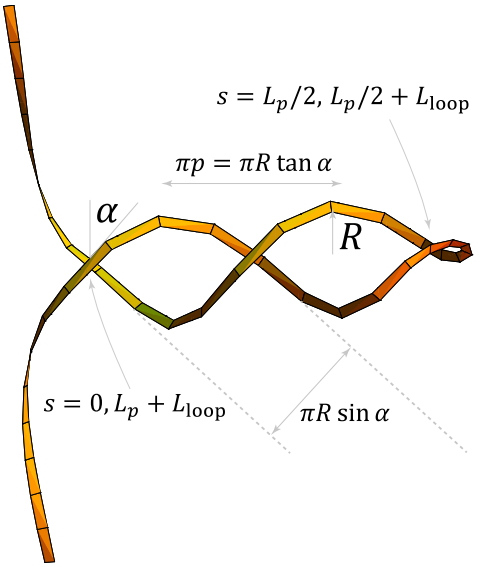


Figure 4.1: A plectoneme, along with the end loop, closing it, and the rest of the straight DNA. The path describing the plectoneme is parametrized by s, going from 0 to $L_p/2$ and from $L_p/2 + L_{\text{loop}}$ to $L_p + L_{\text{loop}}$. The gap is filled by the end loop.

perpendicular direction from the remaining two tails. The two halves circle around each other in a helical path like an old fashioned telephone wire and are connected by an *end loop* (see figure 4.1). A system with a sufficient number of turns n inserted has a plectoneme (plus the tails) as its ground state. This transition typically happens before n_{crit}. Since the molecule is too small to directly observe plectoneme formation, their presence is inferred by measuring the DNA end-to-end extension as a function of n. These curves are called the *turn-extension curves* and the increase in plectoneme length L_p is observed as a linear decrease of the end-to-end extension after the transition.

4.1 The DNA shape

For the shape of the molecule, as seen in figure 4.1, we assume that the tails are given by the homoclinic solutions (eqs. (1.31–1.32)) for $t > t_c \approx 0.80424$. We restrict t since we "attach" the plectoneme at the non-zero point of closest approach of the two tails, which exists only when $t > t_c$. At the point of closest approach the distance between the two tails can be approximated by

$$d_{\text{crit}}(t) = 2\lambda\left(\sqrt{\frac{1-t}{0.3799}} - 0.00112\right). \tag{4.1}$$

Therefore the radius of the plectoneme is given by $R(t) = d_{\text{crit}}(t)/2$. Its path, on the other hand, can be parametrized by

$$\mathbf{r}_p = \begin{pmatrix} (s_0 + s)\sin\alpha \\ -R(t)\cos\left((s_0+s)\frac{\cos\alpha}{R(t)}\right) \\ R(t)\sin\left((s_0+s)\frac{\cos\alpha}{R(t)}\right) \end{pmatrix} \quad \text{for} \quad s \in [0, L_p/2]$$
$$\mathbf{r}_p = \begin{pmatrix} (s_0 + L_p + L_{\text{loop}} - s)\sin\alpha \\ R(t)\cos\left((s_0+L_p+L_{\text{loop}}-s)\frac{\cos\alpha}{R(t)}\right) \\ -R(t)\sin\left((s_0+L_p+L_{\text{loop}}-s)\frac{\cos\alpha}{R(t)}\right) \end{pmatrix} \quad \text{for} \quad s \in [0, L_p/2], \tag{4.2}$$

where α is plectoneme angle (see fig. 4.1) and L_p, L_{loop} are the contour lengths of the plectoneme and of the loop. The starting orientation depends on the homoclinic solution and is set by s_0, chosen so that the tails are attached in a continuous fashion to the plectoneme.

4.2 The writhe

We can first calculate the writhe of the plectoneme by using the tangent of eq (4.2) and the $\hat{x}$-axis in eq. (1.39):

$$\begin{aligned}\omega_1(s) &= \frac{1}{2\pi}\frac{\cos\alpha(\sin\alpha-1)}{R(t)} \quad s\in[0,l_p/2],\\ \omega_2(s) &= \frac{1}{2\pi}\frac{\cos\alpha(\sin\alpha+1)}{R(t)} \quad s\in[l_p/2+l_l,l_p+l_l].\end{aligned} \tag{4.3}$$

This expression neglects end loop and tails; by summing ω_1 and ω_2 one arrives at an "average" writhe density

$$\omega^0(\alpha,t) = \frac{\cos\alpha\sin\alpha}{2\pi R(t)}. \tag{4.4}$$

This expression is very convenient, and it was normally taken to be the writhe density of the plectoneme [49]. However, when we computed the writhe of the tails in eq. (1.35), the reference axis was the $\hat{z}$-axis, parallel to the force F, and not the $\hat{x}$-axis as in the case of eqs. (4.3). To be consistent (the non-locality of the writhe forbids, in fact, to use different reference axes in Fuller formula, eq. (1.39), for different sections of the curve) we compute the writhe of the plectoneme with respect to the $\hat{z}$-axis, resulting in

$$\begin{aligned}\omega_b(s) = \frac{1}{2\pi}\frac{\sin\alpha\cos\alpha}{R(t)}\times\\ \times\left[1-\frac{1}{1+\cos\alpha\cos\left((s+s_0)\frac{\cos\alpha}{R(t)}\right)}\right].\end{aligned} \tag{4.5}$$

This expression is annoying as it is s-dependent. However, while the plectoneme grows, the writhe of the end loop changes, as it changes its orientation. This loop is described by a space curve $\mathbf{r}_0 = (r_x(u), r_y(u), r_z(u))$, $u\in[0,L_{\text{loop}}]$, subject to the conditions, at its boundaries, $\mathbf{r}_0(0) = \mathbf{r}_p(0)$ and $\mathbf{r}_0(L_{\text{loop}}) = \mathbf{r}_p(L_{\text{loop}})$. We also assume (unlike at the boundaries between tails and plectoneme) that the curve is smooth between plectoneme and end loop.

Increasing the contour length by an amount $2s$ causes a rotation of $\mathbf{r}_0$ by an angle $\xi(s) = s\cos\alpha/R(t)$ (see eq. (4.2)) about the $\hat{x}$-axis, inducing an s-dependent change in the writhe of the loop

$$\begin{aligned} Wr_{\text{loop}} =& \frac{1}{2\pi}\int_0^{L_{\text{loop}}} \mathrm{d}u \frac{\cos\xi(s)(t_x(u)\dot{t}_y(u) - \dot{t}_x(u)t_y(u)}{1-\sin\xi(s)t_y(u)+\cos\xi(s)t_z(u)} \\ & -\frac{1}{2\pi}\int_0^{L_{\text{loop}}} \mathrm{d}u \frac{\sin\xi(s)(t_z(u)\dot{t}_x(u) - \dot{t}_z(u)t_x(u)}{1-\sin\xi(s)t_y(u)+\cos\xi(s)t_z(u)}. \end{aligned} \tag{4.6}$$

A change in the plectoneme contour length induces a differential change of this writhe equal to

$$\begin{aligned} \frac{\mathrm{d}Wr_{\text{loop}}}{\mathrm{d}s} =& -\frac{\cos\alpha}{2\pi R(t)}\times \qquad (4.7) \\ & \left(\int_0^{l_l} \mathrm{d}u \frac{\dot{t}_x(u) + \sin\left(s\frac{\cos\alpha}{R(t)}\right)(t_x(u)\dot{t}_y(u) - \dot{t}_x(u)t_y(u))}{\left(1-\sin\left(s\frac{\cos\alpha}{R(t)}\right)t_y(u)+\cos\left(s\frac{\cos\alpha}{R(t)}\right)t_z(u)\right)^2} + \right. \\ & \left. \int_0^{l_l} \mathrm{d}u \frac{\cos\left(s\frac{\cos\alpha}{R(t)}\right)(t_z(u)\dot{t}_x(u) - \dot{t}_z(u)t_x(u))}{\left(1-\sin\left(s\frac{\cos\alpha}{R(t)}\right)t_y(u)+\cos\left(s\frac{\cos\alpha}{R(t)}\right)t_z(u)\right)^2}\right) \\ =& \frac{\cos\alpha}{\pi R(t)} \times \frac{t_x(0)}{1-\sin\left(s\frac{\cos\alpha}{R(t)}\right)t_y(0)+\cos\left(s\frac{\cos\alpha}{R(t)}\right)t_z(0)}, \qquad (4.8) \end{aligned}$$

where we used the unimodularity of the tangent vector and its symmetry: $t_x(0) = -t_x(L_{\text{loop}}), t_{y,z}(0) = t_{y,z}(L_{\text{loop}})$. Making use of the boundary conditions we finally find

$$\frac{\mathrm{d}Wr_{\text{loop}}}{\mathrm{d}s} = \frac{\cos\alpha\sin\alpha}{\pi R(t)}\left(1+\cos\alpha\cos\left((s_0+s)\frac{\cos\alpha}{R(t)}\right)\right)^{-1} \tag{4.9}$$

By adding this differential writhe density to the "bare" writhe density of the plectoneme ((4.5)) (half of it to each strand) we recover the standard writhe density of a plectoneme (4.4), but now with the added bonus that

the remaining writhe of the closing loop is independent of the length of the plectoneme. Since only in the end loop the antipodal points appear along the homotopy, defined by the explicit formation of the plectoneme, we can state that in this sense the writhe is additive:

$$Wr(t,\alpha) = Wr_{\text{loop}}(t) + L_p\omega^0(t,\alpha), \tag{4.10}$$

with Wr_{loop} and ω given by (1.35) and (4.4).

4.3 Mechanical and electrostatic energy

Starting point for the mechanical Hamiltonian of the system is eq. (1.49). Since the *number of turns* is experimentally controlled, we can write

$$H_M = \int_0^{L_c} \mathrm{d}s \left(\frac{A}{2}\dot{\mathbf{t}}_s^2 - \mathbf{f}\cdot\mathbf{t}_s \right) + 2\pi^2 \frac{C}{L_c}\left(n - Wr\right)^2 \tag{4.11}$$

where the writhe Wr is given by eq. (4.10). Using the bending energy of the homoclinic solution, eq. (1.33), and the plectoneme path eq. (4.2), we arrive at

$$E_M = -f(L_c - L_p) + E_{\text{loop}} + E^0_{\text{bend}} + 2\pi^2 \frac{C}{L}\left(n - Wr\right)^2 \tag{4.12}$$

$$E^0_{\text{bend}} = L_p \frac{A}{2}\frac{\cos^4\alpha}{R^2(t)}. \tag{4.13}$$

However, since many experiments are performed at low salt concentration, where the negatively charged DNA is less “screened”, electrostatic interactions change the energy. First the bending persistence length is renormalized according to the OSF theory [52, 63]:

$$l_p = l_p^{(0)} + \frac{\kappa^{-2}}{4Q_B} \tag{4.14}$$

where κ^{-1} is the Debye screening length and Q_B the Bjerrum length

$$Q_B = \frac{q^2}{4\pi\varepsilon_0\varepsilon_r k_B T} \tag{4.15}$$

$$\kappa = \sqrt{8\pi Q_B n_s} \tag{4.16}$$

where q is the elementary charge, ε_0 the vacuum permittivity, ε_r the dielectric constant and n_s the number density of the salt molecules. At $T = 300\,\mathrm{K}$, $Q_B \approx 0.715\,\mathrm{nm}$ and $\kappa = 0.1\sqrt{c_s}$, where c_s is the salt concentration in mM (milliMolar).

In the plectoneme there is another electrostatic effect: the two strands of DNA repeal each other, resulting in an energetic contribution [68]

$$E^0_{\mathrm{el}} = L_p \nu^2_{\mathrm{eff}} \frac{Q_B}{2} \sqrt{\frac{\pi}{\kappa R(t)}} e^{-2\kappa R(t)}\, \mathrm{Z}(\cot\alpha) \tag{4.17}$$

$$\mathrm{Z}(x) = 1 + 0.828x^2 + 0.868x^4$$

if $\cot\alpha < 1$, with ν_{eff} the effective charge density of the centerline of a charged cylinder source of a Debye-Hückel potential that asymptotically coincides in the small potential, far field, region with the non-linear Poisson-Boltzmann potential of that cylinder with a given surface charge (for DNA $2\,e/0.34$ nm, radius 1 nm).

To compute ν_{eff} and R^* as in figure 4.2 we follow [53]. The radius R^* marks the breaking down of the linearized theory, as there the reduced potential equals 1 [53]. The energy E_M changes therefore to

$$E^0 = E_M + E^0_{\mathrm{el}} \tag{4.18}$$

where the superscript indicates that no thermal fluctuations are taken into account up to now and where the persistence length should be taken as in eq. (4.14).

A plectoneme will form when the energy E^0 has a global minimum for $L_p > 0$. Minimization of the energy shows that this *transition point* happens at $n < n_{\mathrm{crit}}$. Moreover the angle α stabilizes between $\pi/2$, where $E_{\mathrm{bend}} = 0$, and $\pi/4$, where Wr is maximized and thus $n - Wr$ is minimized.

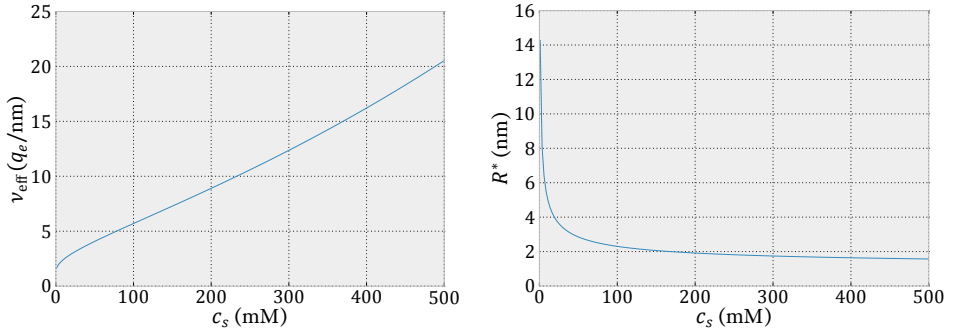


Figure 4.2: ν_{eff} and R^* as a function of the salt concentration c_s.

4.4 Fluctuations

The effect of thermal fluctuations is difficult to treat and complicated to analyze. Without entering in the details, we will explain the origin of each term, referring the reader to the specific papers for the details.

The first contribution from thermal fluctuations is the enhancement of the electrostatic interactions in the plectoneme. There the thermal undulations couple non-linearly with the electrostatic interactions [68]. As a result the interactions are strongly enhanced. In fact assuming the fluctuations to have a Gaussian distribution we can estimate their effect by considering one strand of the plectoneme in the mean field potential of the opposing strand. The undulations in the plectoneme are taken in two directions, along the radius and the pitch. The radial direction is limited by the exponent of E^0_{el} in eq. (4.17), $-2\kappa R(t)$, while in the pitch direction the limit is geometrical (see figure 4.1) so that we will assume that the standard deviation in that direction is fixed and equal to $\sigma_p = \pi R(t) \sin\alpha$. Calling σ_r the standard deviation in the radial direction, the electrostatic interaction becomes

$$E_{\text{el}} = E^0_{\text{el}} e^{4\kappa^2\sigma_r^2}. \tag{4.19}$$

The steep potential limits σ_r to $\approx \kappa^{-1}/2$. Here σ_p is not present as it only affects marginally the electrostatic energy [68]. Confining the DNA in the plectoneme has also an entropic cost. In fact the polymer has a

lower number of configurations as it is restricted in the radial direction by an harmonic potential, and in the pitch direction by a hard wall potential (represented by the consecutive turn of the strand). Burkhardt [6] computed the entropic contribution of such a confinement, but only for the torsion-less case. Recently Emanuel [20] worked out the more difficult case where torque is present. The result is that two deflection lengths appear, one for each confinement direction, $\lambda_{r,p} = (P_b \sigma^2_{r,p})^{1/3}$. The new effective deflection length for the plectoneme as a whole is then

$$\bar{\lambda} = 2\frac{\lambda_r^3\lambda_p + \lambda_r^2\lambda_p^2 + \lambda_r\lambda_p^3}{(\lambda_r + \lambda_p)(\lambda_r^2 + \lambda_p^2)} \tag{4.20}$$

which contributes to a confinement free energy equal to

$$E_C = \frac{3}{8}k_B T(\lambda_r^{-1} + \lambda_p^{-1})L_p. \tag{4.21}$$

Moreover fluctuations inside the plectoneme reduce its contour length by a factor [21]

$$\rho_{\mathrm{pl}} = 1 - \frac{k_B T}{4A}(\lambda_r + \lambda_p) \tag{4.22}$$

that in turn change the bending energy eq. (4.13) and the writhe density of the plectoneme eq. (4.4) to

$$E_{\mathrm{bend}} = E^0_{\mathrm{bend}}\rho^4_{\mathrm{pl}} \tag{4.23}$$

$$\omega(\alpha, t) = \omega^0(\alpha, t)\rho_{\mathrm{pl}}. \tag{4.24}$$

Outside the plectoneme, before the transition point, thermal fluctuations also play a role. In fact, the straight solution $\vartheta, \varphi = 0$ incurs in finite deformation $\mathrm{d}\vartheta, \mathrm{d}\varphi$ on top of it. These deformations, in general, alter the writhe of the chain. As a consequence, in a torsionally constrained setup, the White relation eq. (1.38) implies that the twist in influenced by fluctuations. In practice the torsional persistence length is rescaled to

$$C(\lambda) = \frac{C}{1 + \frac{Ck_BT}{4A\lambda f}} \tag{4.25}$$

where $\lambda = \sqrt{A/f}$ in this case. When computing the torsional energy we should use $C(\lambda)$ instead of C. However there is no reason the λ used in the tails should be reused in the plectoneme. The correct way to do it, in fact, is to use $\bar{\lambda}$ from eq. (4.20) for the torsional energy of the plectoneme. This results in the torsional energy of the system

$$E_T = 2\pi^2 \left(C(\lambda) \frac{Tw_\lambda^2}{(L_c - L_p)} + C(\bar{\lambda}) \frac{Tw_{\bar{\lambda}}^2}{L_p} \right) \tag{4.26}$$

where Tw_λ and $Tw_{\bar{\lambda}}$ are the twist values in the tails and in the plectoneme. The linking number density in plectoneme, $Tw_{\bar{\lambda}}$, and tails, Tw_λ, do not need to be the same. Twist relaxation is fast, as is confirmed by experiments [9]. Since the twist degree of freedom only couples globally, (by means of the White's equation), to the tangential degrees of freedom, we can integrate out the twist fluctuations and simplify the model by equating the twist free energy densities:

$$C(\lambda) \frac{Tw_\lambda^2}{(L_c - L_p)^2} = C(\bar{\lambda}) \frac{Tw_{\bar{\lambda}}^2}{L_p^2}. \tag{4.27}$$

We will use $Tw_\lambda/(L_c - L_p) \equiv tw_\lambda$ as one of the minimization parameters: $Tw_{\bar{\lambda}} \equiv tw_{\bar{\lambda}} L_p$ can be inferred from eq. (4.27). In principle, the end loop should be treated separately from the tails, with yet another λ. However the end loop only affects marginally the straight chain entropic contribution [34], justifying the use of a unique λ for tails and end loop. Therefore when writing Tw_λ and $Wr_{\text{loop}}(t)$ (see eq. (1.35)) we always mean the twist and the writhe of tails and loop together.

When using $Tw_\lambda/(L_c - L_p)$ as a minimization parameter, the length of the plectoneme L_p is given through White's relation eq. (1.38)

$$\begin{aligned} Lk &= (\text{Writhe} + \text{Twist})_{\text{plectoneme}} + (\text{Writhe} + \text{Twist})_{\text{tails}} \\ n &= (\omega(\alpha, t) + tw_{\bar{\lambda}}) L_p + Wr_{\text{loop}} + tw_\lambda (L_c - L_p) \end{aligned} \tag{4.28}$$

from which

$$L_p = \frac{n - Wr_{\text{loop}} - L_c tw_\lambda}{\omega(\alpha, t) + tw_{\bar{\lambda}} - tw_\lambda}. \tag{4.29}$$

Thermal fluctuations in the tails also modify, to lowest order, the $-fL_c$ term in equation (4.12) to

$$E_{\text{tails}} = \left(-f + \frac{k_B T}{\lambda} - \frac{(k_B T)^2}{4A}\right)(L_c - L_p) \tag{4.30}$$

and shorten the end-to-end distance by a factor $\rho_{\text{tail}} = 1 - \lambda k_B T/2A$ according to [47].

The total energy is thus

$$E_{\text{single}} = E_C + E_T + E_{\text{el}} + E_{\text{bend}} + E_{\text{loop}} + E_{\text{tails}} \tag{4.31}$$

4.5 Multi-plectoneme phase

From a purely mechanical point of view the energy cost of the end loop and tails is so high that only one plectoneme will form in the system, its length growing when increasing n. However the prominent role of thermal fluctuations and entropy could increase the contributions of multiple plectonemes, which act in this case as local minima. We call m the number of plectonemes, with total length L_p, given by equation (4.29) with $Wr_{\text{loop}} \to mWr_{\text{loop}}$. The total energy will be

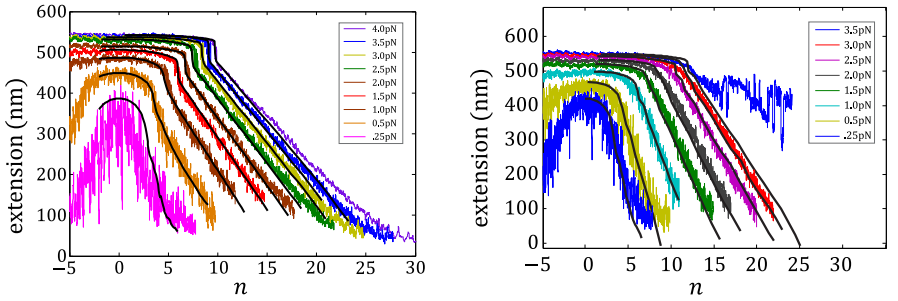
$$E(m) = E_C + E_T + E_{\text{el}} + E_{\text{bend}} + mE_{\text{loop}} + E_{\text{tails}} \tag{4.32}$$

Assuming that L_p grows faster than m, we can neglect E_{loop} in eq. (4.32). This has the advantage that the minimization of the total energy with respect to $\alpha, R(t), \sigma_r$ and tw_λ is m-independent. The partition sum is computed with these values. We choose a hardcore repulsion between plectonemes (for simplicity) and a cutoff $\Lambda = 3.4$ nm (for structural reasons) to calculate[1] the density of states. The resulting partition sum is [21]

$$Z = e^{-E(0)} + \sum_{m=1} G_m e^{-E(m)} \tag{4.33}$$

$$G_m = \frac{\left(\rho_{\text{tail}}(L_c - L - p) - nL_{\text{loop}}\right)^n L_p^{n-1}}{n!\,(n-1)!\,\Lambda^n \Lambda^{n-1}} \tag{4.34}$$

[1]Removing the hardcore repulsion or changing the cutoff in a reasonable range affect the curves below the experimental error.



(a) 320 mM ionic strength. (b) 20 mM ionic strength.

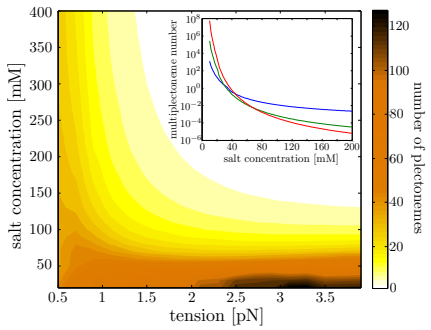
Figure 4.3: Turns-extension plots comparison between the theory and experimental data from [5] for a DNA chain with contour length 600 nm.

where the upper limit is chosen so that $L_p \geq 0$. For long chains the distribution is strongly peaked around an average $\langle m \rangle$. There are 2 ways the extension decreases with increasing linking number, through an increase of plectoneme length and through an increase of the number of plectonemes. At high salt concentrations the single plectoneme configuration becomes the groundstate at finite plectoneme length. The jump as seen in experiments [24, 13] is partly caused by the end loop, partly by the finite size plectoneme. The nature of these configurations differs enough from the former to speak of a multi-plectoneme phase (MP): they affect the slope and the torque after the transition. To characterize the MP we introduce the multi-plectoneme parameter

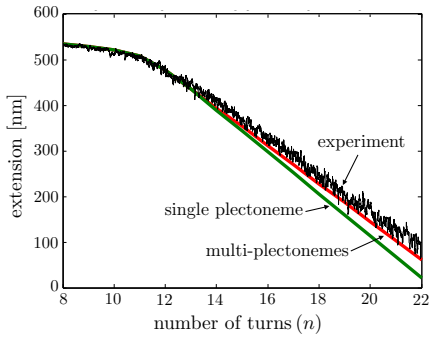
$$\zeta \equiv \exp\left[-\frac{Wr_{\text{loop}}}{k_B T}\left(\frac{E_{\text{loop}}}{Wr_{\text{loop}}} - \frac{\Delta f}{\omega(\alpha,t)}\right)\right]\left(\frac{Wr_{\text{loop}}/L_{\text{loop}}}{\omega(\alpha,t)}\right)^2 \tag{4.35}$$

$$\Delta f = \frac{E_C + E_{\text{el}} + E_{\text{bend}}}{L_p} - \frac{E_{\text{tails}}}{L_c - L_p} \tag{4.36}$$

For $\zeta \ll 1$ the experimental turn-extension plots and torques, are well described by a single plectoneme whereas for $\zeta \approx 1$, the slope is a result of an increase of plectonemes with increasing n. The inset of Fig. 4.4a shows ζ as a function of salt concentration for different tensions.



(a) Phase diagram of the average number of plectonemes as a function of tension and salt concentration for a 7.2 μm long chain. Note the shift of the maximum from low tension at high salt to high tension at low salt. The inset shows the MP parameter vs salt concentration for 1 pN (blue), 2 pN (green) and 3 pN (red).



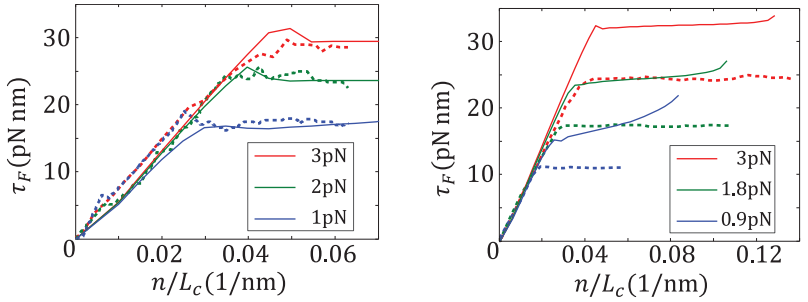
(b) The results of the theory with and without the possibility to form more than one plectoneme are presented alongside the experimental results (3 pN, 20 mM, experimental data from [5]).

Figure 4.4: How many plectonemes and do they make a difference?

4.6 Comparison to experiments

The predicted turn-extension plots of the model agree remarkably well with experiments, see Fig. 4.3. Our model has only two parameters, A and C, both known to some extent from other experiments. The general consensus for A is from 45 to 50nm k_BT. For the numerics we took $A = 50\text{nm}\, k_BT$+ OSF [52] corrections. The value of C influences foremost the transition point. To fit the measurements its value ranges from 100 to 120 nm k_BT. Only for a salt concentration of 20 mM, a lower value of 90 nm k_BT was needed to get the transition point right. Since the plectoneme length starts at 0 at the transition, our approximation of not treating the end loop separately is debatable. Detailed modelling of entropic and electrostatic repulsion within the end loop might improve the model, for example starting from [7], although in the end the proximity of the bifurcation point might invalidate a simple perturbation calculation. For low salt concentrations, older models predict slopes too step [5]. As

shown in Fig. 4.4b the MP phase corrects this picture.



(a) 750 nm DNA chain at 150 mM ionic strength. Comparison between theory and torques directly measured [24].

(b) 5600 nm DNA chain at 100 mM ionic strength. Comparison between theory and inferred torques [48].

Figure 4.5: Predicted versus measured (dashed lines) torque.

In the MP phase the torque of the system is not constant after the transition. This could explain the difference between torques measured in optical tweezer experiments [24] and torques calculated using Maxwell relations in a magnetic tweezer setup [48]. The latter method assumes a constant torque after the transition. However, in the MP phase our theory predicts a non-constant torque. In Fig. 4.5b we show what our model predicts for the data presented in [48]. To facilitate comparison with the original paper, not the linking number, but the supercoiling density is used, defined as the ratio of the linking number density to the linking density of the two strands of free DNA. As can be seen in Fig. 4.5b, the assumption of constant torque underestimates the torque difference between the high and low tension curves. Our model, however, correctly reproduces the direct torque measurements of [24] (see figure 4.5a).

A final consequence of the MP phase is the change in the dynamics of plectonemes. Multiple plectonemes can change their length distribution fast as twist diffusion is fast [9]. This makes a fast diffusion of plectonemes possible also in the crowded environment of the plasmoid in bacteria or through a dense chromatin fiber in eukaryotes. The implications might be important, from cellular processes to transcription to segregation.

Bibliography

[1] A. Arneodo, C. Vaillant, B. Audit, F. Argoul, Y. d'Aubenton Carafa, and C. Thermes. "Multi-scale coding of genomic information: From DNA sequence to genome structure and function". In: *Phys. Rep.* 498.2-3 (2011), pp. 45–188.

[2] A. Bancaud, G. Wagner, N. Conde E Silva, C. Lavelle, H. Wong, J. Mozziconacci, M. Barbi, A. Sivolob, E. Le Cam, L. Mouawad, J.-L. Viovy, J.-M. Victor, and A. Prunell. "Nucleosome chiral transition under positive torsional stress in single chromatin fibers". In: *Mol. Cell.* 27.1 (2007), pp. 135–147.

[3] J. Bednar, R. A. Horowitz, J. Dubochet, and C. L. Woodcock. "Chromatin Conformation and Salt-Induced Compaction - 3-Dimensional Structural Information From Cryoelectron Microscopy". In: *J. Cell. Biol.* 131.6 (1995), pp. 1365–1376.

[4] J. Bednar, R. A. Horowitz, S. Grigoryev, Carruthers, Lenny M, J. C. Hansen, A. J. Koster, and C. L. Woodcock. "Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin". In: *Proc. Natl. Acad. Sci. U.S.A.* 95.24 (1998), pp. 14173–14178.

[5] H. Brutzer, R. Schöpflin, and R. Stehr. "A Theoretical Description of DNA Plectonemes under Tension". In: *Biophys. J.* 98.3 (2010), 470a.

[6] T. Burkhardt. "Free energy of a semiflexible polymer confined along an axis". In: *Journal of Physics. London A Mathematical and General* 28 (1995), pp. L629–L635.

[7] A. G. Cherstvy. “Looping charged elastic rods: applications to protein-induced DNA loop formation”. In: *Eur. Biophys. J. Biophy.* 40.1 (2011), pp. 69–80.

[8] N. Clauvelin, B. Audoly, and S. Neukirch. “Mechanical response of plectonemic DNA: An analytical solution”. In: *Macromolecules* 41.12 (2008), pp. 4479–4483.

[9] A. Crut, D. A. Koster, R. Seidel, C. Wiggins, and N. H. Dekker. “From the Cover: Fast dynamics of supercoiled DNA revealed by single-molecule experiments”. In: *Proc. Natl. Acad. Sci. U.S.A.* 104.29 (2007), pp. 11957–11962.

[10] F. Cui and V. B. Zhurkin. “Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA”. In: *Nucleic Acids Res.* 37.9 (2009), pp. 2818–2829.

[11] Y. Cui and C. Bustamante. “Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure”. In: *Proc. Natl. Acad. Sci. U.S.A.* 97.1 (2000), pp. 127–132.

[12] J. R. Daban and A. Bermudez. “Interdigitated solenoid model for compact chromatin fibers”. In: *Biochemistry-Us* 37.13 (1998), pp. 4299–4304.

[13] B. C. Daniels, S. Forth, M. Y. Sheinin, M. D. Wang, and J. P. Sethna. “Discontinuities at the DNA supercoiling transition”. In: *Phys. Rev. E* 80.4 (2009), p. 040901.

[14] M. Depken and H. Schiessel. “Nucleosome Shape Dictates Chromatin Fiber Structure”. In: *Biophys. J.* 96.3 (2009), pp. 777–784.

[15] P. M. Diesinger and D. W. Heermann. “Average crossing number of Gaussian and equilateral chains with and without excluded volume”. In: *Eur. Phys. J. B* 62.2 (2008), pp. 209–214.

[16] B. Dorigo, T. Schalch, K. Bystricky, and T. J. Richmond. “Chromatin fiber folding: Requirement for the histone H4N-terminal tail”. In: *J. Mol. Biol.* 327.1 (2003), pp. 85–96.

[17] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, and T. J. Richmond. “Nucleosome arrays reveal the two-start organization of the chromatin fiber”. In: *Science* 306.5701 (2004), pp. 1571–1573.

[18] R. P. C. Driessen, H. Meng, R. Shahapure, G. Lanzani, M. Emanuel, U. D. Priyakumar, M. F. White, H. Schiessel, J. van Noort, and R. T. Dame. “Plectoneme formation of double-stranded DNA under torsion”. In: *Nucl. Acids Res.* 41.1 (2013).

[19] J. Dubochet and M. Noll. “Nucleosome Arcs and Helices”. In: *Science* 202.4365 (1978), pp. 280–286.

[20] M. Emanuel. “On the confinement of semiflexible chains under torsion”. In: *In preparation* (2012).

[21] M. Emanuel, G. Lanzani, and H. Schiessel. “Plectoneme formation of double-stranded DNA under torsion”. In: *Submitted* (2012).

[22] J. T. Finch and A. Klug. “Solenoidal model for superstructure in chromatin”. In: *Proc. Natl. Acad. Sci. U.S.A.* 73.6 (1976), pp. 1897–1901.

[23] P. Flory. *Principles of polymer solutions*. Cornell University Press, 1953.

[24] S. Forth, C. Deufel, M. Y. Sheinin, B. Daniels, J. P. Sethna, and M. D. Wang. “Abrupt buckling transition observed during the plectoneme formation of individual DNA molecules”. In: *Phys. Rev. Lett.* 100.14 (2008), p. 148301.

[25] F. B. Fuller. “Decomposition of linking number of a closed ribbon: A problem from molecular biology”. In: *Proc. Natl. Acad. Sci. U.S.A.* 75.8 (1978), pp. 3557–3561.

[26] V. K. Gangaraju and B. Bartholomew. “Mechanisms of ATP dependent chromatin remodeling”. In: *Mutat. Res-Fund. Mol. M.* 618.1-2 (May 2007), pp. 3–17.

[27] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick. “Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions”. In: *Proc. Natl. Acad. Sci. U.S.A.* 106.32 (2009), pp. 13317–13322.

[28] K. Holde van and J. Zlatanova. “Chromatin fiber structure: Where is the problem now?” In: *Semin. Cell. Dev. Biol.* 18.5 (2007), pp. 651–658.

[29] S. C. Kampranis, A. D. Bates, and A. Maxwell. “A model for the mechanism of strand passage by DNA gyrase.” In: *Proc. Natl. Acad. Sci. U.S.A.* 96.15 (July 1999), pp. 8414–8419.

[30] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal. "The DNA-encoded nucleosome organization of a eukaryotic genome". In: *Nature* 458.7236 (2009), 362–U129.

[31] N. Kepper, D. Foethke, R. Stehr, G. Wedemann, and K. Rippe. "Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation". In: *Biophys. J.* 95.8 (2008), pp. 3692–3705.

[32] O. Kratky and G. Porod. "Rontgenuntersuchung Geloster Fadenmolekule". In: *Recl. Trav. Chim. Pay. B.* 68.12 (1949), pp. 1106–1122.

[33] M. Kruithof, F.-T. Chien, A. Routh, C. Logie, D. Rhodes, and J. v. Noort. "Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber". In: *Nat. Struct. Mol. Biol.* 16.5 (2009), pp. 534–540.

[34] I. M. Kulic, H. Mohrbach, R. Thaokar, and H. Schiessel. "Equation of state of looped DNA". In: *Phys. Rev. E* 75.1 (2007), p. 011913.

[35] I. M. Kulic and H. Schiessel. "Chromatin dynamics: Nucleosomes go mobile through twist defects". In: *Phys. Rev. Lett.* 91.14 (2003), p. 148103.

[36] I. M. Kulic and H. Schiessel. "DNA spools under tension". In: *Phys. Rev. Lett.* 92.22 (2004), p. 228101.

[37] A. Lesne and J.-M. Victor. "Chromatin fiber functional organization: Some plausible models". In: *Eur. Phys. J. E* 19.3 (2006), pp. 279–290.

[38] R. Levien. *The elastica: a mathematical history*. Tech. rep. UCB/EECS-2008-103. EECS Department, University of California, Berkeley, 2008. URL: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-103.html`.

[39] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. "Crystal structure of the nucleosome core particle at 2.8 angstrom resolution". In: *Nature* 389.6648 (1997), pp. 251–260.

[40] K. Maeshima, S. Hihara, and M. Eltsov. "Chromatin structure: does the 30-nm fibre exist in vivo?" In: *Curr. Opin. Cell Biol.* 22.3 (June 2010), pp. 291–297.

[41] V. Makarov, S. Dimitrov, V. Smirnov, and I. Pashev. "A Triple Helix Model for the Structure of Chromatin Fiber". In: *Febs. Lett.* 181.2 (1985), pp. 357–361.

[42] B. Mergell, R. Everaers, and H. Schiessel. "Nucleosome interactions in chromatin: fiber stiffening and hairpin formation". In: *Phys. Rev. E* 70.1 (2004), p. 011915.

[43] S. Mihardja, A. J. Spakowitz, Y. Zhang, and C. Bustamante. "Effect of force on mononucleosomal dynamics". In: *Proc. Natl. Acad. Sci. U.S.A.* 103.43 (2006), pp. 15871–15876.

[44] W. Möbius and U. Gerland. "Quantitative Test of the Barrier Nucleosome Model for Statistical Positioning of Nucleosomes Up- and Downstream of Transcription Start Sites". In: *PLoS Comput. Biol.* 6.8 (Aug. 2010), e1000891.

[45] F. Mohammad-Rafiee, I. M. Kulic, and H. Schiessel. "Theory of nucleosome corkscrew sliding in the presence of synthetic DNA ligands." In: *J. Mol. Biol.* 344.1 (Nov. 2004), pp. 47–58.

[46] L. Mollazadeh-Beidokhti, F. Mohammad-Rafiee, and H. Schiessel. "Nucleosome Dynamics between Tension-Induced States". In: *Biophys. J.* 102.10 (2012), pp. 2235–2240.

[47] J. D. Moroz and P. Nelson. "Entropic Elasticity of Twist-Storing Polymers". In: *Macromolecules* 31.18 (Sept. 1998), pp. 6333–6347.

[48] F. Mosconi, J. F. Allemand, D. Bensimon, and V. Croquette. "Measurement of the Torque on a Single Stretched and Twisted DNA Using Magnetic Tweezers". In: *Phys. Rev. Lett.* 102.7 (Feb. 2009), p. 078301.

[49] S. Neukirch. "Extracting DNA twist rigidity from experimental supercoiling data". In: *Phys. Rev. Lett.* 93.19 (2004), p. 198107.

[50] S. Neukirch and J. F. Marko. "Analytical Description of Extension, Torque, and Supercoiling Radius of a Stretched Twisted DNA". In: *Phys. Rev. Lett.* 106.13 (2011), p. 138104.

[51] M. Nizette and A. Goriely. "Towards a classification of Euler–Kirchhoff filaments". In: *J. Math. Phys.* 40.6 (1999), pp. 2830–2866.

[52] T. Odijk. "Polyelectrolytes near the rod limit". In: *J. Polym. Sci. Polym. Phys. Ed.* 15.3 (Mar. 1977), pp. 477–483.

[53] J. Philip and R. A. Wooding. “Solution of the Poisson–Boltzmann Equation about a Cylindrical Particle”. In: *The Journal of Chemical Physics* 113.20 (1970), pp. 953–959.

[54] C. Rivetti, M. Guthold, and C. Bustamante. “Wrapping of DNA around the E.coli RNA polymerase open promoter complex.” In: *Embo. J.* 18.16 (Aug. 1999), pp. 4464–4475.

[55] P. J. J. Robinson, L. Fairall, A. T. van Huynh, and D. Rhodes. “EM measurements define the dimensions of the ”30-nm” chromatin fiber: Evidence for a compact, interdigitated structure”. In: *Proc. Natl. Acad. Sci. U.S.A.* 103.17 (2006), pp. 6506–6511.

[56] A. Routh, S. Sandin, and D. Rhodes. “Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure”. In: *Proc. Natl. Acad. Sci. U.S.A.* 105.26 (July 2008), pp. 8872–8877.

[57] J. SantaLucia. “A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics”. In: *Proc. Natl. Acad. Sci. U.S.A.* 95.4 (Feb. 1998), pp. 1460–1465.

[58] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. “X-ray structure of a tetranucleosome and its implications for the chromatin fibre”. In: *Nature* 436.7047 (2005), pp. 138–141.

[59] M. P. Scheffer, M. Eltsov, and A. S. Frangakis. “Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei”. In: *Proc. Natl. Acad. Sci. U.S.A.* 108.41 (2011), pp. 16992–16997.

[60] H. Schiessel, W. M. Gelbart, and R. Bruinsma. “DNA folding: structural and mechanical properties of the two-angle model for chromatin”. In: *Biophys. J.* 80.4 (2001), pp. 1940–1956.

[61] M. Shogren-Knaak. “Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions”. In: *Science* 311.5762 (Feb. 2006), pp. 844–847.

[62] A. Sivolob and A. Prunell. “Linker histone-dependent organization and dynamics of nucleosome entry/exit DNAs”. In: *J. Mol. Biol.* 331.5 (2003), pp. 1025–1040.

[63] J. Skolnick and M. Fixman. “Electrostatoc Persistence Length of a wormlike polyelectrolyte”. In: *Macromolecules* 10.5 (1977), pp. 944–948.

[64] J. Starostin E. L. edited by Calvo, K. Millett, E. Rawdon, and A. Stasiak. "Physical and numerical models in knot theory including applications to the life sciences". In: World Scientific Publishing: Singapore, 2005. Chap. 26.

[65] B. Sudhanshu, S. Mihardja, E. F. Koslover, S. Mehraeen, C. Bustamante, and A. J. Spakowitz. "Tension-dependent structural deformation alters single-molecule transition kinetics." In: *Proc. Natl. Acad. Sci. U.S.A.* 108.5 (Feb. 2011), pp. 1885–1890.

[66] S. H. Syed, D. Goutte-Gattat, N. Becker, S. Meyer, M. S. Shukla, J. J. Hayes, R. Everaers, D. Angelov, J. Bednar, and S. Dimitrov. "Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome". In: *Proc. Natl. Acad. Sci. U.S.A.* 107.21 (2010), pp. 9620–9625.

[67] O. V. Tsodikov, R. M. Saecker, S. E. Melcher, M. M. Levandoski, D. E. Frank, M. W. Capp, and M. T. Record. "Wrapping of flanking non-operator DNA in lac repressor-operator complexes: implications for DNA looping." In: *J. Mol. Biol.* 294.3 (Dec. 1999), pp. 639–655.

[68] J. Ubbink and T. Odijk. "Electrostatic-Undulatory Theory of Plectonemically Supercoiled DNA". In: *Biophys. J.* 76.5 (1999), pp. 2502–2519.

[69] C. Vaillant, L. Palmeira, G. Chevereau, B. Audit, Y. d'Aubenton Carafa, C. Thermes, and A. Arneodo. "A novel strategy of transcription regulation by intragenic nucleosome ordering". In: *Genome Res.* 20.1 (2010), pp. 59–67.

[70] J. White. "Self-linking and Gauss-integral in higher dimensions". In: *Am. J. Math.* 91.3 (1969), pp. 693–728.

[71] J. Widom and A. Klug. "Structure of the 300A Chromatin Filament:X-Ray Diffraction from Oriented Samples". In: *Cell* 43.1 (1985), pp. 207–213.

[72] S. P. Williams, B. D. Athey, L. J. Muglia, R. S. Schappe, A. H. Gough, and J. P. Langmore. "Chromatin Fibers Are Left-Handed Double Helices with Diameter and Mass Per Unit Length That Depend on Linker Length". In: *Biophys. J.* 49.1 (1986), pp. 233–248.

[73] H. Wong, J.-M. Victor, and J. Mozziconacci. “An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length”. In: *PLoS One* 2.9 (2007).

[74] C. L. Woodcock, S. A. Grigoryev, R. A. Horowitz, and N. Whitaker. “A chromatin folding model that incorporates linker variability generates fibers resembling the native structures”. In: *Proc. Natl. Acad. Sci. U.S.A.* 90.19 (1993), pp. 9021–9025.

[75] A. Worcel, S. Strogatz, and D. Riley. “Structure of Chromatin and the Linking Number of Dna”. In: *P. Natl. Acad. Sci-Biol.* 78.3 (1981), pp. 1461–1465.

Samenvatting

Gimli, het beroemde personage uit Tolkien's meesterwerk, is zo klein als de kleinste 10% van de volwassen wereldpopulatie. En toch is de totale hoeveelheid DNA in een van zijn cellen, wanneer uitgestrekt, langer dan bijna ieder mens. Dit feit is nog opmerkelijker wanneer we ons realiseren dat zijn cellen niet groter zijn dan één miljoenste meter.

Wanneer dingen zeer klein worden, wordt de visualisatie van wat er gebeurt sterk bemoeilijkt. Gelukkig lijkt de verdichting van DNA op de verdichting die het "spul" dat we vinden in boeken ondergaat wanneer het gedrukt wordt. Namelijk, zonder de stroom van woorden op te breken in regels, pagina's, boeken en boekenplanken, zouden we moeten joggen tijdens het lezen. Hetzelfde principe geldt voor DNA: een mechanisme is nodig om de genetische informatie op een efficiënte manier (zonder "joggen" dus) uit het polymeer af te lezen.

Ik schrijf "het lezen van genetische informatie", want DNA heeft nog meer overeenkomsten met een boek: zoals boeken "informatie" bevatten, bevat DNA de instructies (informatie dus) om eiwitten te produceren. Dit is mogelijk dankzij de vier letters, ATCG, de voornaamste bouwstenen van DNA. De letters worden in drietallen gelezen, wat 64 mogelijke combinaties oplevert. Elk van die combinaties zijn gekoppeld aan de 20 aminozuren, de bouwblokken van eiwitten.

De verdichting van DNA is echter niet eenvoudig, omdat deze molecuul keten zich gedraagt als een semi-flexibel polymeer. Een polymeer is een molecuul dat bestaat uit duizenden of miljoenen gelijke eenheden,

de monomeren. In de meeste gevallen kunnen we de algemene eigenschappen van een polymeer beschrijven als een willekeurige wandeling op een vierkant rooster (hoewel polymeren niet twee-dimensionaal zijn), waarbij het exacte model niet van belang is: het hoge aantal monomeren, en dus configuraties, maakt de details van de interactie tussen opeenvolgende monomeren irrelevant. De algemene interactie tussen monomeren is veel meer van belang. Bijvoorbeeld wanneer de monomeren elkaar niet aantrekken en elkaar mogen overlappen, hebben we een "ideaal" polymeer. In het geval dat de monomeren elkaar niet mogen overlappen, en de monomeren elkaar niet aantrekken, spreken we van een "gezwollen" polymeer, omdat de kwadratische eind-tot-eind afstand groter is dan in het "ideale geval". Aan de andere kant van het spectrum vinden we ineengestorte polymeren, waar de monomeren elkaar aantrekken waardoor de gemiddelde kwadratische eind-tot-eind afstand kleiner wordt. Voor lengtes die in dit proefschrift beschouwd worden, behandelen we het DNA als een ideaal polymeer, alhoewel de monomeren elkaar niet kunnen overlappen. Deze vereenvoudiging is mogelijk omdat het DNA molecuul relatief stijf is op een lengte schaal die veel groter dan zijn eigen diameter. De informatie over de richting van het molecuul gaat verloren na ongeveer 50 nm, de buigings persistentie lengte. Bovendien bestaat er ook een torsionele persistentie lengte, ≈ 100 nm; voorbij die lengte raakt de torsionele staat van het molecuul verloren. DNA is ook te vergelijken met een waterslang. Net zoals DNA verzet een waterslang zich tegen buiging en torsionele vervorming. Dit is een krachtige analogie, omdat een waterslang niets anders is dan een *elastica*: zijn 3D pad kan worden beschreven door een draaiende tol, dankzij de bewegings-analogie van Kirchoff.

Met behulp van deze analogie kunnen we een eenvoudig model bouwen om de structuur van de chromatine vezel vast te stellen. De chromatine vezel helpt het DNA te comprimeren, zodat het in de cel past. Deze vezel bestaat uit nucleosomen en onbedekt DNA. De nucleosomen vormen een complex, opgebouwd uit DNA en eiwitten. Gebruik makend van de Kirchoff analogie kunnen we de energie van onbedekt DNA vast stellen, en daarmee de structuur van de chromatine vezel. De verdichting van DNA is echter niet het complete verhaal. Om de genetische code te lezen is het nodig om het DNA los te wikkelen van de nucleosomen. Dit kan gedaan worden door het uitoefenen van kracht op de nucleosomen

of door een combinatie van kracht en torsie. Het is daarom logisch om aan het eind van dit proefschrift te kijken naar het effect van kracht en torsie op onbedekt DNA. Dit is eerder theoretisch onderzocht, maar de resultaten kwamen niet altijd overeen met experimentele data, vooral bij lage zout concentraties. Uiteindelijk blijkt dat het ontstaan van meerdere plectonemen, een geometrische configuratie van het molecuul dat efficiënt torsiespanning kan opheffen, de sleutel is om theorie en experiment op elegante wijze met elkaar in overeenstemming te brengen.

Publications

Out of register: how DNA determines the chromatin fiber geometry
G. Lanzani and H. Schiessel
Europhys. Lett. **97**, 38002-1-6 (2012).

Nucleosome response to tension and torque
G. Lanzani and H. Schiessel
Europhys. Lett., **100**, 48001 (2012).

Multi-plectoneme phase of double stranded DNA under torsion
M. Emanuel, G. Lanzani and H. Schiessel
Submitted to Phys. Rev. E.

Crenarchaeal chromatin proteins Cren7 and Sul7 compact DNA by inducing rigid bends
R. P. C. Driessen, H. Meng, R. Shahapure, G. Lanzani, M. F. White, H. Schiessel, J. van Noort and R. T. Dame
Nucl. Acids Res., **41** (2013).

Curriculum vitæ

On the 8th of March, 1984, I was born in Padua, Italy, where I completed secondary education at *Liceo scientifico Alvise Cornaro* in 2003.

In the same year I started studying physics at *Università degli studi di Padova*, where I became *dottore* in 2006 with the thesis *Poincaré, Einsten, special relativity* under the supervision of prof. dr. Pieralberto Marchetti. There I investigated how much of Einstein's work had already been done by Poincaré, many years before.

In 2006 I moved to the Netherlands, to Leiden, where I graduated as Master of Science in theoretical physics on the thesis *A phase-field model for supercooled glycerol* under the supervision of Prof. dr. ir. Wim van Saarloos in 2008.

In 2008 I started a PhD research project under supervision of Prof. dr. Helmut Schiessel at the Instituut-Lorentz for theoretical physics, which is part of the Leiden Institute of Physics at Leiden University. During this time I was a teaching assistant for the courses *Statistische en Thermische Fysica 2* and *Statistical Physics* by Prof. dr. Helmut Schiessel.

Since 2008 I am happily married to Giuditta and, as of the writing of this curriculum, I am the proud father of Sofia, Pietro and Elena. Besides that I work for KPMG as an advisor.

Acknowledgments

Helmut is smart, patient and insightful. Thanks for being The. Most. Amazing. Boss. Ever.

I wish to thank my wife and kids for allowing me to do unpaid overtime to finish this thesis, and for many other things I am not supposed to write here.

I thank my parents, my brother and my sister for their part in shaping the curious mind that I have and, again, for many other things.

I had fruitful discussions with (without any particular order) Alessandro Broggio, Patrick Rebeschini, Jonathan Edge, Marc Emanuel, Nima Hamedani Radja, Anton Akhmerov, Carlo Beenakker, Ruslan Sepkhanov, Christopher Groth, Peter Prinsen, Wim van Saarloos, Harry Linders, Andrej Mesaros and Zorana Zeravcic.

A big thank you to Raoul Schram for translating the *Samenvatting*.

I've visited many institutes during my career, but nothing beats the Instituut-Lorentz in Leiden for the openness, stimulating work environment and friendliness. There are many factors that make such a great institute, but I thank Fran and Marianne for being among the top contributors.

Index