DNA Mechanics Inside Plectonemes, Nucleosomes and Chromatin Fibers

Lanzani, G.

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

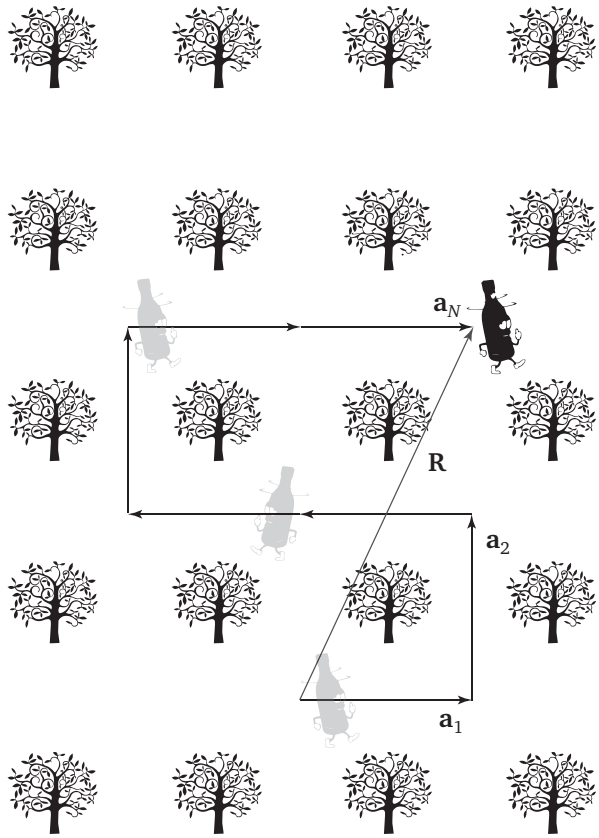


Figure 1.1: A drunk wanderer in a Dutch wood, i.e. a wood of equally spaced and perfectly equal trees. Without an external force, there is a high probability that the wanderer will walk randomly in the wood.

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

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

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

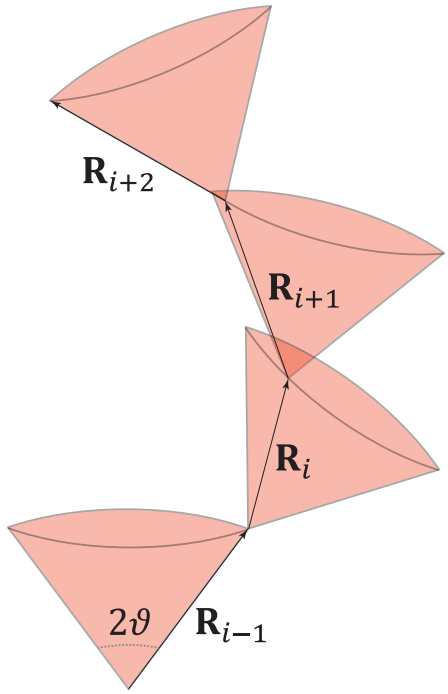


Figure 1.2: A section of the freely rotating chain. Each **R** lies on the surface of its own cone.

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.

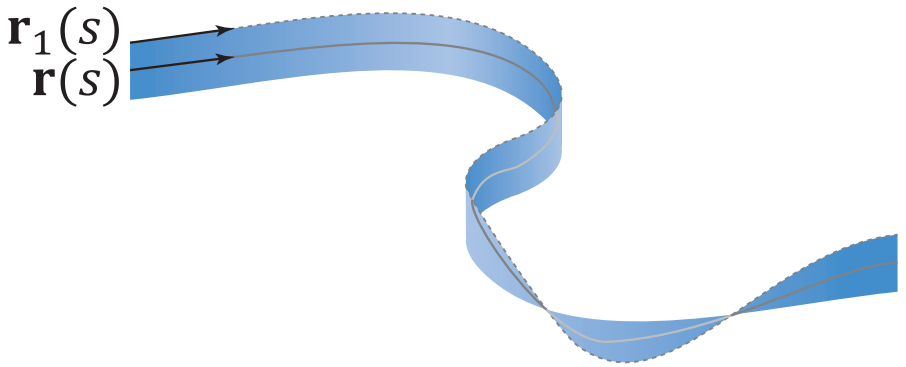


Figure 1.3: An example of a ribbon.

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

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

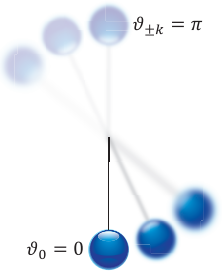


Figure 1.4: The revolving pendulum with the relevant boundary conditions.

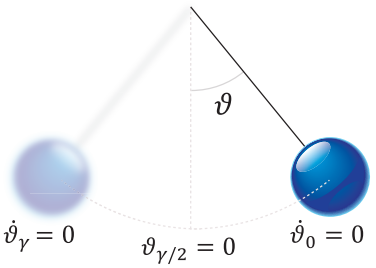


Figure 1.5: The oscillating pendulum with the relevant boundary conditions. Here γ is half the pendulum's period.

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

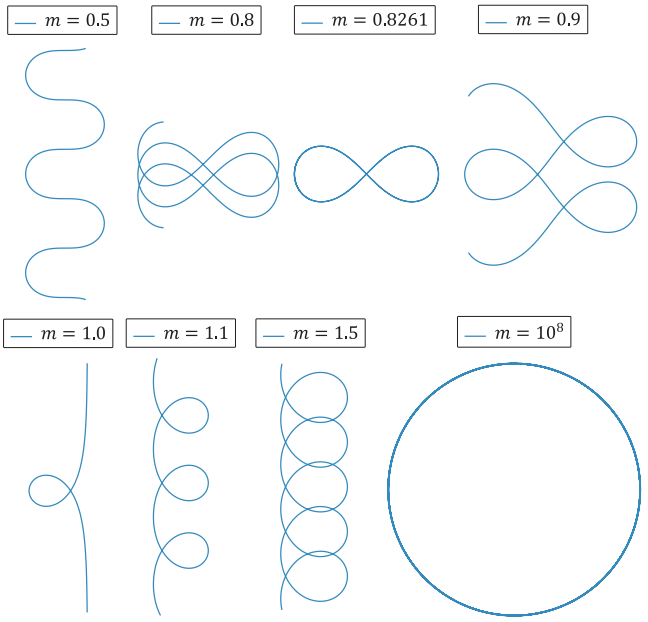


Figure 1.6: The various orbits obtained through integration of eqs. (1.28–1.30). The plot for different m are not in scale.

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.

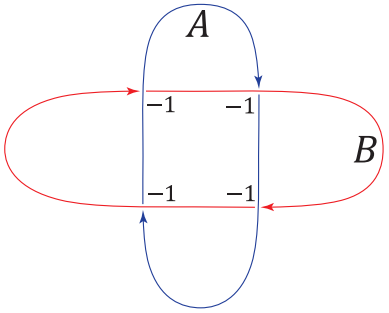


Figure 1.7: An example of two curves with $Lk = -4$.

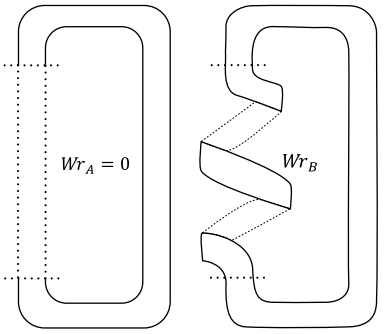


Figure 1.8: An example where Fuller equation can be fruitfully applied.

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

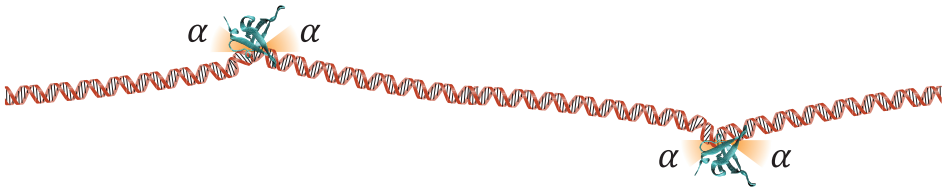


Figure 1.9: How proteins that enforce an angle on a polymer binds to DNA. The situation is analogous to the experiments performed in [18].

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 $\hat{T}$:

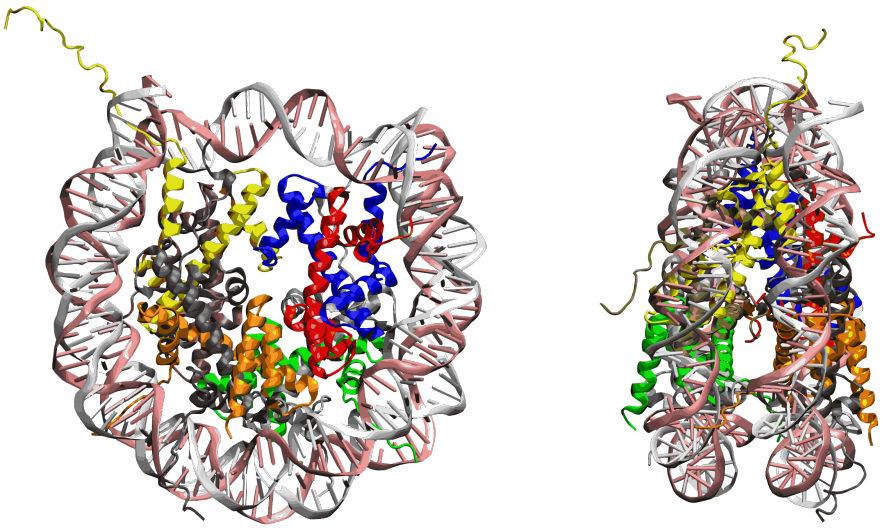
$$(\det \hat{T})|_{k_m} = \frac{\tau_F}{4A^2}\left(AF - \frac{\tau_F^2}{4}\right). \tag{1.55}$$

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

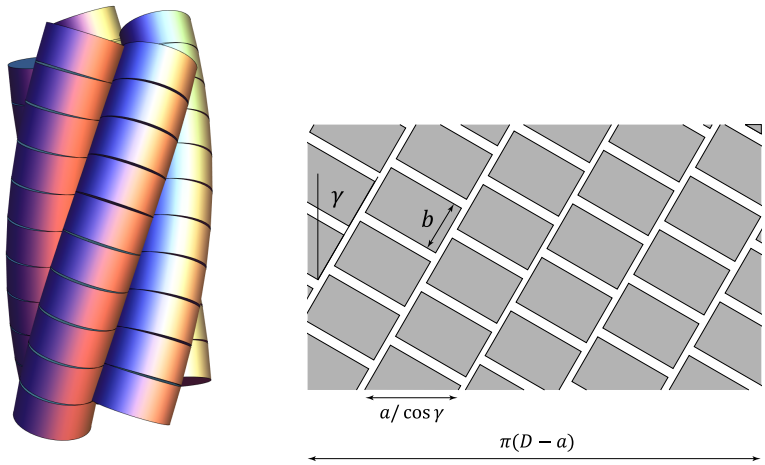
Figure 1.10: The geometry of the nucleosome [39].

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-

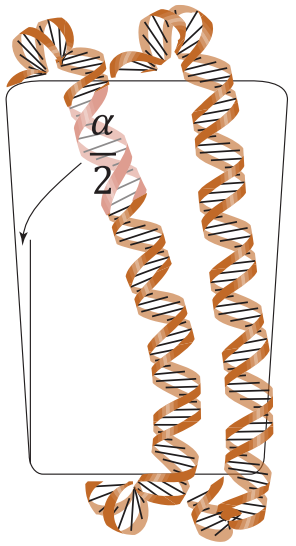


Figure 1.12: A side view of the nucleosome, showing the precise meaning of α.

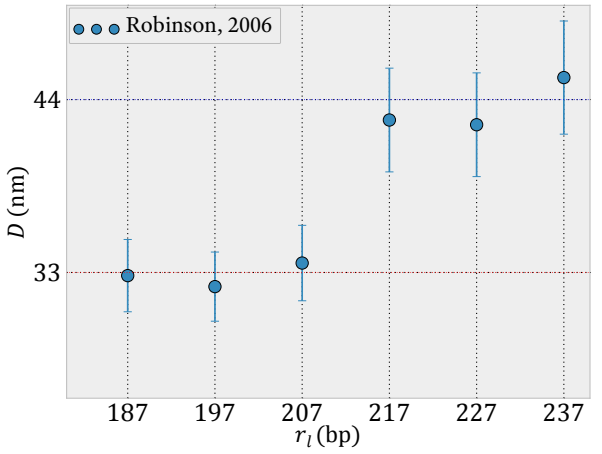


Figure 1.13: The diameters of the chromatin fiber as presented in [55].

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

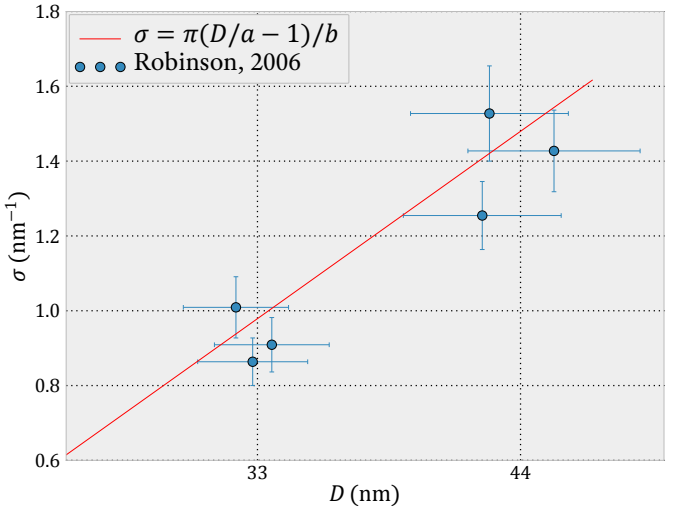


Figure 1.14: A comparison between the experimental data in [55] and equation (1.57).

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

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.

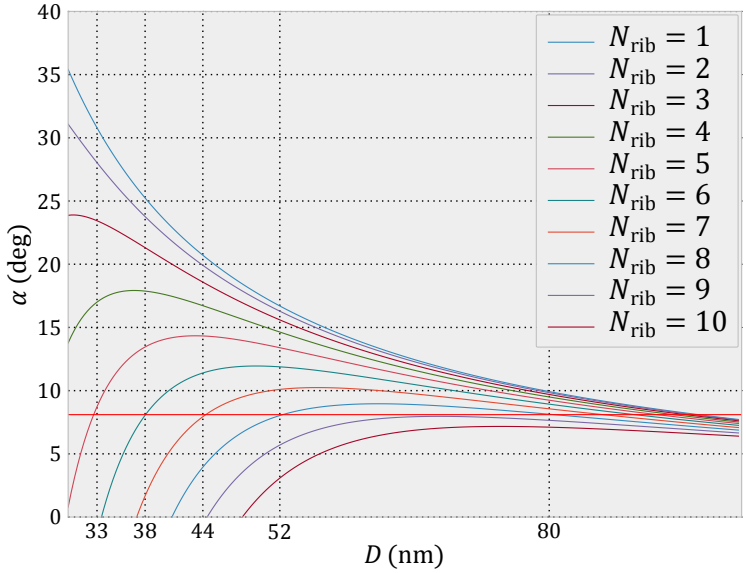


Figure 1.15: A plot of equation (1.59) and of $\alpha \approx 8°$. The intersections between the straight line and the curves mark the diameters compatible with the measured value of α.

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

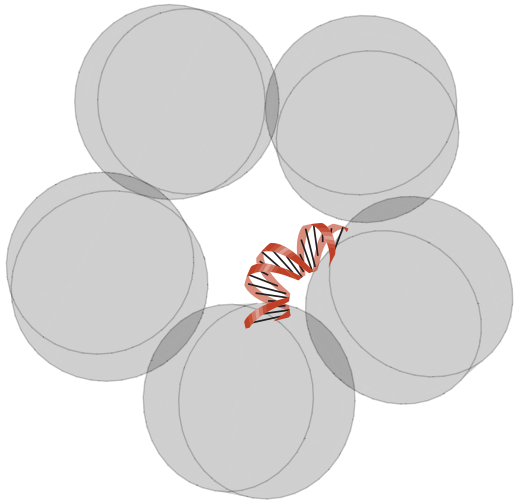


Figure 1.16: The (5, 1) BW.

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.