Introduction to polymer physics

Salvatore Assenza

October 10, 2024

1 Introduction

This document is meant as an introduction to polymer physics aimed at its immediate employment in relevant biophysical problems. Only a few topics are considered, namely the equilibrium conformations of a single polymer under a variety of conditions. The treatment is mostly inspired by the book "Polymer Physics" by M. Rubinstein and R. Colby, although some other elements are present. Although as fundamental as the present selection, topics such as the dynamics of polymers or the physics of concentrated solutions (semidilute regime and polymer melts) are not treated. The textbooks listed at the end of the document may serve as a relevant source to study such topics.

2 Basic definitions

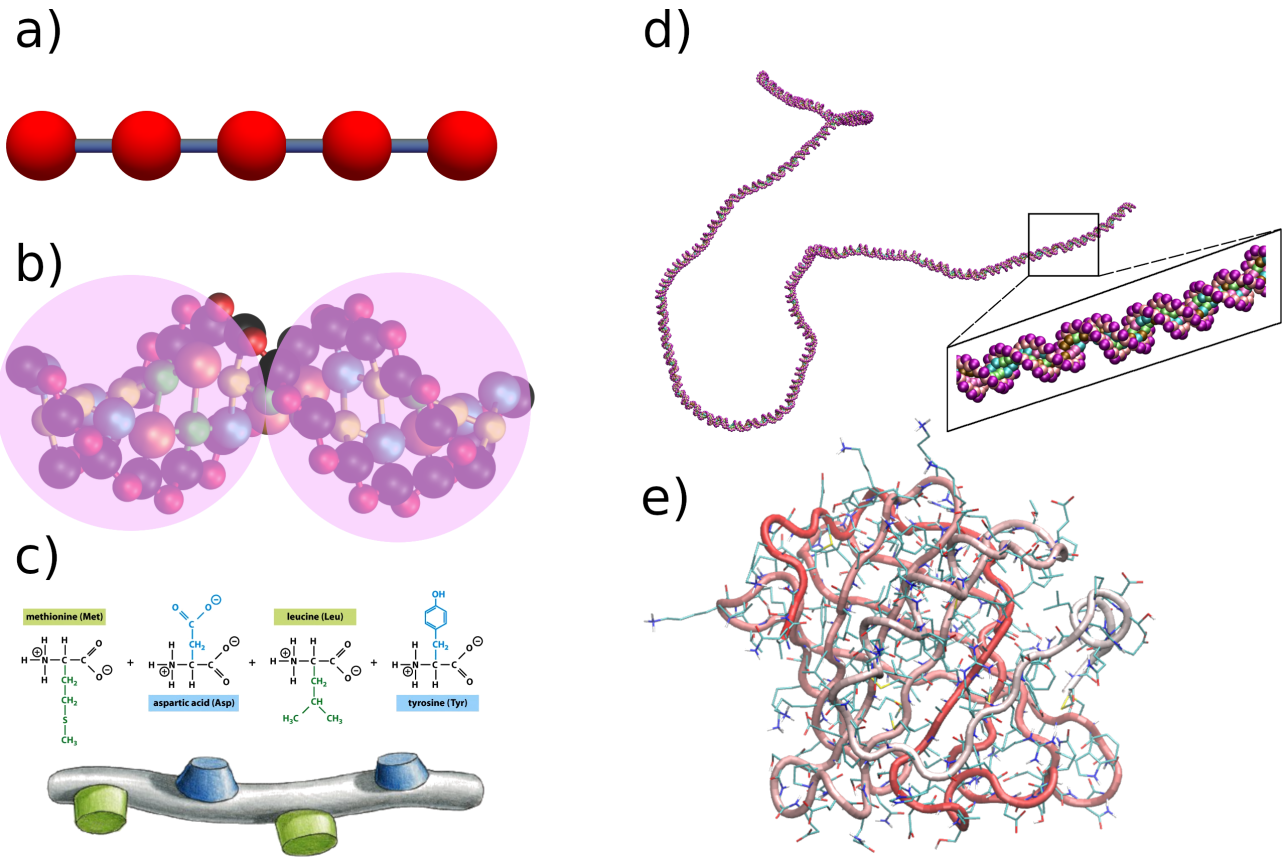

Figure 1: a) Schematic definition of a polymer. b) Short fragment of DNA and corresponding coarse-grained polymer description. c) Representative amino acids and sketch of the protein chain obtained as a backbone with protruding side chains. d) Large-scale snapshot of a DNA conformation, showing the polymer nature of DNA at this scale. e) Representative native structure of a protein, where the line followed by the backbone is highlighted as a tubular string. Panel c was adapted from Alberts et al., "Molecular Biology of the Cell".

By definition, a *polymer* is a macromolecule obtained by attaching basic units to each other as a chain (see Fig.1a). The basic units are called *monomers*. Although in Fig.1a we show for simplicity a polymer where all the monomers are equal to each other, this is not necessarily the case. If all the monomers are equal to each other, one can stress this aspect by means of the word *homopolymer*, which is opposed to *heteropolymer*, used in the case in which monomers are chemically different from each other.

Notable examples of polymers in biophysics are provided by the most important macromolecules in our cells, namely DNA and proteins. In the case of DNA, the microscopic struc-

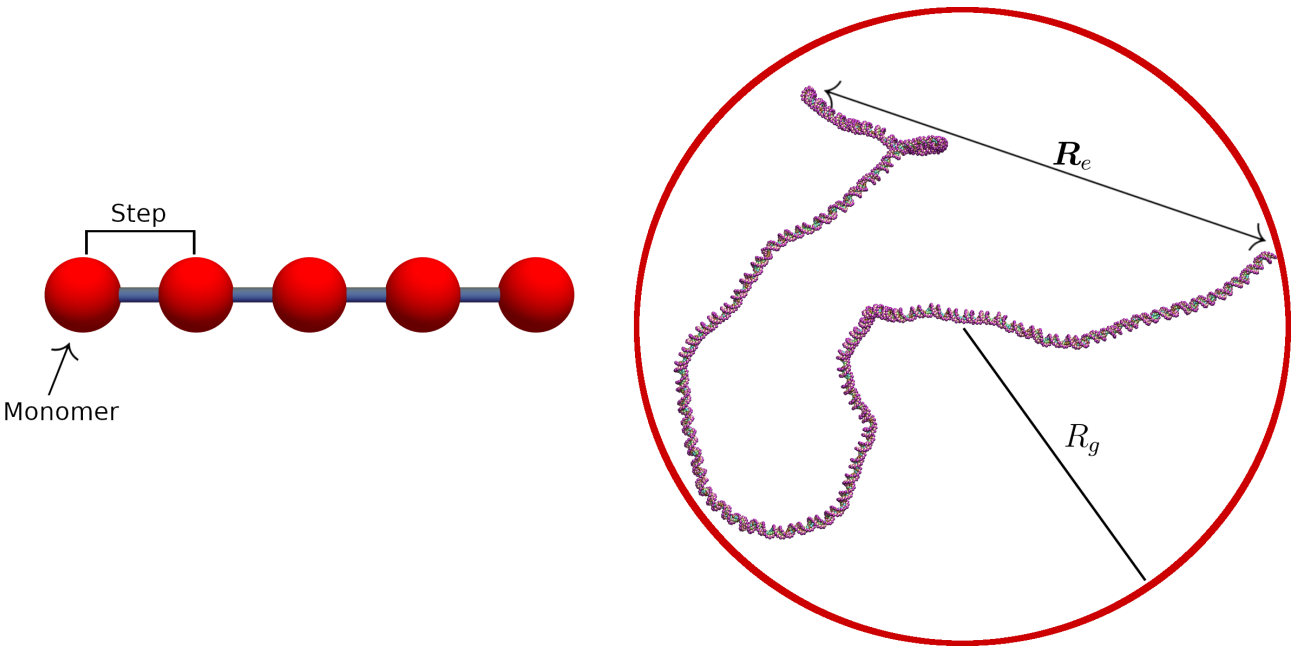

Figure 2: Basic definitions to describe the conformation of a polymer.

ture is more complex, since one has two strands joined to each other from which the bases are protruding towards the inner part of the double helix. Nonetheless, as shown in Fig.1b one can depict the DNA as a polymer by means of a suitable coarse grain (transparent circles), which can provide a useful description of its large-scale properties. In the case of proteins, the monomers are the amino acids, which join each other via a peptidic bond and thus form a backbone from which the side chains of the amino acids are protruding (Fig.1c). In Fig.1d and Fig.1e we show the most salient features of DNA and proteins at the whole-molecule level. Particularly, it becomes evident from Fig.1d that many properties of DNA at this scale can be captured by ascribing it to a string-like object. In the case of a protein, the situation is complicated by the presence of side chains, which interact with each other and with the surrounding water inducing the protein to close into its native structure. In Fig.1e, we show as an example the case of β-lactoglobulin in its native structure, where on the chemical structure we drew the line followed by the backbone as a tubular filament. The color of the tube ranges from white to red in order to ease following the backbone along its contour length. As shown in Fig.2 left, we define as *step* or *bond* the line joining two consecutive monomers. By considering a polymer made of N steps and $N+1$ monomers, we indicate the position of each monomer as $\boldsymbol{r}_i$, with $i = 0, \ldots, N$. The step joining monomers i and $i+1$ is identified by the versor

$$\widehat{\boldsymbol{t}}_i \equiv \frac{\boldsymbol{r}_i - \boldsymbol{r}_{i-1}}{|\boldsymbol{r}_i - \boldsymbol{r}_{i-1}|} \,, \tag{1}$$

which is called *tangent vector* (slighty forcing the nomenclature).

The overall size of the polymer is usually captured by two different quantities. The *end-to-end vector* $\boldsymbol{R}_e$ is the vector joining the two ends of the polymer (see Fig.2 right),

$$\boldsymbol{R}_e \equiv \boldsymbol{r}_N - \boldsymbol{r}_0 \ , \tag{2}$$

and its modulus R_e is called the *end-to-end distance*. We note that R_e might be not very informative about the overall size of the volume filled by the polymer. In order to provide this information, one also defines the *radius of gyration* R_g as

$$R_g^2 \equiv \frac{1}{N+1} \sum_{i=0}^{N} \left(\boldsymbol{r}_i - \boldsymbol{r}_{CM}\right)^2 \ , \tag{3}$$

where

$$\boldsymbol{r}_{CM} \equiv \frac{1}{N+1} \sum_{i=0}^{N} \boldsymbol{r}_i \tag{4}$$

is the position of the center of mass of the polymer[1]. R_g^2 corresponds to the variance of the monomer distribution in space, so that it gives an idea of the typical volume filled by the polymer.

As for any other thermodynamic system, the conformation of a polymer is determined by the combination of the translational entropy of the monomers and their interactions both with each other and with the surrounding solvent. As for the monomer-monomer interactions, it is worth outlining the presence of the so-called *bonded interactions*, which involve monomers close along the backbone of the polymer (Fig.3). Particularly, the bond between consecutive monomers usually has a preferred length, so that it costs energy to displace it from equilibrium (*bond interaction*). Analogously, three consecutive monomers usually display a preferred angular disposition, which can be perturbed only at some energetic cost (*angle interaction*). When four consecutive monomers are considered, there is a tendency for the fourth one to be placed at a certain angle with respect to the plane formed by the first three, giving rise to *dihedral interactions*. In principle one could go on by including five-body interactions, six-body interactions, etc. Nevertheless, this level of approximation is usually enough to describe biomolecules. Apart from the bonded interactions, there are also the so-called *non-bonded interactions* between any two monomers which are close in the euclidean sense (although they might be far from each other in terms of contour length). For instance, one can have electrostatic interactions, van der Waals interactions, excluded-volume interactions, hydrogen bonds, etc. Finally, also the interaction with the surrounding solvent has to be taken into account. Such a complex picture poses great challenges in the theoretical description of polymers. Nevertheless, in spite of such complexity there are many features that can be described by means of very few ingredients, as we will see below.

[1]To simplify the notation, we assumed that all monomers have the same mass m.

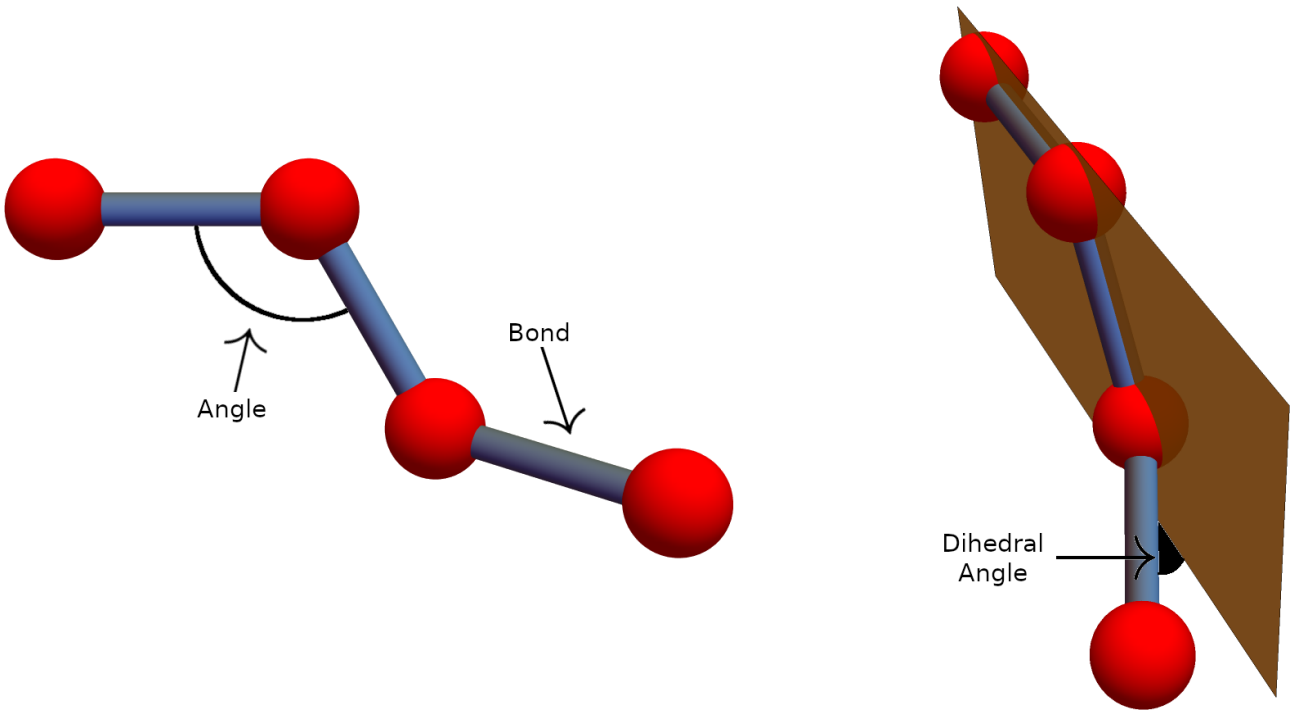

Figure 3: Pictorial description of the bonded interactions within a polymer.

3 Freely-Jointed Chain

The *Freely-Jointed Chain* (FJC) is the minimal model of a polymer. In a FJC, all the non-bonded interactions are neglected. As for the bonded interactions, also angles and dihedrals are neglected. The bonds are introduced in the simplest way possible, namely by assuming that they have a fixed length b. Therefore, the FJC can be thought of as an ideal gas of monomers under the constraint that consecutive monomers have a distance b.

3.1 End-to-end vector

The conformational space of a FJC made of $N+1$ monomers is thus dictated purely by entropy, and is equivalent to a random walk of N steps of size b. Starting from the first monomer $i = 0$ with position $\boldsymbol{r}_0$, the position $\boldsymbol{r}_1$ of the monomer $i = 1$ is taken randomly on the spherical surface centered in $\boldsymbol{r}_0$ and with radius b. Analogously, the position $\boldsymbol{r}_2$ of the monomer $i = 2$ is taken randomly on the spherical surface centered in $\boldsymbol{r}_1$ and with radius b, and so on. It is more convenient to work with the tangent vectors $\widehat{\boldsymbol{t}}_i$, which in the FJC are taken in a random direction independently of each other. The FJC is thus analogous to a diffusive process in which time is substituted by the index i. Based on this analogy, we can thus already expect that on average the end-to-end vector is $\langle \boldsymbol{R}_e \rangle = 0$, while its square is proportional to the number of steps N (i.e. the time in the analogy), $\langle R_e^2 \rangle \propto N$.

In order to make a proper derivation, we note that by construction $\boldsymbol{r}_i - \boldsymbol{r}_{i-1} = b\,\widehat{\boldsymbol{t}}_i$. The end-to-end vector can be written as

$$\boldsymbol{R}_e = \boldsymbol{r}_N - \boldsymbol{r}_0 = (\boldsymbol{r}_N - \boldsymbol{r}_{N-1}) + (\boldsymbol{r}_{N-1} - \boldsymbol{r}_{N-2}) + \cdots + (\boldsymbol{r}_1 - \boldsymbol{r}_0) = \sum_{i=1}^{N} b\,\widehat{\boldsymbol{t}}_i \,. \tag{5}$$

As mentioned above, in the FJC the tangent vectors are independent from each other. One can describe the orientation of a tangent vector in polar coordinates by fixing a certain z axis and write $\widehat{\boldsymbol{t}}_i = (\sin\theta_i \cos\phi_i, \sin\theta_i \sin\phi_i, \cos\theta_i)$. The infinitesimal probability of a given orientation is $\sin\theta_i d\theta_i d\phi_i$. Therefore, one can easily compute the average of each component. For instance,

$$\langle \sin\theta_i \cos\phi_i \rangle = \frac{\int_0^\pi d\theta_i \int_0^{2\pi} d\phi_i \sin^2\theta_i \cos\phi_i}{\int_0^\pi d\theta_i \int_0^{2\pi} d\phi_i \sin\theta_i} = 0 \tag{6}$$

and similarly for the other components. It is easy to see that $\left\langle \widehat{\boldsymbol{t}}_i \right\rangle = 0$, which is expected due to the isotropic symmetry of the system. Now, if we consider the correlation $\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle$, we are faced with two cases. If $i = j$, since for any orientation $\widehat{t}_i^2 = 1$, this is also true on average, i.e. $\langle \widehat{t}_i^2 \rangle = 1$. In contrast, if $i \neq j$, due to the independence of different tangent vectors we can

write $\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = \left\langle \widehat{\boldsymbol{t}}_i \right\rangle \cdot \left\langle \widehat{\boldsymbol{t}}_j \right\rangle = 0$. By introducing Kronecker's delta δ_{ij},

$$\delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \ , \\ 0 & \text{if } i \neq j \ , \end{cases} \tag{7}$$

we can summarize our results by writing

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = \delta_{ij} \ . \tag{8}$$

From these findings on the tangent vector, we can easily obtain that

$$\langle \boldsymbol{R}_e \rangle = b \sum_{i=1}^{N} \left\langle \widehat{\boldsymbol{t}}_i \right\rangle = 0 \ , \tag{9}$$

which again is expected due to the spherical symmetry of the system. As for the average of the squared end-to-end distance, we find

$$\left\langle R_e^2 \right\rangle = \left\langle \left(b \sum_{i=1}^{N} \widehat{\boldsymbol{t}}_i \right) \cdot \left(b \sum_{j=1}^{N} \widehat{\boldsymbol{t}}_j \right) \right\rangle = b^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = b^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{ij} \ . \tag{10}$$

We note that the Kronecker's delta plays in a sum a similar role as the one played by Dirac's delta in an integral. Indeed, for a given value of i, among all the terms in the sum over j a non-zero contribution is found only when $j = i$, in which case $\delta_{ii} = 1$. Hence, we can write

$$\boxed{\left\langle R_e^2 \right\rangle = b^2 \sum_{i=1}^{N} 1 = Nb^2 \ .} \tag{11}$$

As we anticipated based on the analogy with the diffusion process, we find that $\langle R_e^2 \rangle \propto N$.

Up to this point we have computed the first two moments of the distribution $P(\boldsymbol{R}_e)$ of the end-to-end vector, i.e. $\langle \boldsymbol{R}_e \rangle = 0$ and $\langle R_e^2 \rangle = Nb^2$. In the simple case of a FJC, it is actually possible to compute explicitly $P(\boldsymbol{R}_e)$. To this aim, we recall the *central limit theorem*: let $x_i, i = 1, \ldots, N$ be N random variables with the same probability distribution and independent of each other. It does not matter the exact shape of their probability distribution. The only important point is that both the average and the variance are finite, and we indicate them as μ and σ^2 respectively. Then, for $N \to \infty$ the random variable S obtained by summing the various x_i, i.e. $S \equiv \sum_{i=1}^{N} x_i$, is distributed according to a Gaussian function with average $N\mu$ and variance $N\sigma^2$. Note that this holds also for multidimensional variables. In our case, the independent, identically-distributed variables are the tangent vectors multiplied by b. Each component (x,y,z) has average 0 and variance $b^2/3$. The end-do-end vector is defined as their sum, so that we conclude that for large polymers each component of $\boldsymbol{R}_e$ is distributed according

to a Gaussian with average 0 and variance $Nb^2/3$. The distribution of the end-to-end vector is then obtained as the product of the distributions of the three components. In formulas, we thus have

$$\boxed{P(\boldsymbol{R}_e) = \left(\frac{3}{2\pi N b^2}\right)^{\frac{3}{2}} e^{-\frac{3R_e^2}{2Nb^2}} .} \tag{12}$$

Problem. Based on Eq.(12), what is the most probable end-to-end vector $\boldsymbol{R_e}$? What is the most probable value of R_e^2? Are these results in agreement with the computations above, where we did not use $P(\boldsymbol{R}_e)$

Problem. Derive the results of this section for a FJC confined on a plane.

3.2 Radius of gyration

For the FJC it is also possible to compute the average square of the radius of gyration. Rather than making brute-force computations, it is best to first play a bit with the definition of R_g. Substituting Equation (4) in Equation (3), we find (note that $\left(\sum_j \boldsymbol{r}_j\right)^2 = \sum_{j,k} \boldsymbol{r}_j \cdot \boldsymbol{r}_k$)

$$\begin{aligned} R_g^2 &= \frac{1}{N+1}\sum_{i=0}^{N}\left(\boldsymbol{r}_i - \frac{1}{N+1}\sum_{j=0}^{N}\boldsymbol{r}_j\right)^2 = \\ &\frac{1}{N+1}\left[\sum_{i=0}^{N} r_i^2 - \frac{2}{N+1}\sum_{i=0}^{N}\sum_{j=0}^{N}\boldsymbol{r}_i\cdot\boldsymbol{r}_j + \frac{1}{(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\sum_{k=0}^{N}\boldsymbol{r}_j\cdot\boldsymbol{r}_k\right] . \end{aligned} \tag{13}$$

In the last term, we see that argument of the triple sum in independent of i, so that its sum can be immediately carried out to give $N+1$. We thus find

$$R_g^2 = \frac{1}{N+1}\left(\sum_{i=0}^{N} r_i^2 - \frac{2}{N+1}\sum_{i=0}^{N}\sum_{j=0}^{N}\boldsymbol{r}_i\cdot\boldsymbol{r}_j + \frac{1}{N+1}\sum_{j=0}^{N}\sum_{k=0}^{N}\boldsymbol{r}_j\cdot\boldsymbol{r}_k\right) . \tag{14}$$

In the last two terms, the sums are the same (they differ only by the names of the dummy indices), so that

$$R_g^2 = \frac{1}{N+1}\left(\sum_{i=0}^{N} r_i^2 - \frac{1}{N+1}\sum_{i=0}^{N}\sum_{j=0}^{N}\boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) . \tag{15}$$

Now, we can reorganize the first term by summing also over j and dividing by $N+1$:

$$R_g^2 = \frac{1}{N+1}\left(\frac{1}{N+1}\sum_{i=0}^{N}\sum_{j=0}^{N} r_i^2 - \frac{1}{N+1}\sum_{i=0}^{N}\sum_{j=0}^{N}\boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) = \\ \frac{1}{(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\left(r_i^2 - \boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) \ . \tag{16}$$

Now, the double sum does not depend on the name given to the dummy indices, i.e. $\sum_{i,j}\left(r_i^2 - \boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) = \sum_{i,j}\left(r_j^2 - \boldsymbol{r}_i\cdot\boldsymbol{r}_j\right)$. Therefore, we can rewrite the double sum in a symmetric form as

$$R_g^2 = \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\left(r_i^2 - \boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) + \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\left(r_j^2 - \boldsymbol{r}_i\cdot\boldsymbol{r}_j\right) = \\ \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\left(r_i^2 - 2\boldsymbol{r}_i\cdot\boldsymbol{r}_j + r_j^2\right) = \\ \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{N}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 \ . \tag{17}$$

It is convenient to further manipulate the formula by splitting the contributions corresponding to $j<i$ and $j>i$. Therefore, we write (note that for $i=j$ the contribution is zero, so that we can add include it as many times as we please)

$$R_g^2 = \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{i}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 + \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=i}^{N}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 = \\ \frac{1}{2(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{i}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 + \frac{1}{2(N+1)^2}\sum_{j=0}^{N}\sum_{i=0}^{j}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 = \\ \frac{1}{(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{i}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2 \ . \tag{18}$$

With this formula, it is easy to compute the average squared radius of gyration $\left\langle R_g^2\right\rangle$. Indeed, we note that $\boldsymbol{r}_i - \boldsymbol{r}_j$ is the end-to-end vector of the fragment of polymer between j and i, which has $i-j$ steps. Therefore, we find $\left\langle\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2\right\rangle = (i-j)b^2$. Armed with this, we find

$$\left\langle R_g^2\right\rangle = \frac{1}{(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{i}\left\langle\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right)^2\right\rangle = \frac{b^2}{(N+1)^2}\sum_{i=0}^{N}\sum_{j=0}^{i}(i-j) \ . \tag{19}$$

Changing the dummy variable j to $k = i - j$, we find

$$\left\langle R_g^2 \right\rangle = \frac{b^2}{(N+1)^2} \sum_{i=0}^{N} \sum_{k=0}^{i} k = \frac{b^2}{(N+1)^2} \sum_{k=0}^{N} \sum_{i=k}^{N} k = \frac{b^2}{(N+1)^2} \sum_{k=0}^{N} (N-k+1)k \; . \tag{20}$$

In order to finalize our calculation, we need to compute two sums, that we denote as S_1 and S_2. The first one reads[2]

$$S_1 = \sum_{k=0}^{N} k \; . \tag{21}$$

This sum is easily computed by observing that $S_1 = 1 + 2 + \cdots + N = (1+N) + [2 + (N-1)] + [3 + (N-2)] + \ldots$. All the terms in the previous rearrangement are equal to $N+1$. If N is even, the last term to appear is $N/2 + (N/2+1)$, so that there are $N/2$ terms like this. In this case the sum is thus $S_1 = (N+1) \cdot N/2$. If N is odd, there is a spare term corresponding to the central number $(N+1)/2$. In this case there are $(N-1)/2$ couples, so that including also the spare term we find $S_1 = (N+1) \cdot (N-1)/2 + (N+1)/2 = (N+1) \cdot N/2$, which is the same result as in the case of even N. The second sum is

$$S_2 = \sum_{k=0}^{N} k^2 = \sum_{k=1}^{N} k^2 \; . \tag{22}$$

To obtain S_2, we note that according to our finding for S_1, the following equality holds: $\sum_{l=1}^{k} l = k(k+1)/2 \Rightarrow k^2 = 2\sum_{l=1}^{k} l - k$. Thus we can write

$$\begin{aligned}
S_2 &= \sum_{k=1}^{N} \left(2 \sum_{l=1}^{k} l - k \right) = \\
&2 \sum_{k=1}^{N} \sum_{l=1}^{k} l - \sum_{k=1}^{N} k = \\
&2 \sum_{l=1}^{N} \sum_{k=l}^{N} l - S_1 = \\
&2 \sum_{l=1}^{N} [l(N-l+1)] - S_1 = \\
&2NS_1 - 2S_2 + 2S_1 - S_1 = \\
&(2N+1)S_1 - 2S_2 \; .
\end{aligned} \tag{23}$$

[2]According to the legend, an eight-year old Gauss was given this problem for $N = 100$ together with his classmates by a bored professor who wanted to keep them busy for a while. Gauss came up with the solution shown here in a few minutes, resulting in a mixture of awe and anger for the professor.

Comparing the first and last term, we get an equation from which S_2 is easily found:

$$S_2 = \frac{1}{3}(2N+1)S_1 = \frac{1}{6}N(N+1)(2N+1)\ . \tag{24}$$

Armed with these results, substituting in Equation (20) we finally obtain

$$\begin{aligned}
&\left\langle R_g^2 \right\rangle = \frac{b^2}{(N+1)^2}\left(NS_1 - S_2 + S_1\right) = \\
&\frac{b^2}{(N+1)^2}\left[\frac{1}{2}N(N+1)^2 - \frac{1}{6}N(N+1)(2N+1)\right] = \\
&\frac{1}{6}Nb^2\frac{N+2}{N+1}\ .
\end{aligned} \tag{25}$$

Considering $N \gg 1$, we get

$$\boxed{\left\langle R_g^2 \right\rangle = \frac{1}{6}Nb^2}\ , \tag{26}$$

showing the interesting result $\langle R_e^2 \rangle = 6\left\langle R_g^2 \right\rangle$.

4 Real chains: excluded volume and monomer-monomer attractions

One of the most immediate shortcomes of the FJC is the complete absence of interactions between monomers, which can thus unphysically cross each other. Similarly, possible monomer-monomer attractions (such as the interactions holding proteins together) are also neglected. In the present section, we will see how to include these features, at least in an approximate way.

4.1 Cluster expansion of the partition function

In order to account for the interaction between monomers, we must first take a step back into some fundamental results of statistical mechanics. For this, let us first consider a system made of N particles characterized by an energy $U_{\rm tot}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$, which depends in general in a complex way on the coordinates, including multibody terms. The configurational part of the partition function of the system is written as

$$Z = \frac{1}{N!} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N e^{-\frac{U_{\rm tot}(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)}{k_B T}} \tag{27}$$

Following a derivation by Mayer, we consider as a first assumption that the overall potential can be written as the sum of pairwise terms:

$$U_{\rm tot}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \sum_{i<j} U(r_{ij}) \,, \tag{28}$$

where $r_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i|$ and the sum is run over $i < j$ to avoid double counting. In order to fix the ideas, U can be for instance a Lennard-Jones potential, so that it has both an attractive and a repulsive part. Within this assumption, the partition function is written as

$$Z = \frac{1}{N!} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N e^{-\sum_{i<j} \frac{U(r_{ij})}{k_B T}} = \frac{1}{N!} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N \prod_{i<j} e^{-\frac{U(r_{ij})}{k_B T}} \tag{29}$$

At this point, it is convenient to define the Mayer function $f(r)$ as

$$f(r) = e^{-\frac{U(r)}{k_B T}} - 1 \,, \tag{30}$$

which gives the excess Boltzmann weight at distance r with respect $r = \infty$, for which $U = 0$. In other words, $f(r) > 0$ when the position is energetically favourable and $f(r) < 0$ when it is unfavourable. By indicating as $f_{i,j} = f(r_{ij})$, the partition function is now written as

$$Z = \frac{1}{N!} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N \prod_{i<j} (1 + f_{i,j}) \,. \tag{31}$$

It is instructive to write more explicitly the integrand:

$$\prod_{i<j}(1+f_{i,j}) = (1+f_{1,2})\cdot(1+f_{1,3})\cdot(1+f_{1,N})\cdot(1+f_{2,3})\cdot\ \ldots\ \cdot(1+f_{N-1,N})\ . \tag{32}$$

The explicit form suggests that one can rearrange $\prod_{i<j}(1+f_{i,j})$ by collecting all the terms corresponding to a given power of f. For instance, there is only one term corresponding to the 0-th power, obtained by multiplying all the "1" in the previous formula. Similarly, the linear term is obtained by multiplying each $f_{i,j}$ by all the "1" appearing in the other terms. One can thus write:

$$\prod_{i<j}(1+f_{i,j}) = 1+\sum_{i<j} f_{i,j} + \ldots\ . \tag{33}$$

The next term would contain quadratic terms such as $f_{1,2}\cdot f_{1,3}$ or $f_{1,2}\cdot f_{3,13}$. We note that for an ideal gas one has $f_{i,j}=0$, so that the 0-th power term (that is, the "1") corresponds to the non-interacting case. Moreover, after substitution into Eq. (31), only the variables corresponding to the indices of the particular combination are involved in a non-trivial integration. For example, for a term such as $f_{1,7}$, the corresponding integral in the configurational partition function would read

$$\frac{1}{N!}\int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N\ f_{1,7}\,. \tag{34}$$

In the previous formula, all the integration variables $d\boldsymbol{r}_i$ where $i\neq 1,7$ can be readily integrated, each giving a contribution equal to the volume V of the system. The previous formula thus becomes:

$$\frac{V^{N-2}}{N!}\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_7\ f_{1,7}\,. \tag{35}$$

It is thus natural to associate the term $f_{1,7}$ to the couple of particles 1 and 7. Similarly, a term such as $f_{1,7}\cdot f_{3,7}$ would result in

$$\frac{V^{N-3}}{N!}\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_3 \int d\boldsymbol{r}_7\ f_{1,7}\cdot f_{3,7}\,, \tag{36}$$

which is naturally associated to the *cluster* formed by particles 1,3,7. As a further example, for a term such as $f_{1,7}\cdot f_{3,13}$, the corresponding integral in the configurational partition function would read

$$\frac{V^{N-4}}{N!}\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_3 \int d\boldsymbol{r}_7 \int d\boldsymbol{r}_{13}\ f_{1,7}\cdot f_{3,13}\,, \tag{37}$$

which is thus associated to the cluster formed by particles $1,3,7,13$. In this case, however, the independence of $f_{1,7}$ and $f_{3,13}$ from each other (no index is repeated) allows to further rearrange the previous formula as

$$\frac{V^{N-4}}{N!}\left(\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_7\ f_{1,7}\right)\cdot\left(\int d\boldsymbol{r}_3 \int d\boldsymbol{r}_{13}\ f_{3,13}\right)\,, \tag{38}$$

i.e. the independent terms can be factorized. In this case, the cluster $1, 3, 7, 13$ can be thus "decomposed" into two *irreducible clusters* formed by $1, 7$ and $3, 13$.

These observations allow the construction of a theoretical framework called *cluster expansion*, which allows to treat the full configurational partition function (31) in a closed form. This theory is however beyond our scope, so that we will limit our treatment to the case in which one can retain up to the linear term, i.e. the clusters of size 2. In other words, we write

$$Z \simeq \frac{1}{N!} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N \left(1 + \sum_{i<j} f_{i,j} \right) . \tag{39}$$

This approximation is valid if $f_{i,j} \ll 1$, which in practice is obtained for weak potentials $U(r) \ll k_B T$ (compare Eq. (30)). Under this approximation, we thus have

$$Z \simeq \frac{V^N}{N!} + \frac{1}{N!} \sum_{i<j} \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_N f_{i,j} = \frac{V^N}{N!} + \frac{V^{N-2}}{N!} \sum_{i<j} \int d\boldsymbol{r}_i \int d\boldsymbol{r}_j \; f_{i,j} \, . \tag{40}$$

Remembering that $f_{i,j} = f(|\boldsymbol{r}_j - \boldsymbol{r}_i|)$, we can operate the variable substitution $\boldsymbol{r}_j \to \boldsymbol{s} = \boldsymbol{r}_j - \boldsymbol{r}_i$, which yields

$$\int d\boldsymbol{r}_i \int d\boldsymbol{r}_j \; f(|\boldsymbol{r}_j - \boldsymbol{r}_i|) = \int d\boldsymbol{r}_i \int d\boldsymbol{s} \; f(s) = V \cdot \int d\boldsymbol{s} \; f(s) = -V \cdot v \tag{41}$$

where the volume term is obtained by observing that upon the change of variable the integrand is independent of $\boldsymbol{r}_i$, and where we defined $v \equiv -\int d\boldsymbol{s} \; f(s)$. Armed with this, we can rewrite the partition function as

$$Z \simeq \frac{V^N}{N!} + \frac{V^{N-2}}{N!} \sum_{i<j} \int d\boldsymbol{r}_i \int d\boldsymbol{r}_j \; f_{i,j} = \frac{V^N}{N!} - \frac{V^{N-1}}{N!} \sum_{i<j} v \, . \tag{42}$$

The argument of the sum is now independent of the indices, so the sum is trivially obtained by counting the number of times in which the constant term v is repeated. This number of times corresponds to the number of combinations for which $i < j$, equal to $N(N-1)/2$. Assuming $N \gg 1$, we can thus write

$$Z \simeq \frac{V^N}{N!} - \frac{V^{N-1}}{N!} \cdot \frac{1}{2} N^2 v = \frac{V^N}{N!} \left(1 - N^2 \frac{v}{2V} \right) . \tag{43}$$

The corresponding free energy is thus obtained as

$$F = -k_B T \ln Z \simeq -k_B T \left[N \ln V - N \ln N! + \ln \left(1 - N^2 \frac{v}{2V} \right) \right] . \tag{44}$$

Finally, we assume the density to be small enough to treat the term $N^2 v/(2V)$ as a perturbation, which allows to write

$$F \simeq -k_B T \left(N \ln V - N \ln N! - N^2 \frac{v}{2V} \right) . \tag{45}$$

This formula provides an approximation of the configurational free energy accounting for pairwise interactions via the parameter v, which has the units of a volume and is called *excluded volume*, as further commented below.

As a final observation, we note that the pressure can be obtained via a Maxwell relation:

$$p = -\left(\frac{\partial F}{\partial V}\right)_T = \frac{NK_BT}{V} + N^2 k_B T \frac{v}{2V^2} \,. \tag{46}$$

From this formula, we thus find

$$\frac{pV}{k_BT} = 1 + \frac{v}{2} \cdot \frac{N}{V} \,. \tag{47}$$

We see that the pairwise interaction introduces a corrective term to the ideal gas law, proportional to the density N/V. Interestingly, the sign of the change depends on the sign of v, leading to a larger pressure if $v > 0$ and to a smaller pressure if $v < 0$. Eq. (47) corresponds to the first terms of the so-called *Virial expansion*, which can be formally obtained by considering the full treatment of the cluster expansion.

4.2 The excluded volume

In our previous treatment, we have considered a pairwise potential U, which in general has both an attractive and a repulsive part (think e.g. of the Lennard-Jones potential, Fig. 4 top). As discussed in the previous section, the Mayer function $f(r)$ is defined as

$$f(r) = e^{-\frac{U(r)}{k_BT}} - 1 \,, \tag{48}$$

i.e. it gives the excess Boltzmann weight at distance r with respect $r = \infty$, for which $U = 0$. With some approximation, one can ascribe the regime for which $f(r) > 0$ to the attractive part of the potential and the one for which $f(r) < 0$ to the repulsive part. Integrating the Mayer function over the whole space thus gives a measure of the net interaction, i.e. it indicates whether overall the monomers are attracting or repelling each other. Particularly, we defined in the previous section the *excluded volume* v as

$$v \equiv -\int f(r) d\boldsymbol{r} = -4\pi \int_0^\infty r^2 f(r) dr \,. \tag{49}$$

If $v < 0$ there is an overall attraction between the monomers. More precisely, it is energetically more advantageous for the particles to stick together rather than being exposed to the solvent. One thus speaks of *poor solvent conditions*. In this case, the particles are expected to show a tendency to pack. For $v > 0$ there is a net repulsion, i.e. the particles prefer interacting with the solvent rather than with each other. One hence speaks of *good solvent conditions*. In the particular case in which $v = 0$, there is no net preference and one has *θ-solvent conditions*. Note that v depends on temperature. Particularly, for $T \to 0$ one has that $v \to -\infty$, i.e.

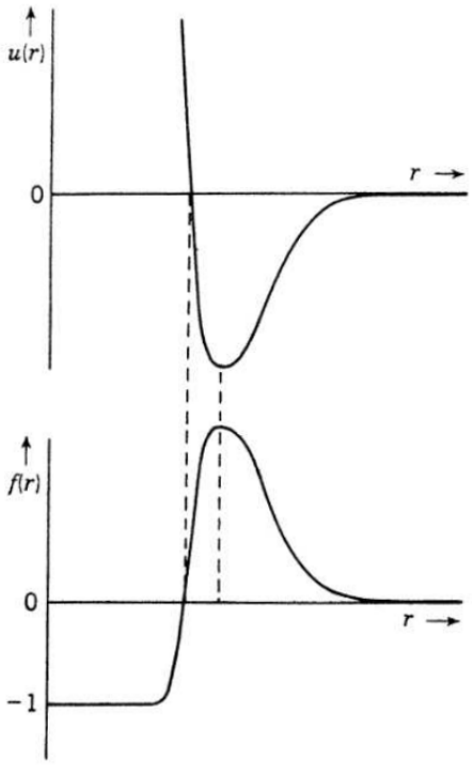

Figure 4: Lennard-Jones potential (top) and corresponding Mayer function (bottom) at a given temperature. Reprinted from T. L. Hill, Statistical Mechanics (McGraw-Hill, New York, 1956).

low temperature implies poor-solvent conditions. Increasing the temperature, the value of v decreases in magnitude (i.e. it approaches zero from below), until it eventually crosses the value $v = 0$ at a temperature called the *θ-temperature* of the system. Further increasing the temperature leads to $v > 0$, i.e. to good-solvent conditions.

The parameter v is called excluded volume because it has the units of a volume and in the case of hard spheres corresponds to the volume that they exclude from each other. Indeed, for hard spheres of size R, the potential is $U(r) = 0$ for $r > 2R$ and $U(r) = \infty$ for $r < 2R$. Correspondingly, the Mayer function takes values $f(r) = 0$ for $r > 2R$ and $f(r) = -1$ for $r < 2R$. The excluded volume v is thus

$$v = 4\pi \int_0^{2R} r^2 dr = \frac{4}{3}\pi \cdot (2R)^3 \,, \tag{50}$$

which is the volume of a sphere of radius $2R$ and corresponds to the volume precluded to an approaching monomer.

4.3 θ solvent

Armed with the tools developed in the two previous sections, we are now ready to treat the case of a polymer in which the monomers interact with each other via a potential $U(r)$. First of all, note that in the case in which $v = 0$ (θ solvent), the net two-body interaction is zero. This does not mean that monomers do not interact at all, but rather that attraction and repulsion at the two-body level cancel out at a statistical level; in other words, overlaps are still forbidden and attraction is still present, but they affect the system only at the next level of perturbation (the third term of the virial expansion). It turns out that for this case the statistics predicted by the FJC model well describe the dependence of the polymer size on the number of steps N.

4.4 Good solvent

Now, let us consider the case in which monomers repel each other, $v > 0$. An exact analytical treatment of this problem is very difficult in two dimensions and (to date) impossible in three dimensions. Therefore, we consider an approximate approach to the problem provided by the so-called *Flory theory* (which is more a "philosophy" in polymer problems rather than a particular theory). Basically, we provide some scaling arguments to provide the entropic and enthalpic contributions to the free energy. As for the interaction part of the free energy, we consider the pairwise contribution to Eq. (45), $k_BT \cdot N^2v/(2V)$, which in practice means that for this aspect we are assuming the monomers to be a gas contained in a certain volume V. The volume is restricted by the overall size R of the polymer[3], i.e. $V \sim R^3$. The interaction part of the free energy is thus written as $U \sim k_BT \cdot N^2v/R^3$, where we are dropping the numerical prefactor (1/2) as usually done in scaling arguments.

It is instructive to describe an alternative way to obtain the interaction term via heuristic arguments, according to which v is interpreted as the volume within which the repulsive interaction takes place with intensity equal to k_BT. For a given monomer i, the probability that any other monomer interacts with it is $v/V \sim v/R^3$. Therefore, the expected number of overlaps between i and other monomers is N times this probability, i.e. Nv/R^3. The total number of overlaps is obtained by summing over all monomers, i.e. by further multipyling by N: N^2v/R^3. The interaction part of the free energy is thus $U \sim k_BTN^2v/R^3$, which is the result obtained by application of the Virial expansion.

As for the entropic contribution to the free energy, we can estimate it by means of the distribution of the end-to-end vector, Equation (12), obtaining $S \sim k_B \ln P(\boldsymbol{R}) \sim -k_BR^2/(Nb^2)$. The Flory free energy thus reads

$$F = U - TS \sim k_BTN^2\frac{v}{R^3} + k_BT\frac{R^2}{Nb^2} \; . \tag{51}$$

[3]We do not distinguish here between R_e and R_g due to the expected similar scaling behavior of the two quantities. We just leave a generic R.

We can appreciate that the enthalpic term is minimized for $R \to \infty$, while the entropic contribution pushes towards $R \to 0$. The value of R is obtained by minimizing F, i.e.

$$\frac{dF}{dR} = 0 \Rightarrow k_B T N^2 \frac{v}{R^4} \sim k_B T \frac{R}{N b^2} \Rightarrow \boxed{R \sim b^{\frac{2}{5}} v^{\frac{1}{5}} N^{\frac{3}{5}} \propto N^{0.6}} \,. \tag{52}$$

This result has to be compared to the FJC case, for which $R \propto N^{1/2}$. The effect of a repulsive excluded volume is thus to increase the exponent, indicating an overall swelling of the polymer, which is what one could have intuitively expected. Numerical results have provided a more precise estimation of the exponent, which is roughly equal to 0.589. Considering the gross simplification of the Flory theory, the estimated value is in pretty good agreement with the exact result!

Problem. Find the Flory approximation for the exponent in two dimensions. How does it compare with the exact result 3/4?

Problem. Find the Flory approximation for the exponent in one dimension. What does it mean to have a polymer with excluded volume in one dimension? Does the value of the exponent make sense compared to the physical intuition?

4.5 Poor solvent

In the opposite case of poor-solvent conditions, one has a net attraction characterized by $v < 0$. The Flory free energy, Equation (51), is minimized in this case by $R = 0$ (as also the enthalpic term is now pushing towards $R \to 0$), which suggests the unphysical result of a polymer collapsing onto itself. In order to get a physically-correct result, we need to go beyond the two-body interaction as captured by the excluded volume. Indeed, in such a compact conformation, repulsive three-body interactions become important. One of the key assumptions behind our development of the Virial expansion (stopping at the two-body interactions) was the low density of the gas. In the case of a poor solvent, the polymer is expected to shrink, so that the "gas" of monomers is expected to have a high density within the volume spanned by the polymer. In terms of the Virial expansion, one has to consider the three-body term in Eq. (45). In terms of dependence on N and V, this term is obtained by further multiplying the two-body term by the density N/V; the role of the excluded volume is instead played by a different parameter, which we indicate as w. It can be shown that $w > 0$, which implies that the three-body term gives always a repulsive contribution (which is just what we need to avoid the collapse of the polymer in a poor solvent!). All in all, the sought term has thus the form $k_B T w N^3/V^2 \sim k_B T w N^3/R^6$.

As for the two-body term, also in this case we can obtain the three-body term via an alternative approach using heuristic arguments. Following the same reasoning as above, for a monomer i the probability that two other monomers lie in its neighborhood is proportional to $1/R^3 \cdot 1/R^3$. Denoting as $w > 0$ the proportionality constant (with the units of a volume squared), one gets for the probability w/R^6. The average number of couples found in the neighborhood are computed by multiplying the probability by the total number of possible

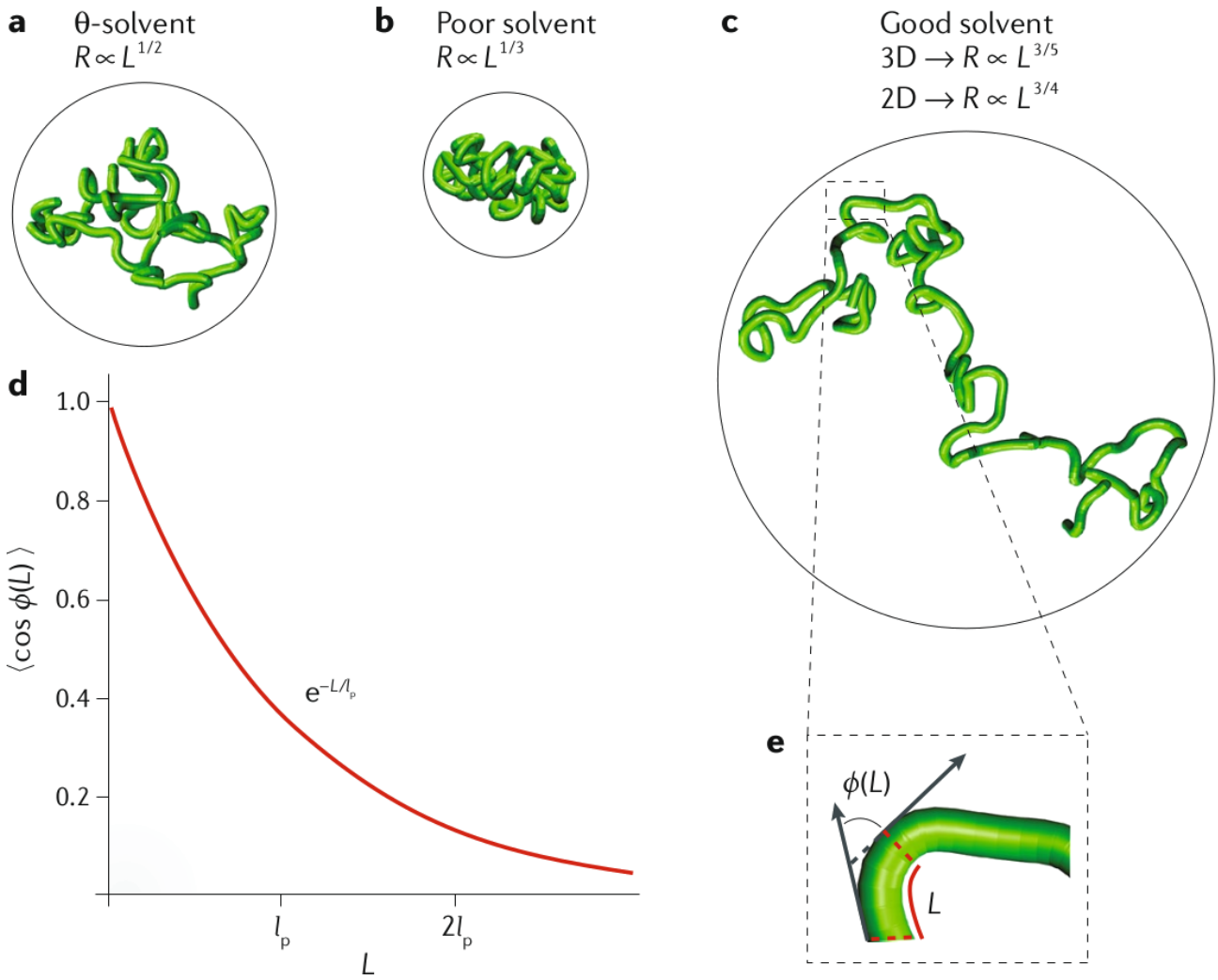

Figure 5: a-c) Scaling behavior of polymer size according to solvent quality. d) Exponential decay of the tangent-tangent correlation function, where the angle ϕ is defined in (e). Reprinted from S. Assenza and R. Mezzenga, "Soft condensed matter physics of foods and macronutrients", *Nature Rev. Phys.* **1**:551 (2019).

couples, which is roughly equal to N^2, i.e. $N^2 w/R^6$. Finally, the total number of three-body overlaps is obtained summing over i, i.e. further multiplying by N: $N^3 w/R^6$. With this term the Flory free energy reads

$$F \sim k_B T N^2 \frac{v}{R^3} + k_B T \frac{R^2}{Nb^2} + k_B T N^3 \frac{w}{R^6} \,. \tag{53}$$

Note that in good-solvent conditions and for $N \gg 1$, the size of the polymer is large and the three-body term becomes negligible, thus obtaining back Equation (51). In the present case of

a poor solvent, we instead expect the polymer to take small values of R, so that it is expected that the entropic term is negligible, i.e. we have

$$F \sim k_B T N^2 \frac{v}{R^3} + k_B T N^3 \frac{w}{R^6} \ . \tag{54}$$

Minimizing with respect to R and remembering that $v < 0$ we find

$$\boxed{R \sim \left(\frac{w}{|v|}\right)^{\frac{1}{3}} N^{\frac{1}{3}} \propto N^{\frac{1}{3}} \ .} \tag{55}$$

As expected, the exponent has become lower than in the case of a FJC. We further note that 1/3 is a physically-intuitive result. If one depicts the monomers as sticky balls with excluded volume, they will aggregate into a roughly spherical shape whose volume is proportional to the number of monomers, i.e. $R^3 \propto N \Rightarrow R \propto N^{1/3}$.

The scaling exponents depending on the solvent quality are summarized pictorially in Fig.5a-c.

4.6 Scaling and biopolymers

In spite of the oversimplification of the polymer models employed above, the scaling predictions from polymer theory are pretty powerful and describe well DNA and, with some caution, proteins.

In Fig.6a we show the scaling behavior of R_g as a function of the contour length for proteins in the presence of the denaturant Guanidinium chloride at various concentrations. Although interactions in proteins are pretty specific, at a first approximation one can think of a globular protein as a polymer in a poor solvent. In the absence of denaturant, the scaling follows indeed the exponent predicted by poor-solvent conditions, Equation (55). Adding the denaturant changes the solvent until, at very large concentrations, good-solvent conditions are achieved and the scaling predicted by Equation (52) is obtained. Not bad for a model based on the "ideal gas of polymers"!

In Fig.6b we report instead the behavior of the end-to-end distance for DNA in three dimensions and lysozyme-based amyloid fibrils (which are elongated aggregates of proteins) in two dimensions. We see that for large values of the contour length the correct exponents for good solvent conditions are obtained (in two dimensions the exponent is 3/4). Nevertheless, at short length scales there is a larger exponent (equal to 1), whose origin will become clear in the next section.

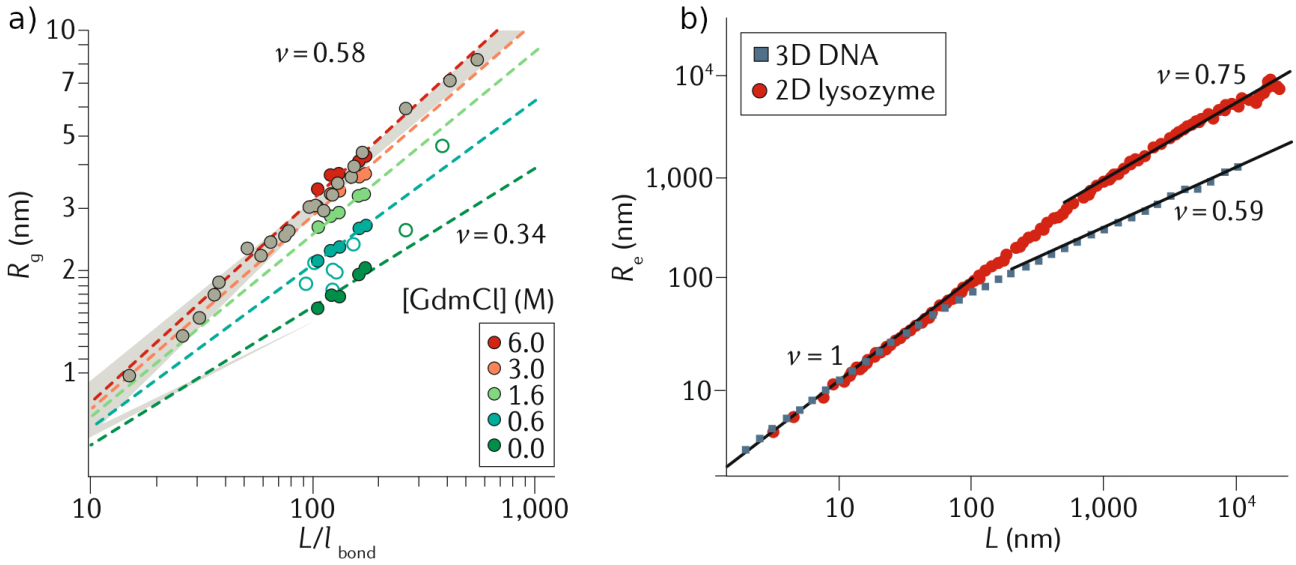

Figure 6: a) Scaling of the radius of gyration experimentally found for a set of proteins under various concentrations of denaturant. b) Scaling of the end-to-end distance vs contour length for DNA in three dimensions and lysozyme-based amyloid fibrils in two dimensions. Reprinted from S. Assenza and R. Mezzenga, "Soft condensed matter physics of foods and macronutrients", *Nature Rev. Phys.* **1**:551 (2019).

5 Adding bending stiffness: the wormlike chain

Another feature overlooked by the FJC is the presence of *bending stiffness*, i.e. the elastic response of the polymer chain against bending stress. Indeed, apart from excluded-volume features, in many cases it is unlikely that a bond takes a strongly different deviation from the one that precedes it, i.e. the angle depicted in Fig.3 left is seldom found to be small.

5.1 The discrete Kratky-Porod model

A simple way to account for the bending stiffness is provided by the *Kratky-Porod* (KP) model. In this model a energy cost is associated to a finite bending angle. The size of each bond is fixed to a value b as in the FJC, but in addition there is the following bending energy:

$$U = -\frac{B}{b}\sum_{i=1}^{N-1} \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \ , \tag{56}$$

where the parameter B is called the *bending stiffness* and has the units of an energy times a length.

5.1.1 Tangent vector correlation function

Differently than for a FJC, now the correlation between two different tangent vectors is not zero. For instance, let us consider the correlation between two consecutive tangent vectors $\widehat{\boldsymbol{t}}_i$ and $\widehat{\boldsymbol{t}}_{i+1}$. In order to compute it, we consider a reference frame where the z axis corresponds to $\widehat{\boldsymbol{t}}_i$, while the x and y axis are suitably chosen on the plane perpendicular to $\widehat{\boldsymbol{t}}_i$. Note that this gives an arbitrarity in the particular orientation that the axis have on the plane. Within this reference frame, the dot product is $\widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} = \cos\theta$. Its average value is obtained as usual by making use of the energy given by Equation (56):

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle = \frac{\int_0^\pi d\theta \sin\theta \cos\theta e^{\frac{\beta B}{b}\cos\theta}}{\int_0^\pi d\theta \sin\theta e^{\frac{\beta B}{b}\cos\theta}} \ . \tag{57}$$

Note that all the tangent vectors other than the ones under consideration can be neglected since they appear in the same way both in the numerator and the denominator. Considering the variable $x = \cos\theta \Rightarrow dx = -\sin\theta d\theta$, the previous integral becomes

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle = \frac{\int_{-1}^1 x e^{\frac{\beta B}{b}x}\, dx}{\int_{-1}^1 e^{\frac{\beta B}{b}x} dx} \ . \tag{58}$$

Using the results $\int_{-1}^1 x e^{cx} dx = 2\cosh c/c - 2\sinh c/c^2$ and $\int_{-1}^1 e^{cx} dx = 2\sinh c/c$, we obtain

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle = \mathcal{L}\left(\frac{\beta B}{b}\right) \ , \tag{59}$$

where $\mathcal{L}(x) \equiv 1/\tanh x - 1/x$ is the Langevin function. As a reminder, the Langevin function monotonically increases between 0 and 1 and has limiting behavior $\mathcal{L}(x) \simeq x/3$ for $x \ll 1$ and $\mathcal{L}(x) \simeq 1 - 1/x$ for $x \gg 1$.

Now, we want to compute the average correlation between two tangent vectors which are two steps apart, i.e. $\widehat{\boldsymbol{t}}_i$ and $\widehat{\boldsymbol{t}}_{i+2}$. To this aim, we first consider an orthonormal reference frame $\widehat{o}_1, \widehat{o}_2, \widehat{\boldsymbol{t}}_{i+1}$ where $\widehat{o}_1$ and $\widehat{o}_2$ are suitably defined. The tangent vector $\widehat{\boldsymbol{t}}_{i+2}$ can be decomposed onto this reference frame as $\widehat{\boldsymbol{t}}_{i+2} = \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1\right) \widehat{o}_1 + \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_2\right) \widehat{o}_2 + \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{\boldsymbol{t}}_{i+1}\right) \widehat{\boldsymbol{t}}_{i+1}$. In this way, the average scalar product with $\widehat{\boldsymbol{t}}_i$ becomes

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+2} \right\rangle = \left\langle \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1\right) \left(\widehat{\boldsymbol{t}}_i \cdot \widehat{o}_1\right) \right\rangle + \left\langle \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_2\right) \left(\widehat{\boldsymbol{t}}_i \cdot \widehat{o}_2\right) \right\rangle + \left\langle \left(\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{\boldsymbol{t}}_{i+1}\right) \left(\widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1}\right) \right\rangle \ . \quad (60)$$

Now, since the energy involves only consecutive terms, the various products in the previous formulas are independent from each other and the averages can be split

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+2} \right\rangle = \left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1 \right\rangle \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{o}_1 \right\rangle + \left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_2 \right\rangle \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{o}_2 \right\rangle + \left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle \ . \quad (61)$$

All the averages involving $\widehat{o}_1$ and $\widehat{o}_2$ are equal to zero. Indeed, let us take as an example $\left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1 \right\rangle$. Within the considered reference frame, $\widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1 = \sin\theta \cos\phi$. Hence, we have

$$\left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{o}_1 \right\rangle = \frac{\int_0^\pi d\theta \int_0^{2\pi} d\phi \sin\theta \cdot \sin\theta \cos\phi e^{\frac{\beta B}{b}\cos\theta}}{\int_0^\pi d\theta \int_0^{2\pi} d\phi \sin\theta e^{\frac{\beta B}{b}\cos\theta}} = 0 \ . \quad (62)$$

The integration over ϕ at the numerator can be immediately carried out and gives zero. Similar arguments apply to all the terms involving $\widehat{o}_1$ and $\widehat{o}_2$ in Equation (61), which thus becomes

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+2} \right\rangle = \left\langle \widehat{\boldsymbol{t}}_{i+2} \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+1} \right\rangle \ . \quad (63)$$

However, these terms involve only consecutive tangent vectors, so that each of them gives $\mathcal{L}(\beta B/b)$. Therefore, we end up with

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_{i+2} \right\rangle = \mathcal{L}^2\left(\frac{\beta B}{b}\right) \ . \quad (64)$$

This approach can be generalized, so that one has for the *tangent vector correlation function*

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = \mathcal{L}^{|i-j|}\left(\frac{\beta B}{b}\right) \ . \quad (65)$$

In other words, the correlation between distant tangent vectors is a memory effect being propagated along the chain. The previous formula can be rearranged as

$$\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = e^{|i-j| \ln \mathcal{L}\left(\frac{\beta B}{b}\right)} \ . \quad (66)$$

By defining

$$\boxed{l_p \equiv -\frac{b}{\ln \mathcal{L}\left(\frac{\beta B}{b}\right)}}\,, \tag{67}$$

the correlation function can be written as

$$\boxed{\left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = e^{-\frac{|i-j|b}{l_p}}\,.} \tag{68}$$

Therefore, as shown in Fig.5, the correlation function decays exponentially with length constant l_p, which is called the *persistence length* of the polymer. Note that $l_p > 0$ since $\mathcal{L} < 1$.

Problem. Another model for polymer bending stiffness is the so-called *freely-rotating chain model*. In this case, the bond length is fixed to a value b as usual, and the bending elasticity is implemented by imposing that two consecutive tangent vectors have a constant angle θ. Show that for the freely-rotating chain the persistence length is

$$l_p = -\frac{b}{\ln \cos\theta}\,. \tag{69}$$

5.1.2 End-to-end distance

The correlation function enables computing the end-to-end distance of the KP model. Indeed, remembering that $\boldsymbol{R}_e = b\sum_i \widehat{\boldsymbol{t}}_i$, we have

$$\left\langle R_e^2 \right\rangle = b^2 \sum_{i=1}^{N}\sum_{j=1}^{N} \left\langle \widehat{\boldsymbol{t}}_i \cdot \widehat{\boldsymbol{t}}_j \right\rangle = b^2 \sum_{i=1}^{N}\sum_{j=1}^{N} e^{-\frac{|i-j|b}{l_p}}\,. \tag{70}$$

The sum of the terms corresponding to $j < i$ is the same as the sum of the terms for which $j > i$. Therefore, the previous formula can be written as

$$\left\langle R_e^2 \right\rangle = 2b^2 \sum_{i=1}^{N}\sum_{j=1}^{i-1} e^{-\frac{(i-j)b}{l_p}} + Nb^2\,, \tag{71}$$

where the second term stems from the distinct case $j = i$, for which $\sum_{i=1}^{N} e^{-\frac{(i-i)b}{l_p}} = \sum_{i=1}^{N} 1 = N$. Now, to ease notation let us go back to the Langevin formula $\mathcal{L} = e^{-\frac{b}{l_p}}$. The first sum in Equation (71) can be written as

$$\sum_{i=1}^{N}\sum_{j=1}^{i-1} \mathcal{L}^{i-j} = \sum_{i=1}^{N}\sum_{k=1}^{i-1} \mathcal{L}^{k}\,, \tag{72}$$

where we made the change of dummy variable $k = i - j$. Swapping the two sums, we thus find

$$\sum_{i=1}^{N}\sum_{j=1}^{i-1}\mathcal{L}^{i-j} = \sum_{k=1}^{N-1}\sum_{i=k+1}^{N}\mathcal{L}^{k} = \sum_{k=1}^{N-1}(N-k)\mathcal{L}^{k} \ . \tag{73}$$

In order to go any further, we need to compute two sums that we call Q_0 and Q_1. The first one is

$$\begin{aligned} Q_0 &= \sum_{k=1}^{N-1}\mathcal{L}^{k} = \\ &\mathcal{L} + \mathcal{L}^2 + \cdots + \mathcal{L}^{N-2} + \mathcal{L}^{N-1} = \\ &\mathcal{L}\left(1 + \mathcal{L} + \cdots + \mathcal{L}^{N-2}\right) = \\ &\mathcal{L}\left(1 + Q_0 - \mathcal{L}^{N-1}\right) \ . \end{aligned} \tag{74}$$

Comparing first and last side gives an equation from which Q_0 can be determined as

$$Q_0 = \mathcal{L}\frac{1-\mathcal{L}^{N-1}}{1-\mathcal{L}} \ . \tag{75}$$

The second sum is

$$\begin{aligned} Q_1 &= \sum_{k=1}^{N-1} k\mathcal{L}^{k} = \\ &\mathcal{L}\sum_{k=1}^{N-1} k\mathcal{L}^{k-1} = \\ &\mathcal{L}\sum_{k=1}^{N-1}\frac{d\mathcal{L}^{k}}{d\mathcal{L}} = \\ &\mathcal{L}\frac{d}{d\mathcal{L}}\sum_{k=1}^{N-1}\mathcal{L}^{k} = \\ &\mathcal{L}\frac{dQ_0}{d\mathcal{L}} = \\ &\frac{\mathcal{L} - N\mathcal{L}^{N} + (N-1)\mathcal{L}^{N+1}}{(1-\mathcal{L})^2} \ . \end{aligned} \tag{76}$$

Considering large N and remembering that $\mathcal{L} < 1$, we get

$$Q_0 \simeq \frac{\mathcal{L}}{1-\mathcal{L}} \qquad \text{and} \qquad Q_1 \simeq \frac{\mathcal{L}}{(1-\mathcal{L})^2} \,. \tag{77}$$

Substituting these results in Equation (73), we find

$$\sum_{i=1}^{N} \sum_{j=1}^{i-1} \mathcal{L}^{i-j} = NQ_0 - Q_1 \simeq N \frac{\mathcal{L}}{1-\mathcal{L}} \,. \tag{78}$$

Hence, Equation (71) becomes

$$\boxed{\left\langle R_e^2 \right\rangle = 2b^2 \cdot N \frac{\mathcal{L}}{1-\mathcal{L}} + Nb^2 = Nb^2 \frac{1+\mathcal{L}}{1-\mathcal{L}} \,.} \tag{79}$$

This result is pretty instructive. It shows that for large N the proportionality $\langle R_e^2 \rangle \propto N$ holds in the same way as in the FJC. Physically, this is due to the fact that the memory induced by the bending decays exponentially, so that tangent vectors which are far enough along the chain will effectively behave as in a FJC. This result is not a specific feature of the KP model, but is valid for any polymer model with bending stiffness, although naturally the multiplicative constant (here equal to $(1+\mathcal{L})/(1-\mathcal{L})$) changes from case to case. The loss of memory suggests that, if one considers a large enough coarse grain of the polymer, i.e. one redefines the monomers and their distance b, the correlation is completely erased. There is a particular choice which preserves the contour length. Indeed, we look for a certain coarse-grained bond size b_K and corresponding number of steps N_K such that the two conditions

$$\begin{cases} Nb^2 \dfrac{1+\mathcal{L}}{1-\mathcal{L}} = N_K b_K^2 \\ Nb = N_K b_K \end{cases} \tag{80}$$

are fulfilled. This system of equations is easily solved, giving

$$\begin{cases} b_K = b \dfrac{1+\mathcal{L}}{1-\mathcal{L}} \\ N_K = N \dfrac{1-\mathcal{L}}{1+\mathcal{L}} \end{cases} \tag{81}$$

The coarse-grained polymer is called the *equivalent Freely-Jointed Chain* of the KP model, and b_K is the corresponding *Kuhn length*.

Problem. Show that for a freely-rotating chain $\langle R_e^2 \rangle = Nb^2(1+\cos\theta)/(1-\cos\theta)$. What is the Kuhn length of this model?

5.2 The wormlike chain

The *wormlike chain* (WLC) is obtained as the continuum limit of the discrete KP model, obtained by considering the limits $b \to 0$, $N \to \infty$ while keeping the contour length $L \equiv Nb$ constant. For this reason, it is often called the *(continuous) Kratky-Porod model*. In order to take the continuum limit, we introduce the arclength coordinate s. Accordingly, we switch to the notation $\widehat{\boldsymbol{t}}(s)$ for the tangent vectors, with $\widehat{\boldsymbol{t}}_i$ corresponding to the value $s = ib$. To write the continuum version of the energy in Equation (56), we observe that

$$\widehat{\boldsymbol{t}}((i+1)b) \simeq \widehat{\boldsymbol{t}}(ib) + b\frac{d\widehat{\boldsymbol{t}}}{ds}(ib) + \frac{1}{2}b^2\frac{d^2\widehat{\boldsymbol{t}}}{ds^2}(ib) \; . \tag{82}$$

In this way, the scalar product is

$$\widehat{\boldsymbol{t}}((i+1)b) \cdot \widehat{\boldsymbol{t}}(ib) \simeq 1 + b\,\widehat{\boldsymbol{t}}(ib) \cdot \frac{d\widehat{\boldsymbol{t}}}{ds}(ib) + \frac{1}{2}b^2\widehat{\boldsymbol{t}}(ib) \cdot \frac{d^2\widehat{\boldsymbol{t}}}{ds^2}(ib) \; . \tag{83}$$

Now, we note that

$$\widehat{\boldsymbol{t}} \cdot \frac{d\widehat{\boldsymbol{t}}}{ds} = \frac{1}{2}\frac{d}{ds}\left(\widehat{t}^2\right) = 0 \tag{84}$$

since $\widehat{t}^2 = 1$ by construction. Moreover, we also have

$$\widehat{\boldsymbol{t}} \cdot \frac{d^2\widehat{\boldsymbol{t}}}{ds^2} = \frac{d}{ds}\left(\widehat{\boldsymbol{t}} \cdot \frac{d\widehat{\boldsymbol{t}}}{ds}\right) - \left(\frac{d\widehat{\boldsymbol{t}}}{ds}\right)^2 = -\left(\frac{d\widehat{\boldsymbol{t}}}{ds}\right)^2 \; . \tag{85}$$

With these results, Equation (83) becomes

$$\widehat{\boldsymbol{t}}((i+1)b) \cdot \widehat{\boldsymbol{t}}(ib) \simeq 1 - \frac{1}{2}b^2\left[\frac{d\widehat{\boldsymbol{t}}}{ds}(ib)\right]^2 \; . \tag{86}$$

Substituting into Equation (56) gives the "hybrid" discrete-continuous form of the energy

$$U = -\frac{B}{b}(N-1) + \frac{B}{b}\frac{1}{2}b^2\sum_{i=1}^{N-1}\left[\frac{d\widehat{\boldsymbol{t}}}{ds}(ib)\right]^2 \; . \tag{87}$$

The first term is a constant and can be eliminated by a proper shifting of the zero of the energy, hence giving:

$$U = \frac{B}{b}\frac{1}{2}b^2\sum_{i=1}^{N-1}\left[\frac{d\widehat{\boldsymbol{t}}}{ds}(ib)\right]^2 \; . \tag{88}$$

The continuum limit enables substituting $b\sum_{i=1}^{N-1} \to \int_0^L ds$, thus obtaining

$$U = \frac{B}{2}\int_0^L \left(\frac{d\widehat{\boldsymbol{t}}}{ds}\right)^2 ds \; . \tag{89}$$

The persistence length can also be obtained by taking the limit $b \to 0$ in Equation (67). Since $\mathcal{L}(x) \simeq 1 - 1/x$ for $x \to \infty$, we find

$$\boxed{l_p \simeq -\frac{b}{\ln\left(1 - \frac{b}{\beta B}\right)} \simeq -\frac{b}{-\frac{b}{\beta B}} = \beta B \ ,} \tag{90}$$

so that the energy can be written as

$$\boxed{U = \frac{1}{2} l_p k_B T \int_0^L \left(\frac{d\widehat{\boldsymbol{t}}}{ds}\right)^2 ds \ .} \tag{91}$$

Naturally, the correlation function has to be adapted to the continuum version:

$$\boxed{\left\langle \widehat{\boldsymbol{t}}(s_1) \cdot \widehat{\boldsymbol{t}}(s_2) \right\rangle = e^{-\frac{|s_1 - s_2|}{l_p}} \ .} \tag{92}$$

Problem. Show that the wormlike chain model is obtained also as the continuum limit of the freely-rotating chain model by taking $b \to 0, \theta \to 0$ and b/θ^2 constant.

5.2.1 End-to-end distance

The end-to-end vector is now written as

$$\boldsymbol{R}_e = \int_0^L \widehat{\boldsymbol{t}}(s) ds \ . \tag{93}$$

The average squared end-to-end distance can be computed in a similar way as above:

$$\begin{aligned}
&\left\langle R_e^2 \right\rangle = \int_0^L ds \int_0^L ds' \left\langle \widehat{\boldsymbol{t}}(s) \cdot \widehat{\boldsymbol{t}}(s') \right\rangle = \\
&\int_0^L ds \int_0^L ds' e^{-\frac{|s-s'|}{l_p}} = \\
&2\int_0^L ds \int_0^s ds' e^{-\frac{s-s'}{l_p}} = \\
&2\int_0^L ds \int_0^s ds'' e^{-\frac{s''}{l_p}} = \\
&2\int_0^L ds'' \int_{s''}^L ds\, e^{-\frac{s''}{l_p}} = \\
&2\int_0^L ds''(L-s'') e^{-\frac{s''}{l_p}} = \\
&2Ll_p \left[1 - \frac{l_p}{L}\left(1 - e^{-\frac{L}{l_p}}\right)\right] .
\end{aligned} \tag{94}$$

Comparing first and last term we thus find

$$\boxed{\left\langle R_e^2 \right\rangle = 2Ll_p \left[1 - \frac{l_p}{L}\left(1 - e^{-\frac{L}{l_p}}\right)\right] .} \tag{95}$$

It is interesting to consider two limiting cases. For $L \ll l_p$, the polymer is expected to be very rigid. Considering the expansion of the exponential up to the second order, one finds $\langle R_e^2 \rangle \simeq L^2$ in this limit. This is expected, since in this limit the polymer just behaves as a rigid rod of length L. Importantly, this implies that at short length scales the scaling exponent for the end-to-end distance is equal to 1, which explains the experimental observations reported in Fig.6b. Indeed, the persistence length of DNA is $l_p \simeq 50$ nm, while for amyloid fibrils $l_p \sim 1000$ nm. In the opposite limit $L \gg l_p$, one finds $\langle R_e^2 \rangle \simeq 2Ll_p \propto L$, which recovers the behavior of a FJC. As mentioned above, this is indeed the case for all polymer models with bending, as long as the correlation function decays exponentially and non-bonded interactions can be excluded.

Problem. Show that the Kuhn length of the WLC is $b_K = 2l_p$.

5.3 Tangent vector correlation function and dimensionality

Equation (92) has been derived in three dimensions. It is interesting both theoretically and for experiments to see what happens when other dimensionalities are considered, particularly

in two dimensions. For example, AFM experiments are conducted by adsorbing the molecule of interest onto a surface and letting it relax on the plane, so that for a proper statistical analysis of polymers in two dimensions it is important to have an understanding of the effect of dimensionality d.

Intuitively, when there are less dimensions the molecule has less degrees of freedom to "employ" for bending. Imagine a very stiff WLC ($L \ll l_p$) in three dimensions, oriented towards the z axis. The slight bending of the rodlike molecule can occur in the perpendicular plane, i.e. in the x and y directions. In contrast, if the same molecule is found on a plane, only bending in the perpendicular line is available, thus it is expected that the tangent vector correlation function decays more slowly. Hence, one should more correctly speak of dimension-dependent persistence length. Nevertheless, it is customary to refer to *the* persistence length l_p as the one corresponding to three dimensions, and appearing in the energy of the WLC, Equation (91). In this section, we will instead denote as $\lambda(d)$ the dimension-dependent length scale of the exponential decay of the correlation function, i.e.

$$\left\langle \widehat{\boldsymbol{t}}(s_1) \cdot \widehat{\boldsymbol{t}}(s_2) \right\rangle = e^{-\frac{|s_1 - s_2|}{\lambda(d)}} \ . \tag{96}$$

Naturally, by construction $\lambda(3) = l_p$.

In order to find a general formula for $\lambda(d)$, we consider a rodlike molecule, for which $L \ll l_p$, and we choose the x axis to correspond to its main direction. For our treatment, it is useful to separate the x component t_x of the tangent vector from the rest, which are collectively denoted as $\boldsymbol{t}_\perp$. Note that $\boldsymbol{t}_\perp$ is a vector with $(d-1)$ components. By construction, the tangent vector has magnitude equal to 1, hence $t_x = \sqrt{1 - t_\perp^2}$. Since we are considering a rodlike molecule oriented towards x, we expect that $t_x \simeq 1$, i.e. that $t_\perp$ is very small. Expanding the square root, we can thus write

$$t_x \simeq 1 - \frac{1}{2} t_\perp^2 \ . \tag{97}$$

Based on this, we see that $dt_x/ds \simeq \boldsymbol{t}_\perp \cdot d\boldsymbol{t}_\perp/ds$. Within our perturbative approach, it is reasonable to assume that $\boldsymbol{t}_\perp$ and its derivative have similar magnitudes, so that we expect that $dt_x/ds \boldsymbol{t}_\perp \cdot d\boldsymbol{t}_\perp/ds \ll |d\boldsymbol{t}_\perp/ds|$. Therefore, we can conclude that the integrand in Equation (91) is $(d\widehat{\boldsymbol{t}}/ds)^2 \simeq (d\boldsymbol{t}_\perp/ds)^2$. In other words, the energy now reads

$$U = \frac{1}{2} l_p k_B T \int_0^L \left(\frac{d\widehat{\boldsymbol{t}}_\perp}{ds} \right)^2 ds \ . \tag{98}$$

It is useful to consider the Fourier series[4] of $\boldsymbol{t}_\perp$:

$$\boldsymbol{t}_\perp(s) = \frac{\boldsymbol{a}_0}{2} + \sum_{n=1}^{\infty} \boldsymbol{a}_n \cos \frac{n\pi s}{L} \ . \tag{100}$$

[4]We are considering only the cosines and a period of $2L$ for the Fourier series, obtained by artificially prolonging the curve. The reason for the latter is that a period equal to L would automatically impose that $\boldsymbol{t}_\perp(0) = \boldsymbol{t}_\perp(L)$, which has no reason to be for a free polymer. Now, one should in principle consider the following

The derivative is

$$\frac{d\boldsymbol{t}_\perp}{ds} = -\frac{\pi}{L}\sum_{n=1}^{\infty} n\,\boldsymbol{a}_n \sin\frac{n\pi s}{L}\ , \tag{101}$$

so that its square is equal to

$$\left(\frac{d\boldsymbol{t}_\perp}{ds}\right)^2 = \frac{\pi^2}{L^2}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} n\,m\,\boldsymbol{a}_n\cdot\boldsymbol{a}_m \sin\frac{n\pi s}{L}\sin\frac{m\pi s}{L}\ . \tag{102}$$

The energy thus reads

$$U = \frac{\pi^2}{2}\frac{l_p}{L^2}k_BT\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} n\,m\,\boldsymbol{a}_n\cdot\boldsymbol{a}_m\int_0^L \sin\frac{n\pi s}{L}\sin\frac{m\pi s}{L}ds\ . \tag{103}$$

The integrand can be rewritten by means of Werner's formulas as

$$\sin\frac{n\pi s}{L}\sin\frac{m\pi s}{L} = \frac{1}{2}\left[\cos\frac{(n-m)\pi s}{L} - \cos\frac{(n+m)\pi s}{L}\right]\ . \tag{104}$$

It is evident that the integral of the cosines is almost always zero. The only case in which there is a contribution is when $n = m$, so that the first cosine is identically equal to 1. Therefore, we conclude that

$$\int_0^L \sin\frac{n\pi s}{L}\sin\frac{m\pi s}{L}ds = \frac{L}{2}\delta_{nm}\ . \tag{105}$$

Substituting in the formula for the energy, we thus find

$$U = \frac{\pi^2}{2}\frac{l_p}{L^2}k_BT\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} n\,m\,\boldsymbol{a}_n\cdot\boldsymbol{a}_m\cdot\frac{L}{2}\delta_{nm}\ , \tag{106}$$

that is

$$\boxed{U = \frac{\pi^2}{4}\frac{l_p}{L}k_BT\sum_{n=1}^{\infty} n^2 a_n^2\ .} \tag{107}$$

The energy is quadratic in the Fourier degrees of freedom $\boldsymbol{a}_n$, which are vectors with $d-1$ components. Applying the equipartition theorem, we find that

$$\frac{\pi^2}{4}\frac{l_p}{L}k_BTn^2\left\langle a_n^2\right\rangle = \frac{d-1}{2}k_BT\ , \tag{108}$$

series

$$\boldsymbol{t}_\perp(s) = \frac{\boldsymbol{a}_0}{2} + \sum_{n=1}^{\infty}\boldsymbol{a}_n\cos\frac{n\pi s}{L} + \sum_{n=1}^{\infty}\boldsymbol{b}_n\sin\frac{n\pi s}{L}\ . \tag{99}$$

However, we can arbitrarily consider a specular prolongation, so that the derivative of $\boldsymbol{t}_\perp(s)$ at $s = 0$ is zero by construction, which in fact imposes that the sines disappear. Anyway, keeping the sines would lead to the same final result.

that is

$$\left\langle a_n^2 \right\rangle = \frac{2(d-1)}{\pi^2 n^2} \frac{L}{l_p} \, . \tag{109}$$

The equipartition theorem also ensures that $\langle \boldsymbol{a}_n \cdot \boldsymbol{a}_m \rangle = 0$ if $n \neq m$. Therefore, we can write in general

$$\langle \boldsymbol{a}_n \cdot \boldsymbol{a}_m \rangle = \frac{2(d-1)}{\pi^2 n^2} \frac{L}{l_p} \delta_{nm} \, . \tag{110}$$

Now, let us consider the tangent vector correlation at the two ends of the rodlike molecule. By definition:

$$\left\langle \widehat{\boldsymbol{t}}(0) \cdot \widehat{\boldsymbol{t}}(L) \right\rangle = e^{-\frac{L}{\lambda(d)}} \simeq 1 - \frac{L}{\lambda(d)} \, . \tag{111}$$

The left-hand side can be written by separating the contributions from t_x and $\boldsymbol{t}_\perp$, hence

$$\langle t_x(0) t_x(L) \rangle + \langle \boldsymbol{t}_\perp(0) \cdot \boldsymbol{t}_\perp(L) \rangle = 1 - \frac{L}{\lambda(d)} \, . \tag{112}$$

Now, from Equation (100), we have that (remember that $\cos n\pi = (-1)^n$)

$$\boldsymbol{t}_\perp(0) = \frac{\boldsymbol{a}_0}{2} + \sum_{n=1}^{\infty} \boldsymbol{a}_n \qquad \text{and} \qquad \boldsymbol{t}_\perp(0) = \frac{\boldsymbol{a}_0}{2} + \sum_{n=1}^{\infty} \boldsymbol{a}_n (-1)^n \tag{113}$$

Therefore, we obtain

$$\begin{aligned}
&\langle \boldsymbol{t}_\perp(0) \cdot \boldsymbol{t}_\perp(L) \rangle = \\
&\frac{\langle a_0^2 \rangle}{4} + \frac{1}{2} \sum_{n=1}^{\infty} \cancel{\langle \boldsymbol{a}_0 \cdot \boldsymbol{a}_n \rangle} + \frac{1}{2} \sum_{n=1}^{\infty} (-1)^n \cancel{\langle \boldsymbol{a}_0 \cdot \boldsymbol{a}_n \rangle} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (-1)^n \langle \boldsymbol{a}_m \cdot \boldsymbol{a}_m \rangle = \\
&\frac{\langle a_0^2 \rangle}{4} + \sum_{n=1}^{\infty} (-1)^n \left\langle a_n^2 \right\rangle = \\
&\frac{\langle a_0^2 \rangle}{4} + \frac{2(d-1)}{\pi^2} \frac{L}{l_p} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = \\
&\frac{\langle a_0^2 \rangle}{4} - \frac{2(d-1)}{\pi^2} \frac{L}{l_p} \cdot \frac{\pi^2}{12} = \\
&\frac{\langle a_0^2 \rangle}{4} - \frac{d-1}{6} \frac{L}{l_p} \, .
\end{aligned} \tag{114}$$

In the previous formula, we have made use of Equation (110) and used the known result $\sum_{n=1}^{\infty} (-1)^n / n^2 = -\pi^2/12$. As for the x component, since $t_x \simeq 1 - \frac{1}{2} t_\perp^2$, we have

$$\langle t_x(0) t_x(L) \rangle \simeq 1 - \frac{1}{2} \left\langle t_\perp^2(0) \right\rangle - \frac{1}{2} \left\langle t_\perp^2(L) \right\rangle \, . \tag{115}$$

Let us compute the two averages. The first one reads

$$\begin{aligned}
&\left\langle t_\perp^2(0)\right\rangle = \frac{\langle a_0^2\rangle}{4} + \sum_{n=1}^{\infty} \cancel{\langle \boldsymbol{a}_0 \cdot \boldsymbol{a}_n \rangle} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \langle \boldsymbol{a}_n \cdot \boldsymbol{a}_m \rangle = \\
&\frac{\langle a_0^2\rangle}{4} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{2(d-1)}{\pi^2 n^2}\frac{L}{l_p}\delta_{nm} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{2(d-1)}{\pi^2}\frac{L}{l_p}\sum_{n=1}^{\infty}\frac{1}{n^2} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{2(d-1)}{\pi^2}\frac{L}{l_p}\cdot\frac{\pi^2}{6} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{d-1}{3}\frac{L}{l_p}\,,
\end{aligned} \tag{116}$$

where we used Equation (110) and the known result $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. The second one reads

$$\begin{aligned}
&\left\langle t_\perp^2(L)\right\rangle = \frac{\langle a_0^2\rangle}{4} + \sum_{n=1}^{\infty}(-1)^n \cancel{\langle \boldsymbol{a}_0 \cdot \boldsymbol{a}_n \rangle} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}(-1)^{n+m} \langle \boldsymbol{a}_n \cdot \boldsymbol{a}_m \rangle = \\
&\frac{\langle a_0^2\rangle}{4} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}(-1)^{n+m}\frac{2(d-1)}{\pi^2 n^2}\frac{L}{l_p}\delta_{nm} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{2(d-1)}{\pi^2}\frac{L}{l_p}\sum_{n=1}^{\infty}\frac{(-1)^{2n}}{n^2} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{2(d-1)}{\pi^2}\frac{L}{l_p}\sum_{n=1}^{\infty}\frac{1}{n^2} = \\
&\frac{\langle a_0^2\rangle}{4} + \frac{d-1}{3}\frac{L}{l_p} = \left\langle t_\perp^2(0)\right\rangle\,.
\end{aligned} \tag{117}$$

Note that $\langle t_\perp^2(0)\rangle = \langle t_\perp^2(L)\rangle$, as expected by symmetry. Using these results, we can compute the x contribution to the tangent vector correlation function:

$$\langle t_x(0)t_x(L)\rangle \simeq 1 - \frac{1}{2}\left\langle t_\perp^2(0)\right\rangle - \frac{1}{2}\left\langle t_\perp^2(L)\right\rangle = 1 - \left\langle t_\perp^2(0)\right\rangle = 1 - \frac{\langle a_0^2\rangle}{4} - \frac{d-1}{3}\frac{L}{l_p}\,. \tag{118}$$

Inserting Equation (114) and Equation (118) in Equation (112), we obtain

$$\left(1-\frac{\langle a_0^2\rangle}{4}-\frac{d-1}{3}\frac{L}{l_p}\right)+\left(\frac{\langle a_0^2\rangle}{4}-\frac{d-1}{6}\frac{L}{l_p}\right)=1-\frac{L}{\lambda(d)}\Rightarrow$$

$$1-\frac{d-1}{2}\frac{L}{l_p}=1-\frac{L}{\lambda(d)}\Rightarrow \tag{119}$$

$$\boxed{\lambda(d)=\frac{2}{d-1}l_p}\,.$$

This is the final result that we were looking for. Note that for $d=3$ we recover $\lambda(3)=l_p$, as it should. An intuitive explanation is that λ is proportional to the available degrees of freedom for bending, which are $d-1$. This, together with the condition $\lambda(3)=l_p$, gives immediately our final result. Finally, we note that for the experimentally-relevant case $d=2$, we obtain $\lambda(2)=2l_p$, i.e. the tangent vector correlation function decays with the contour length more slowly by a factor of two when compared to three dimensions. We thus have

$$\boxed{\left\langle \widehat{\boldsymbol{t}}(s_1)\cdot\widehat{\boldsymbol{t}}(s_2)\right\rangle = e^{-\frac{|s_1-s_2|}{2l_p}}} \qquad \text{in two dimensions .} \tag{120}$$

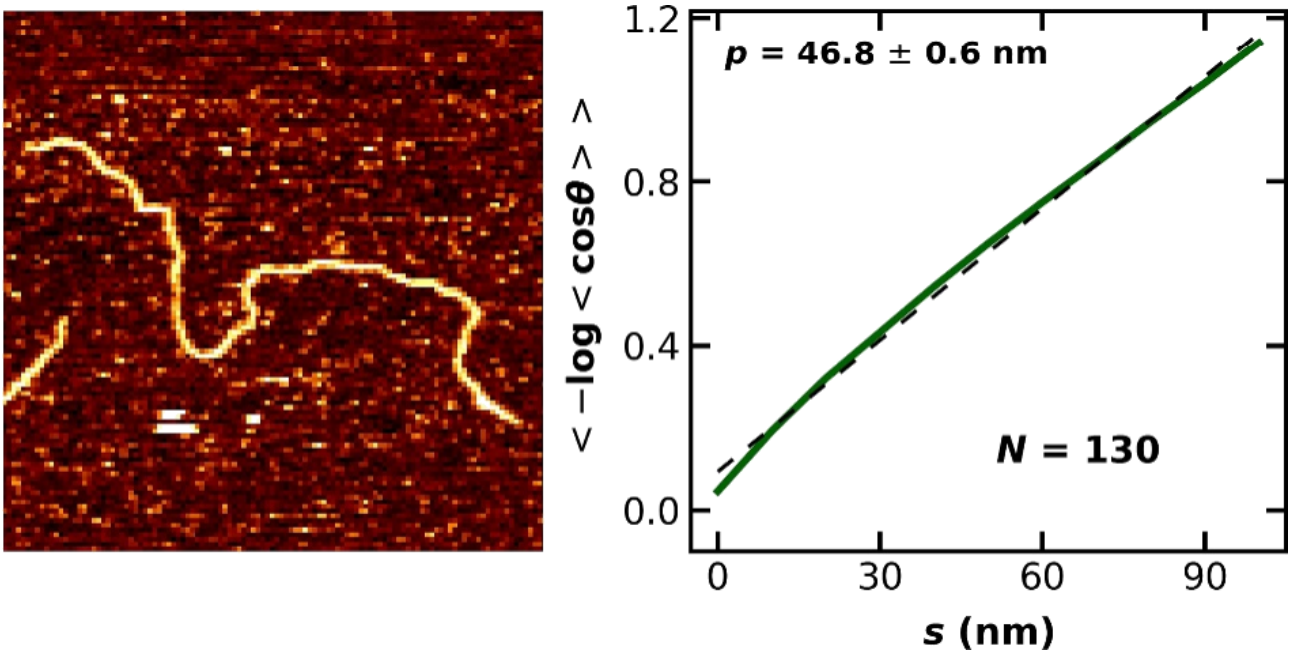

Figure 7: Left: AFM image of a DNA molecule absorbed onto a mica surface. Right: $-\ln\left(\left\langle \widehat{\boldsymbol{t}}(s_0)\cdot\widehat{\boldsymbol{t}}(s_0+s)\right\rangle\right)$ as a function of the contour-length distance s. Adapted from P. R. Heenan and T. T. Perkins, *ACS Nano* **13**:4220 (2019).

As an example of the application of this result, in Fig.7 we report a representative AFM image of DNA adsorbed onto mica and the corresponding correlation function as a function of the contour length. In the considered case, it was found that $l_p \simeq 47$ nm, which is in the consensus range for DNA.

6 Polymer response to a stretching force

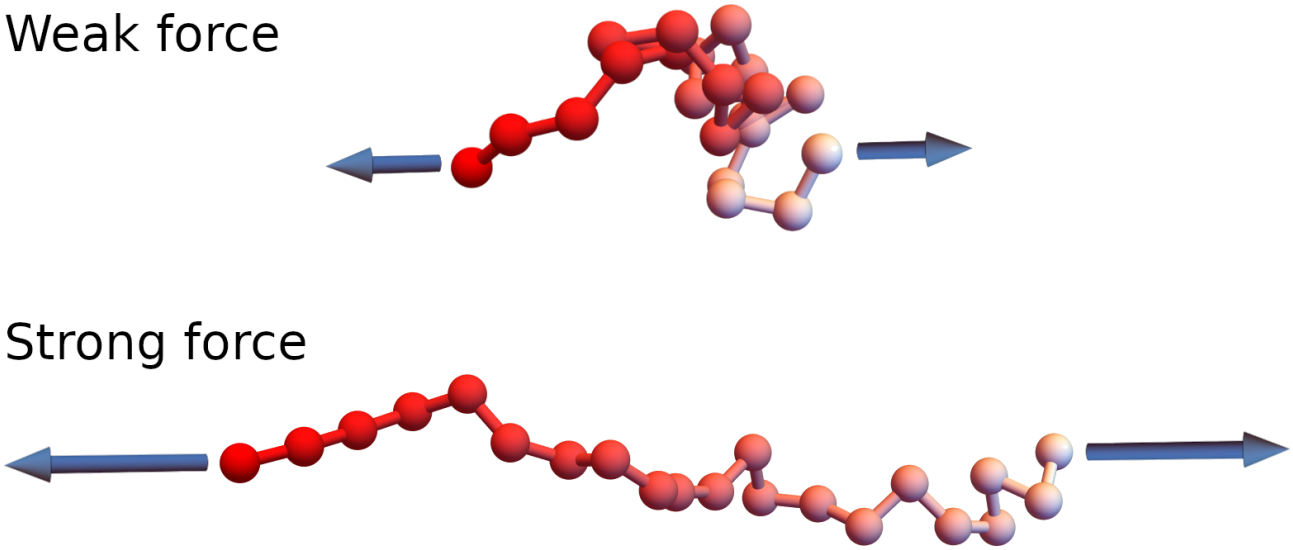

Figure 8: Representative snapshots of polymer conformations under the action of a couple of opposing forces.

In this section, we study the response of a polymer under the presence of two opposing forces with constant magnitude f applied to the end monomers (Fig.8). In order to simplify our wording, we will always speak of "a pulling force", but you have to keep in mind that there is an opposing one on the other end of the polymer.

6.1 Pulling a Freely-Jointed Chain

6.1.1 Free-energy approach

We start our treatment by considering the simple case of a FJC. In this regard, remember that the entropic contribution to the free energy is k_BTR^2/Nb^2, where R is identified with the end to end distance. From Equation (12), the entropic contribution F_{ent} to the free energy of a FJC is

$$F_{\text{ent}} = -k_BT \ln P(\boldsymbol{R}_e) = \frac{3}{2} k_BT \frac{R_e^2}{Nb^2} \ . \tag{121}$$

Under the presence of a pulling force, the enthalpic contribution to the free energy is minus the work performed by the force, $-\boldsymbol{f} \cdot \boldsymbol{R}_e$. The free energy to be minimized is thus

$$F = \frac{3}{2} k_BT \frac{R_e^2}{Nb^2} - \boldsymbol{f} \cdot \boldsymbol{R}_e \ . \tag{122}$$

The minimizing end-to-end vector is obtained by imposing $\nabla F = 0$:

$$\frac{3k_BT}{Nb^2}\boldsymbol{R}_e - \boldsymbol{f} = 0 \Rightarrow \boxed{\boldsymbol{R}_e = \frac{Nb^2}{3k_BT}\boldsymbol{f}}\,. \tag{123}$$

As one could have expected, this result shows that the average end-to-end vector is aligned with the external force. Moreover, its magnitude is linearly dependent on f, i.e. the FJC behaves as a spring with constant $k_{\text{ent}} = 3k_BT/(Nb^2)$. Note that this elastic response comes purely from entropy (as suggested by the presence of k_BT in the elastic constant), so that in the literature it is often said that the polymer is behaving as an *entropic spring*.

6.1.2 Scaling approach

The entropic spring can also be obtained by means of a scaling argument ("tension blobs"). Indeed, we note that on average any fragment of the polymer is receiving the same couple of forces applied to the end. Thinking in terms of length scales, we note that the energy corresponding to the work of the force increases with the size of the considered fragment as $f\xi$, where ξ is the end-to-end distance of the fragment under consideration, which contains a certain number g of monomers. This suggests that the force overcomes entropy only at large scales, while at small scales the entropy dominates. As usual, we can estimate the entropic contribution as k_BT, so that entropy dominates whenever $k_BT \gg f\xi$. Under this condition, the FJC corresponding to the fragment is practically unperturbed by the force, so that the usual FJC statistics hold, i.e. $\xi \sim bg^{1/2} \Rightarrow g \sim (\xi/b)^2$. In contrast, for the length scales such that $k_BT \ll f\xi$ the chain is practically aligned with the force. The threshold between these two regimes is obtained for the length scale $\xi^* \sim k_BT/f$. Hence, one can intuitively think of the conformation of the polymer as a straight arrangement of "blobs" of size ξ^*, which are internally organized as an unperturbed FJC. Quantitatively, the number of monomers in a blobs is $g^* \sim (\xi^*/b)^2$, so that there are in total $N/g^* \sim N(b/\xi^*)^2$ blobs. Since all the blobs are aligned with the force and each of them has size ξ^*, we conclude that the extension of the polymer is $R \sim (N/g^*)\xi^* \sim Nb^2/\xi^* \sim (Nb^2/k_BT)f$, which is the same result as Equation (123) apart from a numerical prefactor.

Problem. By means of the scaling approach, estimate the response of a polymer in a good solvent where for simplicity the excluded volume is $v = b^3$. Compared to the FJC, is it easier or harder to stretch such a polymer? How can you explain the result intuitively?

6.1.3 Exact solution

The entropic spring is by no means the exact result. This can be easily understood by observing that for $f \to \infty$ Equation (123) predicts that $R_e \to \infty$. However, the maximum theoretical value of R_e corresponds to the contour length Nb, so that something is up. In this regard, we note that F_{ent} corresponds to the entropy of a FJC *in the absence of the pulling force*. The

presence of $\boldsymbol{f}$ modifies the conformational space of the polymer and its entropy with it. It is reasonable to assume that small forces only slightly perturb the probability distribution of $\boldsymbol{R}_e$, so that we expect that the entropic spring correctly describes the response of a FJC at small forces. What is the meaning of "small"? An answer is suggested by the scaling argument considered above. The size of a blob was $\xi^* \sim k_B T/f$ and it was assumed that within a blob the usual statistics of a FJC holds. An implicit assumption of this argument is that the blob contains enough monomers to actually form a FJC on its own! Therefore, we expect the argument to work as long as $\xi^* \gg b$, i.e. for forces such that $k_B T/f \gg b \Rightarrow f \ll k_B T/b$.

Luckily enough, it is possible to compute exactly the response of a FJC to a force. Even more strikingly, we do not even need that $N \gg 1$, i.e. the result holds for any chain length. The calculation is based on the observation that the end-to-end vector can be written as

$$\boldsymbol{R}_e = b \sum_{i=1}^{N} \widehat{\boldsymbol{t}}_i \; . \tag{124}$$

The energy is $-\boldsymbol{f} \cdot \boldsymbol{R}_e$, so that the partition function $\mathcal{Z}_N(f)$ of the system can be written as[5]

$$\begin{aligned} \mathcal{Z}_N(f) &= \int d\widehat{\boldsymbol{t}}_1 \int d\widehat{\boldsymbol{t}}_2 \cdots \int d\widehat{\boldsymbol{t}}_N e^{\beta \boldsymbol{f} \cdot \boldsymbol{R}_e} = \\ & \int d\widehat{\boldsymbol{t}}_1 \int d\widehat{\boldsymbol{t}}_2 \cdots \int d\widehat{\boldsymbol{t}}_N e^{\sum_{i=1}^{N} \beta b \boldsymbol{f} \cdot \widehat{\boldsymbol{t}}_i} = \\ & \left(\int d\widehat{\boldsymbol{t}}_1 e^{\beta b \boldsymbol{f} \cdot \widehat{\boldsymbol{t}}_1} \right) \left(\int d\widehat{\boldsymbol{t}}_2 e^{\beta b \boldsymbol{f} \cdot \widehat{\boldsymbol{t}}_2} \right) \ldots \left(\int d\widehat{\boldsymbol{t}}_1 e^{\beta b \boldsymbol{f} \cdot \widehat{\boldsymbol{t}}_N} \right) \; . \end{aligned} \tag{125}$$

All the terms in parenthesis are equal to each other, so that

$$\mathcal{Z}_N(f) = [\mathcal{Z}_1(f)]^N \tag{126}$$

with

$$\mathcal{Z}_1(f) = \int d\widehat{\boldsymbol{t}} \, e^{\beta b \boldsymbol{f} \cdot \widehat{\boldsymbol{t}}} \tag{127}$$

corresponding to the response of a single step. In a sense, the FJC is the ideal gas of polymers, where all the particles (in this case, the tangent vectors) are independent of each other!

The computation of $\mathcal{Z}_1$ is straightforward and it is analogous to the response of a dipole in an electric field. The tangent vector has unit size, so that we can write $d\widehat{\boldsymbol{t}} = \sin\theta d\theta d\phi$. Considering a reference frame with the z axis oriented along $\boldsymbol{f}$, we obtain

$$\mathcal{Z}_1(f) = \int_0^{\pi} d\theta \int_0^{2\pi} d\phi \, e^{\beta b f \cos\theta} \sin\theta = 2\pi \int_{-1}^{1} e^{\beta b f x} dx = 4\pi \frac{\sinh \beta b f}{\beta b f} \; , \tag{128}$$

[5] we omit the overall position of the center of mass, since it does not affect the response to the force

where in the second step we substituted $x = \cos\theta$. The overall partition function of the chain is thus

$$\mathcal{Z}_N(f) = \left(4\pi \frac{\sinh \beta b f}{\beta b f}\right)^N \,. \tag{129}$$

The average end-to-end vector is obtained as

$$\langle \boldsymbol{R}_e \rangle = \frac{1}{\mathcal{Z}_N} \int d\widehat{\boldsymbol{t}}_1 \int d\widehat{\boldsymbol{t}}_2 \cdots \int d\widehat{\boldsymbol{t}}_N \boldsymbol{R}_e e^{\beta \boldsymbol{f}\cdot\boldsymbol{R}_e} \,. \tag{130}$$

We note that the chain of integrals at the numerator is equal to $(\nabla_{\boldsymbol{f}} \mathcal{Z}_N)/\beta = k_B T \nabla_{\boldsymbol{f}} \mathcal{Z}_N$, where $\nabla_{\boldsymbol{f}} = \left(\frac{\partial}{\partial f_x}, \frac{\partial}{\partial f_y}, \frac{\partial}{\partial f_z}\right)$. We also note that

$$\nabla_{\boldsymbol{f}} f = \nabla_{\boldsymbol{f}} \sqrt{f_x^2 + f_y^2 + f_z^2} = \left(\frac{f_x}{f}, \frac{f_y}{f}, \frac{f_z}{f}\right) = \frac{\boldsymbol{f}}{f} = \widehat{\boldsymbol{z}} \,, \tag{131}$$

since we chose the reference frame with z oriented along the direction of the force. Based on these observations, we can thus write

$$\begin{aligned}
&\langle \boldsymbol{R}_e \rangle = \frac{1}{\mathcal{Z}_N} \cdot k_B T \nabla_{\boldsymbol{f}} \mathcal{Z}_N = \\
&k_B T \cdot \left(\frac{\beta b f}{\sinh \beta b f}\right)^N \frac{1}{(4\pi)^N} \cdot \frac{\partial \mathcal{Z}_N}{\partial (\beta b f)} \cdot \frac{\partial (\beta b f)}{\partial f} \cdot \nabla_{\boldsymbol{f}} f = \\
&k_B T \cdot \left(\frac{\beta b f}{\sinh \beta b f}\right)^N \frac{1}{\cancel{(4\pi)^N}} \cdot \cancel{(4\pi)^N} N \left(\frac{\sinh \beta b f}{\beta b f}\right)^{N-1} \left[\frac{\cosh \beta b f}{\beta b f} - \frac{\sinh \beta b f}{(\beta b f)^2}\right] \cdot \frac{\partial (\beta b f)}{\partial f} \cdot \widehat{\boldsymbol{z}} = \\
&N \cancel{k_B T} \cdot \frac{\beta b f}{\sinh \beta b f} \left[\frac{\cosh \beta b f}{\beta b f} - \frac{\sinh \beta b f}{(\beta b f)^2}\right] \cdot \cancel{\beta} b \, \widehat{\boldsymbol{z}} = \\
&N b \left(\frac{1}{\tanh \beta b f} - \frac{1}{\beta b f}\right) \widehat{\boldsymbol{z}} \,.
\end{aligned} \tag{132}$$

We note that $\mathcal{L}(\beta b f) = \frac{1}{\tanh \beta b f} - \frac{1}{\beta b f}$ is Langevin's function. The average end-to-end vector is thus

$$\boxed{\langle \boldsymbol{R}_e \rangle = N b \mathcal{L}(\beta b f)\, \widehat{\boldsymbol{z}} \,.} \tag{133}$$

First of all, we note that $\langle \boldsymbol{R}_e \rangle$ is parallel to $\boldsymbol{f}$, as it should. Second, since $\mathcal{L} < 1$ at any point, the end-to-end distance is always lower than the contour length Nb. Moreover $\mathcal{L}(\infty) = 1$, i.e. this upper bound is reached asymptotically at infinite forces. In the opposite limit of $\beta b f \ll 1$, we have $\mathcal{L}(\beta b f) \simeq \beta b f/3$, i.e. $\langle \boldsymbol{R}_e \rangle \simeq N b^2/(3 k_B T)\boldsymbol{f}$, which corresponds to Equation (123). This confirms out physical intuition that the approach considered above, based on the entropy of an

unperturbed FJC, is valid at small forces. Also, we note that the condition for the approximate result to hold is that $\beta b f \ll 1 \Rightarrow f \ll k_B T/b$, which corresponds to what predicted based on our scaling considerations.

6.2 Pulling a Wormlike Chain

In the case of a WLC, the calculation is not as easy as for the FJC, due to the presence of the bending energy. Indeed, at present there is no exact solution to the problem, although several approximate results have been shown to give reasonable agreement to the numerical solution. However, they are not general, as it is assumed that the chain is very long, $L \gg l_p$. These results are based on interpolating between the limits of weak and strong forces, which can be computed analytically.

6.2.1 Weak-force regime

For weak forces, we can use the same free-energy approach as above. Note that, since $L \gg l_p$, one can consider the equivalent FJC with bond size equal to the Kuhn length $b_K = 2l_p$ and $N_K = L/b_K$ bonds, which will have the same entropy of the WLC in the unperturbed case. As a result, we can apply directly Equation (123) and write[6]

$$\langle R_e \rangle \simeq \frac{N_K b_K^2}{3k_B T} f = \frac{2Ll_p}{3k_B T} f \Rightarrow \boxed{\frac{\langle R_e \rangle}{L} = \frac{2}{3}\beta f l_p}. \tag{134}$$

Note that the equivalent FJC also indicates the meaning of "weak force", indicating that $\beta b_K f \ll 1 \Rightarrow \beta f l_p \ll 1$. To better understand this regime intuitively, we can resort to the blob picture: for small forces, the tension blobs are so large that, even by coarse-graining the chain by considering the Kuhn monomers, the Maxwell demon in charge of pulling will not detect any difference. This picture breaks down as soon as the force is strong enough for the tension blob to become comparable to l_p (or b_K).

Problem. Find the weak-force response for the discrete Kratky-Porod model and for the freely-rotating chain model.

6.2.2 Strong-force regime

In the strong-force regime, one considers the opposite limit in which the polymer is practically straight. This enables performing a Fourier series expansion analogous to the one performed above. However, in the present case we can assume that the action of the pulling force at the

[6]Forcing a bit the notation, we are writing here $\langle R_e \rangle$ to denote the magnitude of the average end-to-end vector $\langle \boldsymbol{R}_e \rangle$, *not* to indicate the average end-to-end distance. This difference is subtle but important: at force $f = 0$, the former is zero, while the latter is not.

two ends imposes that the tangent vector is parallel to the force itself for $s = 0, L$. This means that rather than using Equation (99), the sine Fourier expansion is considered

$$\boldsymbol{t}_{\perp}(s) = \sum_{n=1}^{\infty} \boldsymbol{b}_n \sin \frac{n\pi s}{L} \ . \tag{135}$$

By proceeding in the same way, and remembering the Werner formula $\cos\alpha\cos\beta = [\cos(\alpha-\beta)+\cos(\alpha+\beta)]/2$, it is straightforward to show that the bending energy is still given by Equation (107), although now with the sine amplitudes b_n instead of a_n. The total energy of the system is obtained by adding $-\boldsymbol{R}_e \cdot \boldsymbol{f}$. Choosing a reference frame with the x axis oriented along the force, as done above, we have $-\boldsymbol{R}_e \cdot \boldsymbol{f} = -(R_e)_x f$. However, by definition of tangent vector, we also have that $(R_e)_x = \int_0^L t_x(s)ds$. Hence, the total energy (bending plus work performed by the force) is

$$U = \frac{\pi^2}{4}\frac{l_p}{L} k_B T \sum_{n=1}^{\infty} n^2 b_n^2 - f \cdot \int_0^L t_x(s)ds \ . \tag{136}$$

Remembering that $t_x = 1 - t_{\perp}^2/2$, we can write for the pulling term

$$-f \cdot \int_0^L t_x(s)ds = -fL + \frac{1}{2} f \cdot \int_0^L t_{\perp}^2(s)ds \ . \tag{137}$$

Using Equation (135) and proceeding as usual, we obtain

$$-f \cdot \int_0^L t_x(s)ds = -fL + \frac{1}{2} f \cdot \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \boldsymbol{b}_n \cdot \boldsymbol{b}_m \int_0^L \sin\frac{n\pi s}{L}\sin\frac{m\pi s}{L} = -fL + \frac{1}{4} fL \sum_{n=1}^{\infty} b_n^2 \ . \tag{138}$$

The first term can be neglected by suitably shifting the zero of the energy. We thus obtain

$$U = \frac{\pi^2}{4}\frac{l_p}{L} k_B T \sum_{n=1}^{\infty} n^2 b_n^2 + \frac{1}{4} fL \sum_{n=1}^{\infty} b_n^2 = \sum_{n=1}^{\infty} \left(\frac{\pi^2}{4}\frac{l_p}{L} k_B T n^2 + \frac{1}{4} fL \right) b_n^2 \ . \tag{139}$$

Using the equipartition theorem, we thus obtain

$$\left(\frac{\pi^2}{4}\frac{l_p}{L} k_B T n^2 + \frac{1}{4} fL \right) \left\langle b_n^2 \right\rangle = 2 \cdot \frac{1}{2} k_B T \Rightarrow \left\langle b_n^2 \right\rangle = \frac{4}{\pi^2 \frac{l_p}{L} n^2 + \beta f L} \ . \tag{140}$$

The previous formula can be further rearranged as

$$\left\langle b_n^2 \right\rangle = \frac{4L}{\pi^2 l_p} \frac{1}{n^2 + c} \ , \tag{141}$$

where

$$c \equiv \beta f L \cdot \frac{L}{\pi^2 l_p} \gg 1 \ . \tag{142}$$

Now, the average elongation of the WLC along the force is[7]

$$\langle R_e \rangle = \int_0^L \langle t_x(s) \rangle \, ds = L \left(1 - \frac{1}{4} \sum_{n=1}^{\infty} \langle b_n^2 \rangle \right) \ . \tag{143}$$

A known result is that

$$\sum_{n=1}^{\infty} \frac{1}{n^2 + c} = \frac{-1 + \frac{\sqrt{c}\pi}{\tanh\left(\sqrt{c}\pi\right)}}{2c} \ . \tag{144}$$

Since $c \gg 1$, we can keep the leading term and write

$$\sum_{n=1}^{\infty} \frac{1}{n^2 + c} \simeq \frac{\sqrt{c}\pi}{2c} = \frac{\pi}{2\sqrt{c}} \ . \tag{145}$$

The elongation is thus

$$\langle R_e \rangle = L \left(1 - \frac{1}{4} \cdot \frac{4L}{\pi^2 l_p} \cdot \frac{\pi}{2\sqrt{c}} \right) = L \left(1 - \frac{L}{2\pi l_p} \cdot \sqrt{\frac{\pi^2 l_p}{\beta f L^2}} \right) = L \left(1 - \sqrt{\frac{1}{4\beta f l_p}} \right) \ . \tag{146}$$

Taking only first and last side of the previous formula gives

$$\boxed{\langle R_e \rangle = L \left(1 - \sqrt{\frac{1}{4\beta f l_p}} \right)} \ . \tag{147}$$

6.2.3 General formulas

Based on Equation (134) and Equation (147), an interpolation formula was derived in J. Marko and E. Siggia, *Macromolecules* **28**:8759 (1995), connecting the combination $\beta f l_p$ to the average elongation $\langle R_e \rangle$:

$$\beta f l_p = \frac{\langle R_e \rangle}{L} + \frac{1}{4 \left(1 - \frac{\langle R_e \rangle}{L} \right)^2} - \frac{1}{4} \ . \tag{148}$$

This formula has the right limiting behaviors, retrieving Equation (134) for $\beta f l_p \ll 1$ and Equation (147) for $\beta f l_p \gg 1$. Moreover, a comparison with the numerical solution of the problem shows that the maximum error is equal to roughly 15%. Based on this, a more precise formula was obtained in Bouchiat et al., *Biophys. J.* **76**:409 (1999):

$$\beta f l_p = \frac{\langle R_e \rangle}{L} + \frac{1}{4 \left(1 - \frac{\langle R_e \rangle}{L} \right)^2} - \frac{1}{4} + \sum_{i=2}^{7} \alpha_i \left(\frac{\langle R_e \rangle}{L} \right)^i \ , \tag{149}$$

[7] The same comment as above yields: here, $\langle R_e \rangle$ denotes the magnitude of the average end-to-end vector $\langle \boldsymbol{R}_e \rangle$.

where $\alpha_2 = -0.5164228, \alpha_3 = -2.737418, \alpha_4 = 16.07497, \alpha_5 = -38.87607, \alpha_6 = 39.49944$ and $\alpha_7 = -14.17718$. In Fig.9, we show the comparison between experimental data obtained by pulling DNA in an opical-tweezers experiment and theory. Particularly, Equation (148) was used by fitting L and l_p (continuous line), while Equation (133) (dashed line) was fitted by tuning N and b in order to best reproduce the weak-force regime. It is evident that the WLC describes better the experimental data, which are captured quantitatively at all forces. Quite generally the WLC formula works better for biomolecules, although this is not always the case. It is also instructive to compare the FJC and the WLC. For a given force, the FJC is more extended, meaning that it is in general softer than a WLC. This makes sense, since the continuous character of the WLC introduces more modes that need to be aligned to the force, while in the case of the FJC the lowest possible length scale is set by the bond size b.

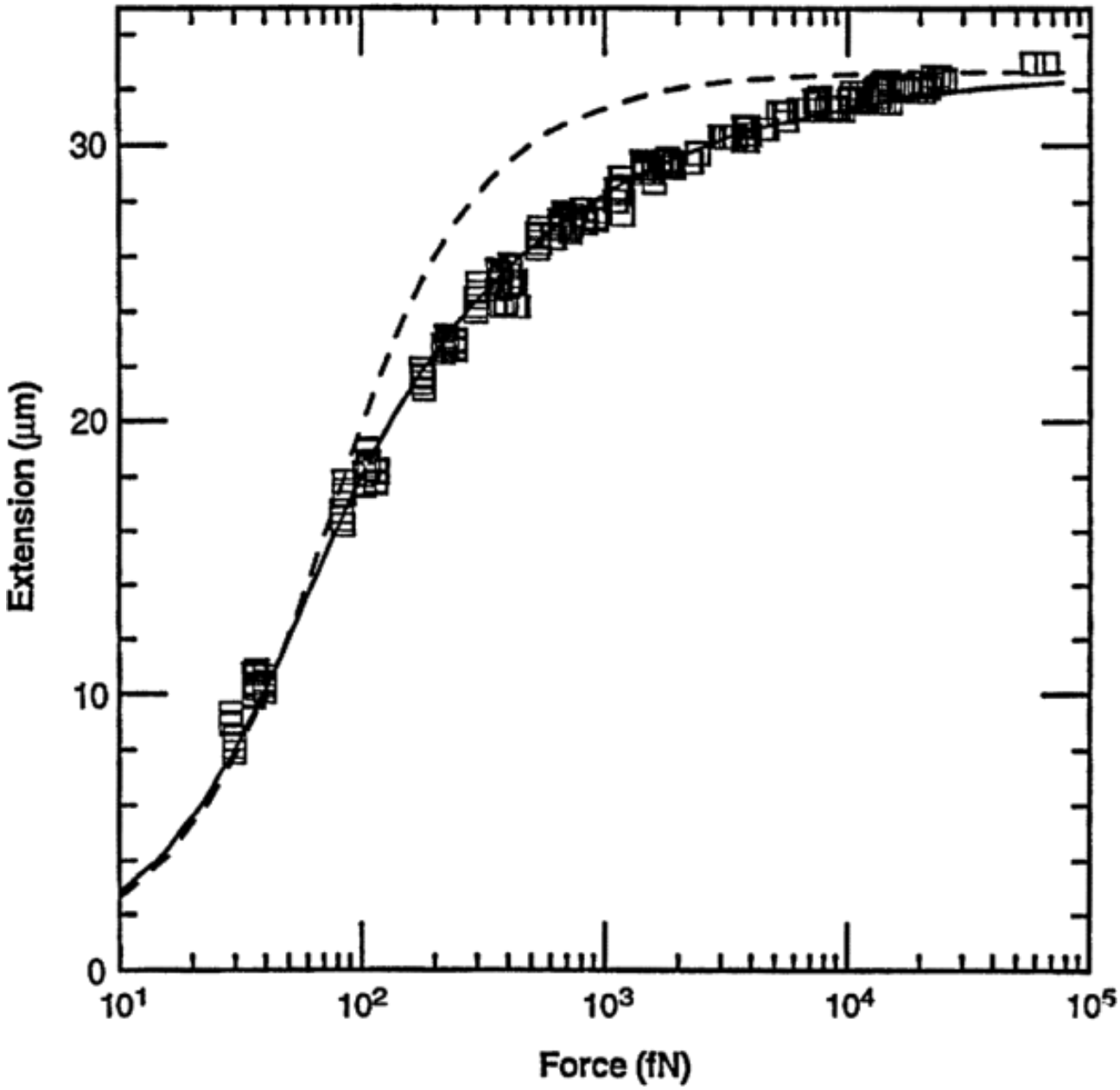

Figure 9: Fit of experimental pulling curves of DNA by means of Equation (148) (continuous line) shows that the WLC can capture quantitatively the response of DNA to a pulling force. The fit was performed by tuning the contour length ($L \simeq 3280$ nm) and the persistence length ($l_p \simeq 53$ nm). Note that the value of l_p is in good agreement with the value extracted from the decay of the tangent vector correlation function (Fig.7). The dashed line shows the FJC formula, Equation (133), optimized to reproduce the weak-force behavior. Reprinted from A. Marantan and L. Mahadevan, *Am. J. Phys.* **86**:86 (2018).

Further reading

- M. Rubinstein, R. Colby, "Polymer Physics", Oxford University Press (2003), Chapters 2,3.
- P.-G. de Gennes, "Scaling Concepts in Polymer Physics", Cornell University Press (1979).
- M. Doi and S. Edwards, "The Theory of Polymer Dynamics", Oxford University Press (1986).
- J. Marko and E. Siggia, *Macromolecules* **28**:8759 (1995).
- S. Assenza and R. Mezzenga, *Nature Rev. Phys.* **1**:551 (2019).